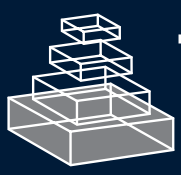
frontiers

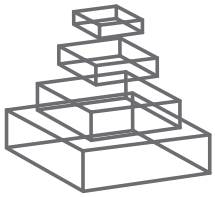
RESEARCH TOPICS

EMBODIED AND GROUNDED COGNITION

Hosted by
Anna M. Borghi and Diane Pecher



frontiers in PSYCHOLOGY



frontiers

FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2012
Frontiers Media SA.
All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, as well as all content on this site is the exclusive property of Frontiers. Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Articles and other user-contributed materials may be downloaded and reproduced subject to any copyright or other notices. No financial payment or reward may be given for any such reproduction except to the author(s) of the article concerned.

As author or other contributor you grant permission to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-013-3
DOI 10.3389/978-2-88919-013-3

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

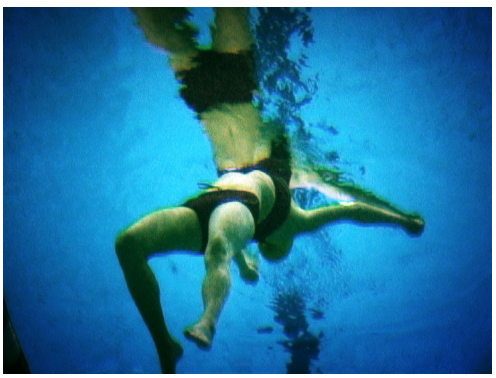
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

EMBODIED AND GROUNDED COGNITION

Hosted By
Anna M. Borghi, University of Bologna & Institute of Cognitive Sciences and Technologies, ISTC, Rome, Italy
Diane Pecher, Erasmus University Rotterdam, Netherlands



ZimmerFrei, *Watersleep*, frame da video, 2003

In the last 10-15 years, the "embodied" and "grounded" cognition approach has become widespread in all fields related to cognitive science, such as cognitive and social psychology, neuroscience, philosophy, anthropology, computational modelling and robotics. According to this approach, our cognitive activity is grounded in sensory-motor processes and situated in specific contexts and situations. Therefore, in this view, concepts consist of the reactivation of the same neural pattern that is present when we perceive and/or interact with the objects they refer to. In the same way, understanding language would imply forming a mental simulation of what is linguistically described. This simulation would entail the recruitment of the same neurons that are activated when actually acting or perceiving the situation, action, emotion, object or entity described by language. In the last years a lot of evidence has been collected in favour of EC and GC view.

The aim of this Research Topic is twofold. First, it intends to give an idea of the field of embodied and grounded cognition in its broadness. We therefore intend to invite scientists of different disciplines (anthropology, philosophy, linguistics, cognitive and social psychology, neuroscience, computer science) to submit their proposals.

The second aim of this Research Topicis to focus on some challenges that in our opinion embodied and grounded theories of cognition need to face. First, we believe that one important challenge for EC and GC views is to account for the way the so-called "abstract concepts" and abstract words are represented. Evidence on the representation of concrete concepts and words is compelling, whereas evidence on abstract concepts representation is still too scarce and limited to restricted domains. We therefore welcome proposals dealing with this complex issue.

Second, we think that embodied and grounded theories of cognition would need to formulate more precise hypotheses, and that in general within the field a larger theoretical effort should be made. It is striking that, even if a lot of work in the field of computational modelling and robotics starts from an embodied approach, experimental and modelling work on embodied cognition remain somehow separate. We therefore invite researchers to submit papers proposing models which might help to explain phenomena as well as to constrain and specify in more detail the predictions and the claims advanced within the framework of EC and GC theories

Table of Contents

05 ***Introduction to the Special Topic Embodied and Grounded Cognition***
Anna M. Borghi and Diane Pecher

08 ***Some Insults are Easier to Detect: The Embodied Insult Detection Effect***
Michele Wellsby, Paul D. Siakaluk, Penny M. Pexman and William J. Owen

21 ***Switching Modalities in A Sentence Verification Task: ERP Evidence for Embodied Language Processing***
Lea A. Hald, Julie-Ann Marshall, Dirk P. Janssen and Alan Garnham

36 ***Modality Switching in a Property Verification Task: An ERP Study of What Happens When Candles Flicker after High Heels Click***
Jennifer Collins, Diane Pecher, René Zeelenberg and Seana Coulson

46 ***Top-Down and Bottom-Up Contributions to Understanding Sentences Describing Objects in Motion***
Shirley-Ann Rueschemeyer, Arthur M. Glenberg, Michael P. Kaschak, Karsten Mueller and Angela D. Friederici

57 ***Context Effects in Embodied Lexical-Semantic Processing***
Wessel O. van Dam, Shirley-Ann Rueschemeyer, Oliver Lindemann and Harold Bekkering

63 ***Deriving Motor Primitives Through Action Segmentation***
Paul E. Hemeren and Serge Thill

74 ***The Embodied Statistician***
Elizabeth R. Marsh and Arthur M. Glenberg

79 ***Embodied Conceptual Combination***
Dermot Lynott and Louise Connell

93 ***How "social" is the social Simon effect?***
Thomas Dolk, Bernhard Hommel, Lorenza S. Colzato, Simone Schütz-Bosbach, Wolfgang Prinz and Roman Liepelt

102 ***The Virtual Co-Actor: The Social Simon Effect does not Rely on Online Feedback from the Other***
Elize Vlainic, Roman Liepelt, Lorenza S. Colzato, Wolfgang Prinz and Bernhard Hommel

108 ***The Two Forms of Visuo-Spatial Perspective Taking are Differently Embodied and Subserve Different Spatial Prepositions***
Klaus Kessler and Hannah Rutherford

120 ***Embodied Language Comprehension Requires an Enactivist Paradigm of Cognition***
Michiel van Elk, Marc Slors and Harold Bekkering

129 ***On the need for embodied and dis-embodied cognition***
Guy Dove

142 ***A Connectionist Approach to Embodied Conceptual Metaphor***
Stephen J. Flusberg, Paul H. Thibodeau, Daniel A. Sternberg and Jeremy J. Glick

153 ***Are temporal concepts embodied? A challenge for cognitive neuroscience***
Alexander Kranjec and Anjan Chatterjee

162 ***Numbers in Space: Differences between Concrete and Abstract Situations***
Diane Pecher and Inge Boot

173 ***Manipulating Objects and Telling Words: A Study on Concrete and Abstract Words Acquisition***
Anna M. Borghi, Andrea Flumini, Felice Cimatti, Davide Marocco and Claudia Scorolli

187 ***Architecture of Explanatory Inference in the Human Prefrontal Cortex***
Aron K. Barbey and Richard Patterson

196 ***The Mechanics of Embodiment: A Dialog on Embodiment and Computational Modeling***
Giovanni Pezzulo, Lawrence W. Barsalou, Angelo Cangelosi, Martin H. Fischer, Ken McRae and Michael J. Spivey

217 ***Credit Assignment in Multiple Goal Embodied Visuomotor Behavior***
Constantin A. Rothkopf and Dana H. Ballard

230 ***A Computational Model of the Lexical-Semantic System Based on a Grounded Cognition Approach***
Mauro Ursino, Cristiano Cuppini and Elisa Magosso

249 ***The Other Half of the Embodied Mind***
Domenico Parisi

257 ***The Neuroscience of Storing and Molding Tool Action Concepts: How "Plastic" is Grounded Cognition?***
J. C. Mizelle and Lewis A. Wheaton

frontiers in
PSYCHOLOGY

EDITORIAL
published: 17 August 2011
doi: 10.3389/fpsyg.2011.00187

Introduction to the special topic Embodied and Grounded Cognition

Anna M. Borghi[1]*† and Diane Pecher[2]*†

[1] University of Bologna and National Research Council, Rome, Italy
[2] Psychology Department, Erasmus University Rotterdam, Rotterdam, Netherlands
*Correspondence: anna.borghi@gmail.com; pecher@fsw.eur.nl
†Anna M. Borghi and Diane Pecher have contributed equally to this work.

INTRODUCTION

In the last 10–15 years, the embodied and grounded (E and G) cognition approach has become widespread in all fields related to cognitive (neuro) science, and a lot of evidence has been collected. The approach proposes that cognitive activity is grounded in sensory–motor processes and situated in specific contexts and situations.

This special topic had two aims: first, give an idea of the field in its broadness. Second, focus on some challenges for E and G theories. The first important challenge is to account for understanding abstract concepts and words. Evidence on the representation of concrete concepts is compelling, whereas evidence on abstract concepts is still scarce and limited to restricted domains. A second important challenge concerns the role of computational models. E and G theories of cognition need to formulate more precise hypotheses, and models help to constrain and specify in more detail the predictions and the claims advanced.

THE FIELD IN ITS BROADNESS

Although the importance of sensory–motor grounding had already become apparent in philosophy and linguistics, only after a couple of influential theoretical papers in the late 1990s did cognitive psychology get involved seriously (Glenberg, 1997; Zwaan and Radvansky, 1998; Barsalou, 1999; Pulvermüller, 1999). The idea that cognitive processes, such as those involved in language and memory, are grounded in the same systems as those used for perception and action has received much empirical support. This special topic presents a sample of the new and exciting empirical work in this field.

Grounding of language comprehension has been among the most compelling demonstrations. Several papers further investigated the grounded or embodied nature of a variety of linguistic issues. Wellsby et al. (2010) showed that people recognize embodied insults faster than less embodied insults. Hald et al. (2011) and Collins et al. (2011) using ERPs, both showed that the modality switch effect reflects modality differences in early meaning activation due to different sensory systems involved in the mental simulation. In an fMRI study, Rueschemeyer et al. (2010) showed that when people process motion language they simulate seeing motion and preparing for the actions that the situation requires. Thus, mental simulations are immediately affected by the sentence context. A similar conclusion is drawn by van Dam et al. (2010) who showed that the immediate linguistic context modulates whether object names activate action simulations. In contrast, Hemeren and Thill (2011) found that visual perception of functional object use is little affected by knowledge of the object identity. Instead, participants used kinematic information such as velocity, acceleration, and changes in direction to segment the action. Marsh and Glenberg (2010) propose that people learn grammar by imitation, in particular neuromuscular tuning. Lynott and Connell (2010) propose a grounded model (embodied conceptual combination, ECCo) for new combinations of familiar concepts such as *elephant complaint*. They argue that people mesh motor and perceptual affordances either in a destructive or non-destructive manner.

How do we understand other people? The social Simon effect supports the idea that we understand other people's actions by representing them as our own. Dolk et al. (2011) and Vlainic et al. (2010) found that online information from another person is not necessary, suggesting that the effect is not due to co-representation of the other's action but rather to awareness of a second location that highlights the congruency between stimulus location and response. Kessler and Rutherford (2010) showed that whether representations of other people's perspective are embodied depends on difficulty. They argue that the degree of symmetry determines whether perspective taking is needed.

THE CHALLENGES

THE CHALLENGE TO ACCOUNT FOR ABSTRACT CONCEPTS

Most studies of E and G cognition have focused on concrete objects and actions. People can also represent and reason about abstract concepts that do not have many sensory–motor features, however, and there is not much evidence yet on grounding of abstract concepts in sensory–motor systems (for a review, see Pecher et al., 2011). Thus, critics have argued that current evidence does not fully account for abstract representation.

Some authors propose a theoretical solution. van Elk et al. (2010) challenged the reliance on representations and proposed an *enactivist* approach. They argued that the view of representations as simulations or re-enactment of previous experiences opens two problems. First, the necessity of sensory–motor systems for cognition has been disputed (e.g., Mahon and Caramazza, 2008). Second, it fails to explain concepts beyond our motor repertoire, such as animal actions, or abstract words. The authors proposed that sensory–motor brain areas underlie prediction of actions, arguing in favor of a more procedural view of cognition. In contrast, Dove believed that "the notion of representation is too useful to give up." Dove proposed to use the term *dis-embodiment*. Language is dis-embodied because its sensory–motor features are unrelated to its meaning. According to Dove (2011), this dual functionality of language is at the basis of generalization and abstraction.

Several papers addressed conceptual metaphor theory. Flusberg et al. (2010) presented a computational connectionist model

of grounding time in the representation of space. Kranjec and Chatterjee (2010) challenged the view that time is grounded in space and investigated grounding of temporal concepts directly in neural areas dedicated to time perception. Beside conceptual metaphor theory, new views emerge that argue for multiple representations and assign a specific role to language (e.g., Barsalou et al., 2008). Views differ, however, on the role of symbolic associations. Whereas Dove (2011) proposes that symbolic processing complements sensory–motor processing, others have argued that symbolic associations merely provide shortcuts for more efficient task performance (Barsalou et al., 2008). Pecher and Boot (2011) found that spatial attention was affected by number magnitude, but, contrary to what conceptual metaphor theory would predict, only for numbers in concrete contexts. Associative shortcuts might suffice for abstract magnitudes, whereas concrete magnitudes require deeper processing. Borghi et al. (2011) investigated how the kind of acquisition influenced representation of concrete and abstract novel words. Verbal information is more crucial to represent abstract concepts, manual information to represent concrete concepts. This is in line with the proposal that Words are social Tools (Borghi and Cimatti, 2009): labels and explanations are particularly helpful to learn abstract words, as they provide the glue helping us to keep together multifaceted experiences.

Barbey and Patterson (2011) addressed abstraction from a cognitive neuroscience perspective. They reviewed evidence that PFC extracts statistical regularities across experiences; these regularities are the basis for building abstract rules, such as causal beliefs. The underlying mechanism is given by a distributed neural network across both modality specific and associative areas. This mechanism can be activated also when stimuli are not present, thus causing a causal simulation of an event.

THE CHALLENGE OF COMPUTATIONAL MODELS

The need for computational models emerges clearly. Pezzulo et al. (2011) argued that embodied robotics is the platform to test embodied computational models independent from the kind of modeling style – dynamic models, Bayesian approaches, or connectionist models. Robots provide cognitive process models with constraints because their sensory–motor potential and limitations can be made similar to those of humans.

Instead of a system that has a full representation of the environment, Rothkopf and Ballard (2010) proposed a system of simple visuomotor modules, learned through reinforcement, allowing rapid access to behavioral primitives. The authors illustrate a credit assignment algorithm that solves the problem of calibrating different visuomotor modules while pursuing multiple goals. In contrast, Ursino et al. (2010) presented a two-store model of representation: a semantic store of topologically organized sensory–motor features and a lexicon. Because attractors are not steady states but rather synchronized oscillations multiple objects can be represented at the same time. A time dependent Hebbian learning rule allows the model to learn the relationship between words and object features.

Parisi (2011) argued that current theories are limited because they ignore emotions, "the other half of the embodied mind." He argued that robots endowed with emotional circuits reach higher level of performance and have better survival probability compared to robots without emotion. Mizelle and Wheaton (2010) criticized current theories for ignoring flexibility. Their modular MSAG model represents high plasticity and variability in tool selection and use.

CONCLUSION

We believe this special topic has opened new perspectives, and gives us some indications of where the field of E and G cognition is going. First of all, it testifies that the domain has expanded rapidly. At the same time, however, researchers will have to solve some open issues, such as the questions of necessity and abstraction. This special topic presents many empirical, theoretical, and modeling approaches to these challenges. It is a proof of how lively and open the field is.

REFERENCES

Barbey, A., and Patterson, R. (2011). Architecture of explanatory inference in the human prefrontal cortex. *Front. Psychol.* 2:162. doi: 10.3389/fpsyg.2011.00162

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford: Oxford University Press), 245–283.

Borghi, A. M., and Cimatti, F. (2009). "Words as tools and the problem of abstract words meanings," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 2304–2309.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., and Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Front. Psychol.* 2:15. doi: 10.3389/fpsyg.2011.00015

Collins, J., Pecher, D., Zeelenberg, R., and Coulson, S. (2011). Modality switching in a property verification task: an ERP study of what happens when candles flicker after high heels click. *Front. Psychol.* 2:10. doi: 10.3389/fpsyg.2011.00010

Dolk, T., Hommel, B., Colzato, L. S., Schutz-Bosbach, S., Prinz, W., and Liepelt, R. (2011). How 'social' is the social Simon effect? *Front. Psychol.* 2:84. doi: 10.3389/fpsyg.2011.00084

Dove, G. (2011). On the need for embodied and dis-embodied cognition. *Front. Psychol.* 1:242. doi: 10.3389/fpsyg.2010.00242

Flusberg, S. J., Thibodeau, P. H., Sternberg, D. A., and Glick, J. J. (2010). A connectionist approach to embodied conceptual metaphor. *Front. Psychol.* 1:197. doi: 10.3389/fpsyg.2010.00197

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Hald, L. A., Marshall, J., Janssen, D. P., and Alan, G. (2011). Switching modalities in a sentence verification task: ERP evidence for embodied language processing. *Front. Psychol.* 2:45. doi: 10.3389/fpsyg.2011.00045

Hemeren, P. E., and Thill, S. (2011). Deriving motor primitives through action segmentation. *Front. Psychol.* 1:243. doi: 10.3389/fpsyg.2010.00243

Kessler, K., and Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* 1:213. doi: 10.3389/fpsyg.2010.00213

Kranjec, A., and Chatterjee, A. (2010). Are temporal concepts embodied? A challenge for cognitive neuroscience. *Front. Psychol.* 1:240. doi: 10.3389/fpsyg.2010.00240

Lynott, D., and Connell, L. (2010). Embodied conceptual combination. *Front. Psychol.* 1:212. doi: 10.3389/fpsyg.2010.00212

Mahon, B., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Marsh, E. R., and Glenberg, A. M. (2010). The embodied statistician. *Front. Psychol.* 1:184. doi: 10.3389/fpsyg.2010.00184

Mizelle, J. C., and Wheaton, L. A. (2010). The neuroscience of storing and molding tool action concepts: How "plastic" is grounded cognition? *Front. Psychol.* 1:195. doi: 10.3389/fpsyg.2010.00195

Parisi, D. (2011). The other half of the embodied mind. *Front. Psychol.* 2:69. doi: 10.3389/fpsyg.2011.00069

Pecher, D., and Boot, I. (2011). Numbers in space: differences between concrete and abstract situations. *Front. Psychol.* 2:121. doi: 10.3389/fpsyg.2011.00121

Pecher, D., Boot, I., and van Dantzig, S. (2011). "Abstract concepts: sensory-motor grounding, metaphors, and beyond," in *The Psychology of Learning and Motivation*, Vol. 54, ed. B. Ross (Burlington: Academic Press), 217–248.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: a dialogue on embodiment and com-

putational modeling. *Front. Psychol.* 2:5. doi: 10.3389/fpsyg.2011.00005

Pulvermüller, F. (1999). Words in the brain's language. *Behav. Brain Sci. 22*, 253–336.

Rothkopf, C. A., and Ballard, D. H. (2010). Credit assignment in multiple goal embodied visuomotor behavior. *Front. Psychol*. 1:173. doi: 10.3389/fpsyg.2010.00173

Rueschemeyer, S., Glenberg, A. M., Kaschak, M. P., Mueller, K., and Friederici, A. D. (2010). Top-down and bottom-up contributions to understanding sentences describing objects in motion. *Front. Psychol.* 1:183. doi: 10.3389/fpsyg.2010.00183

Ursino, M., Cuppini, C., and Magosso, E. (2010). A computational model of the lexical-semantic system, based on a grounded cognition approach. *Front. Psychol.* 1:221. doi: 10.3389/fpsyg.2010.00221

van Dam, W. O., Rueschemeyer, S.-A., Lindemann, O., and Bekkering, H. (2010). Context effects in embodied lexical-semantic processing. *Front. Psychol.* 1:150. doi: 10.3389/fpsyg.2010.00150

van Elk, M., Slors, M., and Bekkering, H. (2010). Embodied language understanding requires an enactivist paradigm of cognition. *Front. Psychol.* 1:234. doi: 10.3389/fpsyg.2010.00234

Vlainic, E., Liepelt, R., Colzato, L. S., Prinz, W., and Hommel, B. (2010). The virtual co-actor: the social Simon effect does not rely on online feedback from the other. *Front. Psychol.* 1:208. doi: 10.3389/fpsyg.2010.00208

Wellsby, M., Siakaluk, P. D., Pexman, P. M., and Owen, W. J. (2010). Some insults are easier to detect: the embodied insult detection effect. *Front. Psychol.* 1:198. doi: 10.3389/fpsyg.2010.00198

Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185.

Received: 29 June 2011; accepted: 22 July 2011; published online: 17 August 2011.

*Citation: Borghi AM and Pecher D (2011) Introduction to the special topic Embodied and Grounded Cognition. Front. Psychology **2**:187. doi: 10.3389/fpsyg.2011.00187*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Borghi and Pecher. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 16 November 2010
doi: 10.3389/fpsyg.2010.00198

Some insults are easier to detect: the embodied insult detection effect

Michele Wellsby[1], Paul D. Siakaluk[1]*, Penny M. Pexman[2] and William J. Owen[1]

[1] *Psychology Department, University of Northern British Columbia, Prince George, BC, Canada*
[2] *Department of Psychology, University of Calgary, Calgary, AB, Canada*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Michael Kaschak, Florida State University, USA
Shirley-Ann Rueschemeyer, Radboud University Nijmegen, Netherlands

***Correspondence:**
Paul D. Siakaluk, Psychology Department, University of Northern British Columbia, Prince George, BC, Canada V2N 4Z9.
e-mail: siakaluk@unbc.ca

In the present research we examined the effects of bodily experience on processing of insults in a series of semantic categorization tasks we call insult detection tasks (i.e., participants decided whether presented stimuli were insults or not). Two types of insults were used: more embodied insults (e.g., *asswipe*, *ugly*), and less embodied insults (e.g., *cheapskate*, *twit*), as well as non-insults. In Experiments 1 and 2 the non-insults did not form a single, coherent category (e.g., *airbase*, *polka*), whereas in Experiment 3 all the non-insults were compliments (e.g., *eyeful*, *honest*). Regardless of type of non-insult used, we observed facilitatory embodied insult effects such that more embodied insults were responded to faster and recalled more often than less embodied insults. In Experiment 4 we used a larger set of insults as stimuli, which allowed hierarchical multiple regression analyses. These analyses revealed that bodily experience ratings accounted for a significant amount of unique response latency, response error, and recall variability for responses to insults, even with several other predictor variables (e.g., frequency, offensiveness, imageability) included in the analyses: responses were faster and more accurate, and there was greater recall for relatively more embodied insults. These results demonstrate that conceptual knowledge of insults is grounded in knowledge gained through bodily experience.

Keywords: embodied cognition, insult processing, conceptual processing, mental simulation

The emerging viewpoint of embodied cognition holds that cognitive processes are deeply rooted in bodily interactions with the environment (Wilson, 2002). That is, bodily interactions with the environment are integral to the acquisition of knowledge and to the development of cognitive processes that bear on that knowledge (Barsalou, 1999). The embodied cognition viewpoint contrasts with the long held classical cognitive viewpoint (known as *cognitivism*), which claims that cognitive processes are not related to bodily interactions with the environment; rather, cognitive processes are proposed to be non-embodied in the sense that they are independent from knowledge gained from bodily experience. As such, by the cognitivist account, cognitive processing involves the manipulation of abstract symbols via rules in the mind (Barsalou, 1999; Cowart, 2004).

Barsalou (1999) developed an embodied cognition account called perceptual symbol systems theory. According to this theory, bodily interactions with the environment are crucial to many cognitive processes. He suggested that the modality-specific neural systems used for perception and action are also used to represent concepts in the brain, through the process of simulation. In other words, conceptual processing is, in a fundamental way, grounded in neural systems involved in perceptual and motor processing. More specifically, simulation is the partial neural re-enactment of perceptual, motor, and introspective states acquired during bodily experiences with the environment (Barsalou, 2008). For example, when we have bodily experience interacting with a particular car, the brain captures states across all sensory, motor, and introspective modalities (e.g., what the car looks and smells like, the bodily "feel" of getting into and driving the car, the fear that may result from nearly getting into an accident, etc.) and integrates them into a multimodal neural representation stored in memory. Later, when knowledge is needed to represent the category *CAR*, these multimodal neural representations are partially reactivated to simulate how the brain represented the original experiences. In other words, off-line cognition (i.e., cognizing about an object or event that is not currently present) is body-based (Wilson, 2002): When we think about an object or event, we are neurally re-experiencing the sensory, motor, and introspective components associated with that object or event from previous bodily experience.

The notion of simulation has been invoked to explain recent demonstrations that knowledge gained through bodily experience influences linguistic processing. The literature on sentence processing contains several such examples, including the action-sentence compatibility effect (Glenberg and Kaschak, 2002; Glenberg et al., 2008) and what we will call the *visibility-sentence compatibility effect* (Yaxley and Zwaan, 2007).

Using a sentence verification task (i.e., does the sentence make sense?), Glenberg et al. (2008) reported an interaction between type of sentence (whether it referred to something being transferred toward or away from one's body) and type of response (whether it required moving one's hand toward or away from one's body), such that responses to toward sentences were faster when they required a toward response as compared to an away response, and responses to away sentences were faster when they required an away response as compared to a toward response. Interestingly, this action-sentence compatibility effect was observed for sentences that described

transfer of either concrete objects (e.g., papers) or abstract concepts (e.g., responsibilities). Glenberg et al. explained their findings in the following way. When participants read a sentence that involved something being transferred away from them, for example, they created a simulation of either the concrete object or abstract concept going away from them. They were faster at making a button press that required moving their hand away from their bodies because the away action was congruent with the direction of transfer in the simulation. Conversely, in the away condition, participants were slower at making a button press that required moving their hand toward their bodies, because this toward action was incongruent with the direction of transfer in the simulation.

Using a sentence–picture verification task (i.e., participants read a sentence and then verified if a pictured object was in that sentence), Yaxley and Zwaan (2007) reported an interaction between type of visibility described in a sentence (clear or unclear) and type of picture resolution (clear or unclear). An example of a clear sentence is "Through the clean goggles, the skier could easily identify the moose," and an example of an unclear sentence is "Through the fogged goggles, the skier could hardly identify the moose." The clear pictures were presented at 100% resolution, whereas the unclear pictures were presented at 50% resolution (pictures in this condition resembled a snowy picture on television). Yaxley and Zwaan reported that when participants read a clear sentence, clear pictures were verified faster than unclear pictures. Remarkably, when participants read an unclear sentence, unclear pictures were verified faster than clear pictures. Yaxley and Zwaan proposed that when participants read the sentences, the sentences were understood by creating simulations of implied perceptual visibility; thus, when reading clear sentences, participants simulated an unobstructed viewpoint, and could, for example, simulate clearly seeing a moose. When the participants then saw a clear picture of a moose, this matched the simulation they had created, and response latencies were faster as compared to the unclearly presented picture of the moose, which was incongruent with the simulation. Alternatively, when participants read unclear sentences, they simulated an obstructed view, where they could barely make out a moose. When they then saw an unclear picture of a moose, this matched the simulation they had created, and responses were faster to the unclear picture as compared to the clear picture.

A second area of linguistic processing where embodied knowledge is demonstrated to have an influence is in the understanding of metaphors. Neisser (2003) stated that metaphoric thought is an act of imagination, which is mediated by human embodiment. Lakoff and Johnson (1999) also proposed that abstract concepts are grounded metaphorically in embodied knowledge. They argued that people possess extensive knowledge about their bodies and environments, and that abstract concepts draw on this knowledge metaphorically. One example Lakoff and Johnson used to explain the role that bodily experience plays in the comprehension of metaphors is the metaphor *BAD IS STINKY*. When someone says, "That book stinks!," we understand that she is stating her belief that the content of the book is bad, and not referring to the actual smell of the book. Our understanding of this metaphor, according to Lakoff and Johnson, is mediated by the sensory domain of smell and our having the primary experience of being repelled by foul-smelling objects. We can use this knowledge to make an evaluation about something abstract, such as the content of a book.

Wilson and Gibbs (2007) recently demonstrated how bodily knowledge facilitates conceptual processing of abstract metaphors. They examined the role of simulation in the understanding of abstract metaphors for which the associated expressions are impossible to physically perform, such as *swallow your pride* or *push the argument*. Similar to Lakoff and Johnson (1999), they reasoned that simulations for abstract metaphors are based on sensorimotor knowledge gained through physically interacting with objects, such as *swallow your food* and *push the carriage*. They reported that abstract metaphors were understood faster when participants made or imagined making a congruent body movement just before reading a metaphorical phrase (e.g., making or imagining making a pushing movement just before reading *push the argument*), than when they either made or imagined making an incongruent body movement (e.g., making or imagining making a swallowing movement just before reading *push the argument*), or did not make or imagine making any body movement (i.e., they simply read the metaphorical phrase).

A third area of linguistic processing in which effects of embodied knowledge have been demonstrated is word recognition. One example is the object manipulation effect reported by Myung et al. (2006). Myung et al. used an auditory lexical decision task in which they presented primes that were either related or unrelated to targets, where relatedness was defined by overlap in manipulation features. For example, the prime *typewriter* would be related to the target *piano* because they share common manipulation features (i.e., using both hands, with fingers in a curled position that press downward), whereas the prime *typewriter* would not be related to the target *screwdriver* because they have different manipulation features (unlike *typewriter*, *screwdriver* has the manipulation features of using one hand, with fingers in a grasping position, and twisting of the wrist). Myung et al. reported that responses were faster when targets shared manipulation features with their primes (e.g., *typewriter–piano* versus *typewriter–screwdriver*), and proposed the following explanation for their results. When participants heard the prime, they simulated the type of physical manipulation associated with the object. For example, when participants heard the prime *typewriter*, they simulated how a human would physically manipulate a typewriter (i.e., fingers curled and pressing downward). When participants then had to decide whether *piano* was a real word or not, they were quick to decide that the required response was "yes," because the simulation they associated with *typewriter* was similar to the simulation they associated with *piano*. However, when participants heard the prime word *typewriter* and then had to decide whether *screwdriver* was a real word or not, the responses were slower, because the simulation of manipulating a typewriter differs from the simulation of manipulating a screwdriver.

A second example of embodied knowledge influencing word recognition is the body–object interaction (BOI) effect. BOI is a variable that assesses the ease with which a human body can physically interact with a word's referent. Facilitatory BOI effects (i.e., faster responding to words rated high in BOI, such as *mask*, than to words rated low in BOI, such as *ship*) have been reported in visual lexical decision and phonological lexical decision tasks (Siakaluk et al., 2008a; Tillotson et al., 2008) and in semantic categorization tasks (Siakaluk et al., 2008b; Wellsby et al., in press). The BOI effect

is consistent with the notion that words rated high in BOI elicit richer motor simulations of how human bodies physically interact with their referents, and that these richer motor simulations allow for more efficient responding to the words that refer to them. Taken together, the BOI effect and the object manipulation effect suggest that knowledge based on bodily experience is an important part of the semantic representations of words.

In the word recognition research reviewed above (i.e., research examining object manipulation and BOI effects), the stimuli of interest were all concrete concepts; that is, the words used referred to concrete entities like *piano*, *typewriter*, *mask*, and *ship*. There has been much less research on the role of bodily experience in the processing of words referring to more abstract concepts; that is, words for which the referents or meanings are less tangible. Indeed, one aim of this special issue was to address processing of abstract concepts. In the present research we approached this subject by examining processing of a particular kind of abstract meaning; that is, the negative inferences conveyed by insults.

An insult is a verbal expression that conveys a negative (e.g., offensive, degrading) meaning. Many insults, such as *stupid* and *ugly*, are abstract in the sense that they express a quality or attribute distinct from any particular person. That is, they qualify people in a negative manner (e.g., *she is stupid* or *he is ugly*). This way of conceptualizing the idea of abstractness as separable from any specific object or instance is what is often meant when cognitive scientists talk about something being abstract. However, there are other ways of conceptualizing the abstractness of insults. One such way is the use of insults, such as *lunatic* and *asswipe*, in a metaphorical manner. That is, knowledge gained from one context (e.g., general semantic knowledge, bodily experience) is applied to a different context (e.g., a person's intellectual ability or moral character) so that certain negative resemblances are inferred. Importantly, however, knowledge used from the original context cannot be *literally* true in the inferred context. For the insult *lunatic*, general semantic knowledge that a lunatic is someone of unsound mind (as perhaps judged from a psychiatric or legal viewpoint) can be used to infer that someone who is not really of unsound mind (from a psychiatric or legal viewpoint) nonetheless possesses certain, say, intellectual qualities or attributes that are less than desirable. For the insult *asswipe*, knowledge gained from certain bodily experiences can be used to infer that someone, who cannot literally be an asswipe, nonetheless possesses certain, say, moral qualities or attributes that are (to put it mildly) less than desirable. We propose that this second way of conceptualizing abstractness is highly relevant for the present study.

Thus, to be as clear as possible, we are proposing a conceptualization of abstractness in a perhaps less traditional, but nonetheless, we believe, theoretically interesting and legitimate manner. Our conceptualization of abstractness is in fact very similar to that of Lakoff and Johnson (1999) and Wilson and Gibbs (2007), described above. That is, knowledge gained from one context may be abstracted and used in another context, in which the knowledge is applied metaphorically because it cannot be applied literally. More pertinently to the question of embodiment, the negative inferences for many insults, such as *asswipe*, will be derived from a bodily (i.e., sensorimotor) context and applied to a different context (e.g., making implications about someone's intelligence or moral character). The research question of interest for the present study is therefore whether this kind of knowledge gained from bodily experience influences semantic processing of insults.

We propose that if the framework of perceptual symbol systems theory were applied to processing insults, two important assumptions can be made. The first assumption is that sensory, motor, kinesthetic, and proprioceptive knowledge gained from bodily experience may be used to understand the implied meaning of some insults (i.e., more embodied insults, such as *numbskull*) more so than other insults (i.e., less embodied insults, such as *idiot*). For example, when someone is called a *numbskull*, comprehension of the implied meaning may comprise simulating prior experiences of numbness (e.g., in one's arms or legs), despite the fact that there is no implication of numbness literally occurring in the person's skull at the time the insult is given. Thus, for more embodied insults like *numbskull*, bodily knowledge (in this case of knowing what it is like to feel numbness and any resulting deficits in functionality of the associated body part) is likely to be a very salient component of their simulations. We should also emphasize here that although the majority of the more embodied insults used in our Experiments 1–3 below have body part components (as *numbskull* does), this does not imply that more embodied insults *must* have a body part component. We propose that more embodied insults, such as *ugly* (e.g., one may experience ugliness because of having a scrape mark on one's face) and *weak* (e.g., one may experience weakness because of not being able to lift a heavy object), while not including mention of body parts, also elicit greater bodily knowledge in their simulations than less embodied insults. All of these insults, whether they include mention of body parts or not, are more embodied because participants rate bodily experience as important to understanding their meanings.[1]

The second assumption we make is that emotional and introspective systems contain knowledge that is highly relevant to processing insults, and that these forms of knowledge are therefore involved in the simulation process. That is, although when someone is called an *idiot*, it is difficult to imagine what type of bodily knowledge may be simulated that would aid in comprehension, it is likely the case that this would cause negative emotional reactions (e.g., shame, anger, embarrassment), similar to when someone is called a *numbskull*. In addition, contextual considerations (e.g., being called an *idiot* or a *numbskull* immediately after committing a faux pas) would activate knowledge in introspective systems that would contribute to understanding that an insult has been given. In summary, our proposal is that: (a) simulating knowledge gained through emotional and introspective systems is likely to be just as important to comprehending the intentions behind more embodied insults and less embodied insults, but (b) simulating knowledge gained through sensory, motor, kinesthetic, and proprioceptive systems is likely to be much more important to comprehending the intentions behind more embodied insults than less embodied insults.

Siakaluk et al. (in press) examined the influence of knowledge gained through bodily experience on processing of insults in a variant of the emotional Stroop task. Previous research has

[1]Note that we expanded the item set in Experiment 4 to include many more embodied insults that did not mention a body part, and the same pattern of results was observed as in Experiments 1–3.

found that response latencies to name the font color of emotionally laden words (e.g., taboo words such as *rape*) are slower than response latencies to name the font color of non-emotional words (e.g., *run*). It has been proposed that this emotional Stroop effect occurs because reactions (e.g., embarrassment, shock) to emotionally laden words engage processing resources needed for color naming (MacKay et al., 2004). In their variant of the emotional Stroop task, Siakaluk et al. (in press) used more embodied insults (e.g., *asswipe*), less embodied insults (e.g., *cheapskate*), and non-insults (e.g., *hardwood*) as stimuli. They first collected bodily experience ratings for their insult stimuli, which measured the extent to which the meaning of each insult referenced or alluded to knowledge gained through human bodily experience (see below for a more detailed explanation). In other words, the bodily experience ratings gauged how important participants thought having a body was to understanding the meanings of the insults. Siakaluk et al. (in press) then conducted multiple regression analyses on color naming latencies and reported two intriguing findings. First, there was an insult Stroop effect, such that color naming latencies were slower to the insults than to the non-insults. Second, the bodily experience ratings accounted for a significant amount of unique color naming latency variability for the insults, even after the variability attributed to other predictors (e.g., print frequency, offensiveness, number of meanings) was statistically removed. Moreover, the effect of bodily experience was inhibitory such that color naming latencies were slower for the more embodied insults than for the less embodied insults.

Siakaluk et al. (in press) proposed that the inhibitory embodied insult Stroop effect can be accommodated by Barsalou's (1999) perceptual symbol systems theory. That is, when more embodied insults are presented in the Stroop task, they activate richer simulations of knowledge gained through sensory, motor, kinesthetic, and proprioceptive systems than do less embodied insults. In other words, when participants view a more embodied insult, they simulate the bodily knowledge associated with the insult (e.g., when viewing the insult *numbskull*, participants simulate the knowledge of what it feels like when a body part is numb and any resulting deficits in functionality of the associated body part). These richer sensorimotor simulations that are activated by more embodied insults lead to more efficient linguistic processing, which ultimately leads to greater inhibition of color naming responses (see also Cohen et al., 1990).

THE PRESENT RESEARCH

In the present research we examined the influence of bodily knowledge in the processing of insults but in a much more direct way than in Siakaluk et al. (in press). That is, Siakaluk et al.'s Stroop study showed longer color naming latencies for relatively more embodied insults, and it was assumed that this effect arose because, for relatively more embodied insults, the insult meanings were harder to ignore (thus inhibiting color naming responding). Since the Stroop task was used in that study, however, one cannot actually be sure that participants were processing the insult meanings of the stimuli. Thus, one could assert that it is not clear that the "embodied" effects observed in Siakaluk et al. (in press) were generated in the process of deriving *insult meanings*. The present study addresses this issue more directly by examining the effects of bodily experience on processing of insults in a task we call the insult detection task.

There is much research demonstrating that richer semantic representations elicit faster responding in semantic categorization tasks (e.g., Pexman et al., 2008; Siakaluk et al., 2003, 2008b), of which the insult detection task is a special case (because it involves categorizing stimuli as belonging or not belonging to the decision category "insult"). It has been argued that richer semantic representations afford faster settling in the semantic units (e.g., Pexman et al., 2008) and, hence, faster semantic decision latencies. We therefore predicted a *facilitatory* effect for more embodied insults in the insult detection task, such that response latencies will be *faster* to more embodied insults than to less embodied insults. If this prediction is borne out, then it will provide support for Siakaluk et al.'s (in press) conclusion that it was indeed the activation of the meanings of the insults that inhibited Stroop task performance in their study. In the present study, richer sensorimotor simulations should provide more evidence that an item is a positive instance of the category "insult": thus, *faster* response latencies should be associated with more embodied insults. In contrast, in Siakaluk et al.'s (in press) Stroop task, task demands required that the participants pronounce the font color, and richer sensorimotor simulations should hinder participants' ability to pronounce the font color; thus, *slower* response latencies should be associated with more embodied insults, which was what was observed.

To ensure the best possible opportunity to observe facilitatory effects of bodily experience in the insult detection task, we used a go/no-go procedure (in which participants respond only to the insult stimuli), rather than a yes/no procedure (in which participants respond to both the insult and non-insult stimuli). We did this because semantic effects in semantic categorization tasks have been most robust using the go/no-go procedure, most likely because this procedure encourages more extensive processing (see, e.g., Siakaluk et al., 2003, 2007). In addition, participants completed a surprise recall task immediately after they finished the insult detection task. We included the surprise recall task in the present study in order to test an additional research question; that is, whether effects of bodily experience on insult processing are very fleeting (in which case they should not influence subsequent memory for the stimuli) or longer lasting (in which case they should influence subsequent memory for the stimuli). We predicted a *facilitatory* effect for more embodied insults in the surprise recall task, such that participants should recall a greater percentage of more embodied insults than less embodied insults. We made this prediction because the more detailed simulations that should be elicited for the more embodied insults should afford more elaborative processing at encoding and hence facilitate recall during the surprise recall testing phase.

EXPERIMENT 1

METHOD

Participants

Thirty-three undergraduate students from the University of Northern British Columbia (UNBC) participated in the experiment for bonus course credit. All were native English speakers and reported normal or corrected-to-normal vision. None of these individuals participated in any of the other experiments.

Stimuli

Siakaluk et al. (in press) had a group of 40 UNBC undergraduate students rate 178 insults on a 1–7 Likert scale regarding how important they thought bodily experience would be to understanding the insult. The following instructions were provided to the participants to read while the researcher read them aloud:

> Please read each insult carefully and decide to what extent the meaning of the insult references or alludes to knowledge gained through human bodily experience and interactions. That is, please rate how important you think having a body contributes to understanding what each insult means.

After reading the instructions, the participants were given two examples to help them understand how to rate the insults. The researcher explained that the insult *bossy* would likely be given a relatively low rating because it is not clear how bodily experience would contribute much to its meaning, whereas the insult *numbskull* would likely be given a relatively high rating because experience with numbness in our bodies may help contribute to its intended meaning.

Thirty-six insults were selected from Siakaluk et al. (in press) for use in the present experiment: 18 of the insults (e.g., *asswipe*) were rated as being high in bodily experience (henceforth referred to as more embodied insults) and the other 18 insults (e.g., *cheapskate*) were rated as being low in bodily experience (henceforth referred to as less embodied insults). The insults consisted of either one or two words. The two groups of insults were matched on length (i.e., number of letters), print frequency (using HAL log-frequency values from the English Lexicon Project database; Balota et al., 2007), rated offensiveness (on a 1–7 scale), rated frequency of usage (on a 1–7 scale), rated number of meanings (on a 0–2 scale), and rated imageability (on a 1–7 scale) (all $ps > 0.35$). The descriptive statistics for the insults are presented in **Table 1**. Thirty-six non-insults (e.g., *airbase*) were selected from Siakaluk et al. (in press) to be used as control items. The non-insults also consisted of either one or two words, and were matched to the insults on printed length. The experimental stimuli are listed in the Appendix.

Apparatus and procedure

The stimulus items were presented one at a time on a computer display driven by a Pentium-class microcomputer running DirectRT software. The participants' task was to determine whether the stimulus items were insults or not. Before starting the insult detection task, the participants were primed with instructions indicating that when each stimulus item was presented they should imagine themselves in a confrontational situation where someone says to them, "You are _____" or "You are a(n) _____." A trial was initiated by a fixation marker that appeared at the center of the computer display. The fixation marker was presented for 1 s, and was then replaced by a stimulus item. Participants were asked to respond only to the stimulus items that were insults, and to respond as quickly and accurately as possible by pressing the "?" key on the computer keyboard. If participants judged a stimulus item to not be an insult, they were asked not to make any key press response. If no response was made, stimulus items remained on the computer display for 2.5 s, and were then removed and replaced by the fixation marker. The 72 stimuli were presented in a different random order to each participant.

Before beginning the experiment, each participant first completed 20 practice trials. The practice trials consisted of 10 insult stimuli (five more embodied insults and five less embodied insults) and 10 non-insult stimuli. All practice stimuli were similar in normative print frequency to the experimental items.

After participants completed the insult detection task, they completed a surprise recall task. Participants were given 3 min to recall as many stimulus items as they could remember from the insult detection task. Participants typed their responses into an Excel spreadsheet. Responses were coded as correct even if they were not spelled correctly (e.g., *asswhole* or *wanna be* were coded as correct for the insults *asshole* and *wannabe*, respectively). Synonyms, however, were not coded as correct.

RESULTS AND DISCUSSION

For all the experiments reported in the present study the following procedure for removal of outliers was used. For each participant, response latencies greater than or less than 2.5 SD from the cell mean of each condition were treated as outliers. Across participants, this resulted in the removal of a total of 38 observations (3.20% of the data) from the data set in Experiment 1. Response errors for the insult stimuli comprised only 2.35% of trials, and, as such, the response error data were not analyzed. The mean response latencies of correct responses, mean error percentages, and correct recall percentages for all stimulus types are presented in **Table 2**. For the first three experiments reported in the present study bodily experience was a within-subject manipulation, and unless otherwise indicated, all effects are significant at $p < 0.05$.

There was a significant effect of bodily experience in the response latency data, $t(32) = 2.68$, SEM = 10.65, partial $\eta^2 = 0.18$, and in the surprise recall data, $t(32) = 2.54$, SEM = 2.39, partial $\eta^2 = 0.17$. In the response latency data, responses to the more embodied insults were an average of 28 ms faster than responses to the less embodied insults. In the surprise recall data, the percentage of items correctly recalled was 6.06% higher for more embodied insults than for less embodied insults.

Table 1 | Mean characteristics and standard errors (in parentheses) for the insult stimuli used in Experiments 1–3.

Insult type	Bodily exp	Length	Print freq	Offensive	Usage freq	NoM	Image
More embodied	5.2 (0.14)	7.1 (0.44)	5.8 (0.96)	3.2 (0.18)	3.5 (0.29)	1.2 (0.03)	3.5 (0.11)
Less embodied	2.5 (0.11)	6.5 (0.44)	6.1 (0.86)	3.2 (0.19)	3.6 (0.22)	1.2 (0.03)	3.3 (0.16)

Note. Bodily exp, bodily experience rating; Length, number of letters; Print freq, print frequency using HAL log-frequency values; Offensive, rating of offensiveness; Usage freq, rated frequency of usage; NoM, rated number of meanings; Image, imageability.

These results show faster response latencies and more accurate recall for more embodied insults as compared to less embodied insults and thus provide support for the idea that knowledge gained through bodily experience facilitates responding in a task (i.e., the insult detection task) in which richer sensorimotor simulations should be useful. Moreover, the response latency data are consistent with the facilitatory BOI effect on response latencies in semantic categorization (Siakaluk et al., 2008b; Wellsby et al., in press). The facilitatory effects of bodily experience on the surprise recall data are intriguing, as they demonstrate that bodily experience effects persist after the insult detection task is finished and are also influential in memory for insults.

One potential problem with Experiment 1 is the fact that 12 of the more embodied insults mentioned a body part, whereas only four of the non-insults mentioned a body part (if one includes the non-insult *denture*). It is possible that the mention of a body part in a stimulus item could have been used as a cue by participants to decide that the stimulus item was an insult. Thus, it may not have been the case that participants were creating richer sensorimotor simulations for the more embodied insults, which led to faster response latencies to these items. Instead, participants may have noticed that any stimulus item that mentioned a body part was an insult, and partly based their responses on that factor. (See above, however, for our proposal that the inclusion of a body part is not necessary for an item to be considered or judged a more embodied insult). To address this issue, we conducted Experiment 2 in which each more embodied insult that mentioned a body part (e.g., *asswipe*) was matched with a non-insult that also mentioned a body part (e.g., *shinpad*). Two outcomes were possible with this methodological change. First, if the mention of a body part was serving as an important cue in Experiment 1 that the item was an insult, then we should not find an effect of bodily experience in Experiment 2, because the mention of a body part in Experiment 2 does not reliably indicate that the item is an insult. Alternatively, if the mention of a body part was not responsible for the results in Experiment 1, but instead the more embodied insults were responded to faster than the less embodied insults because they elicited richer sensorimotor simulations, then facilitatory effects of bodily experience should again be observed in Experiment 2.

EXPERIMENT 2

METHOD

Participants

Thirty-three UNBC undergraduate students participated in the experiment for bonus course credit. All were native English speakers and reported normal or corrected-to-normal vision. None of these individuals participated in any of the other experiments.

Stimuli

The insult stimuli were those used in Experiment 1. Twenty-four of the non-insult stimuli were the same as those used in Experiment 1, and the remaining 12 were selected for this experiment. In this experiment there were as many non-insult stimuli that mentioned a body part (e.g., *shinpad*) as there were insult stimuli that mentioned a body part. The non-insult stimuli again consisted of either one or two words, and were matched as closely as possible to the insult stimuli on printed length. The new set of non-insult stimuli are listed in the Appendix.

Apparatus and procedure

The apparatus and procedure were identical to those used in Experiment 1.

RESULTS AND DISCUSSION

There were 32 observations (2.69% of the data) treated as outliers and removed from the data set in Experiment 2. Response errors for the insult stimuli comprised only 1.30% of trials, and, as such, the response error data were not analyzed. The mean response latencies of correct responses, mean error percentages, and correct recall percentages for all stimulus types are presented in **Table 2**.

There was a significant effect of bodily experience in the response latency data, $t(32) = 3.37$, SEM = 9.96, partial $\eta^2 = 0.26$, and in the surprise recall data, $t(32) = 2.86$, SEM = 2.30, partial $\eta^2 = 0.20$. In the response latency data, responses to the more embodied insults were an average of 33 ms faster than responses to the less embodied insults. In the surprise recall data, the percentage of items correctly recalled was 6.56% higher for more embodied insults than for less embodied insults.

The findings that participants again responded faster to and correctly recalled more of the more embodied insults than the less embodied insults indicates that the hypothesis that mere mention of a body part in a stimulus item was used as a cue for insult categorization (and thus produced the observed facilitatory effects) in Experiment 1 can be discounted. Therefore, the results again provide support for the notion that the facilitatory effects of bodily experience are due to richer elicited sensorimotor simulations, which facilitate detection of more embodied insults during the insult detection task and also cause better encoding and later retrieval for these items.

In Experiments 1 and 2 participants had to distinguish insulting stimuli from non-insulting stimuli, and the non-insult stimuli were not selected from any one single, coherent category. It is possible that this creates a more difficult decision than a situation in which the non-insults are from one coherent category (e.g., compliments). In a similar vein, Hino et al. (2002) suggested that semantic categorization tasks vary in the extent to which they require more versus less analytic processing. Hino et al. demonstrated that by changing the decision category participants could be encouraged to engage in more versus less analytic processing. The argument here would be that, even when the decision category (insult versus non-insult) remains the same, similar modulation might be invoked as a function of the non-insults presented. By this analysis, the insult detection tasks used in Experiments 1 and 2, with heterogeneous non-insults, could have evoked relatively analytic processing to determine whether a presented stimulus is an insult or not. Presumably, richer sensorimotor simulations facilitated this relatively extensive analytical processing. The fact that we observed faster processing of more embodied insults in Experiments 1 and 2 is consistent with this analysis.

It is possible that with a more coherent set of non-insults participants could perform the insult detection task with a relatively more cursory or shallow analysis of a small set of target features. In order to test this possibility, in the present Experiment 3 we modified the non-insult stimuli so that all of them were compliments. With this more coherent set of non-insults it may be possible for participants to categorize the stimuli based on shallower processing (e.g., the

Table 2 | Mean response latencies (in ms) and standard errors, mean response error percentages and standard errors, and percentage words correctly recalled and standard errors for Experiments 1–3.

	Experiment 1		Experiment 2		Experiment 3	
Insult type	***M***	**SE**	***M***	**SE**	***M***	**SE**
RESPONSE LATENCIES						
More embodied	734	31.2	711	25.5	762	26.1
Less embodied	762	37.4	744	24.4	809	31.1
Bodily experience effect	28	–	33	–	47	–
Control items	–	–	–	–	–	–
RESPONSE ERROR PERCENTAGES						
More embodied	1.71	0.8	1.41	0.4	0.36	0.3
Less embodied	2.91	0.9	1.20	0.5	2.07	0.7
Bodily experience effect	1.20	–	–0.21	–	1.71	–
Control items	3.06	0.7	3.23	0.8	3.31	0.7
PERCENTAGE WORDS CORRECTLY RECALLED						
More embodied	26.76	2.0	26.42	2.1	20.03	1.7
Less embodied	20.70	1.6	19.86	2.0	14.64	1.6
Bodily experience effect	6.06	–	6.56	–	5.39	–
Control items	7.24	1.0	8.33	1.2	16.3	1.5

emotional content of the stimuli – insults are negative, whereas compliments are positive). Therefore, in Experiment 3 we predicted two possible outcomes. First, there may not be a significant difference in response latencies between the more embodied and the less embodied insults. This null effect could arise if responses are indeed made after a relatively cursory or shallow analysis of a small set of target features, which may make the need for using simulations largely unnecessary. Second, there may be a small yet significant facilitatory effect of bodily experience. Participants may rely on a small set of target features to decide whether each stimulus is an insult or not, but simulations may still be used to make insult versus compliment decisions, which should lead to faster response latencies to the more embodied insults.

EXPERIMENT 3

METHOD

Participants

Thirty-three UNBC undergraduate students participated in the experiment for bonus course credit. All were native English speakers and reported normal or corrected-to-normal vision. None of these individuals participated in any of the other experiments.

Stimuli

The insult stimuli were those used in Experiment 1. The non-insult stimuli consisted of compliments (e.g., *hard worker*, *level headed*), and there were an equal number of insults and compliments that mentioned a body part. The compliments consisted of either one or two words, and were matched as closely as possible to the insult stimuli on printed length. The compliment stimuli are listed in the Appendix.

Apparatus and procedure

The apparatus and procedure were identical to those used in Experiment 1.

RESULTS AND DISCUSSION

There were 36 observations (3.03% of the data) treated as outliers and removed from the data set in Experiment 3. Response errors for the insult stimuli comprised only 1.22% of trials, and, as such, the response error data were not analyzed. The mean response latencies of correct responses, mean error percentages, and correct recall percentages for all stimulus types are presented in **Table 2**.

There was a significant effect of bodily experience in the response latency data, $t(32) = 3.11$, SEM = 15.19, partial $\eta^2 = 0.23$, and in the surprise recall data, $t(32) = 2.75$, SEM = 1.96, partial $\eta^2 = 0.19$. In the response latency data, responses to the more embodied insults were an average of 47 ms faster than responses to the less embodied insults. In the surprise recall data, the percentage of items correctly recalled was 5.39% higher for more embodied insults than for less embodied insults.

We suggested above that it may be possible to induce a more cursory or shallow analysis in an insult detection task by using a coherent set of non-insults (i.e., compliments), because responses could be partially or primarily based on a small set of target features, such as emotional content. We predicted that if such was the case, then there could be an attenuated or null effect of bodily experience in Experiment 3. However, the results for Experiment 3 are clear: bodily experience again facilitated responding and recall.

The findings from Experiments 1–3 are consistent with the idea that sensorimotor simulations were an important component of how insults were processed, regardless of the type of non-insults used. First, the effect size for the response latency data of the present experiment (partial $\eta^2 = 0.23$) is comparable to the effect sizes for the response latency data in Experiments 1 and 2 (partial $\eta^2 = 0.18$ and partial $\eta^2 = 0.26$, respectively). Second, the effect size for the recall data of the present experiment (partial $\eta^2 = 0.19$) is comparable to the effect sizes for the recall data in Experiments 1 and 2 (partial $\eta^2 = 0.17$ and partial $\eta^2 = 0.20$, respectively).

One potential criticism of the previous three experiments is that they used a relatively small stimulus set (even though the more embodied and less embodied insults were carefully matched on numerous confound variables). In order to address this potential criticism, Experiment 4 was designed to investigate whether the facilitatory effect of bodily experience observed in Experiments 1–3 could be extended to a larger set of items. To do this, we conducted another insult detection task, identical in procedure to the previous experiments, but consisting of a much larger number of items (from Siakaluk et al., in press). In this new stimulus set the more embodied insults and less embodied insults were not matched on any of the variables that were controlled in the previous experiments; instead, we controlled these variables statistically, using hierarchical multiple regression analyses for the response latency, response error, and recall data. On the first step of each analysis we entered the control variables of printed length, morphological complexity, HAL log-frequency, and frequency of usage, number of meanings, offensiveness, and imageability ratings as the predictors, and on the second step we entered bodily experience ratings as the predictor. We predicted that bodily experience ratings would account for a significant amount of unique response latency, response error, and recall variability, above and beyond the variability accounted for by the control variables.

EXPERIMENT 4

METHOD

Participants

Thirty-seven UNBC undergraduate students participated in the experiment for bonus course credit. All were native English speakers and reported normal or corrected-to-normal vision. None of these individuals participated in any of the other experiments. Data for seven of these participants were excluded from the analyses because they had error rates of more than 25% to the insult stimuli.

Stimuli

The full stimulus set from Siakaluk et al. (in press) was used in this experiment. Therefore, there were 178 insults (88 more embodied insults and 90 less embodied insults), and 179 non-insults. The experimental stimuli are available from the corresponding author.

Apparatus and procedure

The same procedure used in Experiments 1–3 was also used in Experiment 4, with the following exception: here participants were allowed to take a break after every 90 trials (the last block of trials always had 88 trials). The stimuli were presented in a different random order to each participant.

RESULTS AND DISCUSSION

Data for 19 of the insult stimuli (12 more embodied insults and 7 less embodied insults) were excluded from the analyses because response error rates for those items were greater than 25%. There were 145 observations (3.29% of the data) treated as response latency outliers and removed from the data set in Experiment 4.

Hierarchical multiple regression analyses were conducted to examine whether significant amounts of unique response latency, response error, and recall variability could be accounted for by bodily experience ratings. As noted, for each analysis, the control variables entered in step one were printed length, morphological complexity, HAL log-frequency, and ratings of frequency of usage, number of meanings, offensiveness, and imageability. Bodily experience ratings were entered in step two. Zero-order correlations between the predictor variables and the criterion variables are shown in **Table 3**. The multiple regression results for the response latency, response error, and recall data are presented in **Tables 4, 5, and 6**, respectively. Of most importance for the present study is the fact that the hypothesis that the bodily experience ratings would account for a significant amount of unique response latency, response error, and recall variability, above and beyond the variability accounted for by the control variables, was supported. Two statistically significant results are of immediate interest. First, for the response latency and error analyses the semi-partial correlation between bodily experience ratings and both criterion variables was negative, indicating that responses to relatively more embodied insults were faster and more accurate. Second, for the recall analysis the semi-partial correlation between bodily experience and percent recall was positive, indicating that greater recall occurred for relatively more embodied insults. In summary, the results from the present experiment replicate and extend the results from the first three experiments to a much larger stimulus set.

Table 3 | Zero-order correlations between the predictor variables and the criterion variables in Experiment 4.

Measure	1	2	3	4	5	6	7	8	9	10	11
1. IDT response latency	–										
2. IDT response error rate	0.47**	–									
3. IDT recall	–0.23**	–0.06	–								
4. Printed length	0.49**	0.18*	–0.12	–							
5. Morphological complexity	0.37**	0.11	–0.04	0.73**	–						
6. HAL log-frequency	0.05	0.03	0.03	0.11	–0.01	–					
7. Frequency ratings	–0.59**	–0.29**	0.34**	–0.41**	–0.36**	0.04	–				
8. Number of meanings	–0.10	–0.07	0.15	–0.16*	–0.13	0.38**	0.28**	–			
9. Offensiveness ratings	–0.24**	–0.43**	0.18*	0.10	0.04	0.08	0.37**	0.13	–		
10. Imageability ratings	0.09	0.16*	0.24**	–0.09	–0.05	0.04	0.10	0.43**	–0.02	–	
11. Bodily experience ratings	0.18*	0.03	0.17*	0.35**	0.36**	0.04	–0.31**	–0.02	–0.02	0.31**	–

Note. IDT, insult detection task; HAL, hyperspace analog to language.

$*p < 0.05$, $**p < 0.01$.

Table 4 | Results of hierarchical multiple regression analysis for insult detection task response latency data for Experiment 4.

Variable	*B*	SEB	β	sr	R^2	ΔR^2
Step 1 (control variables)					0.446***	
Step 2					0.487***	0.021*
Control variables						
Printed length	16.94	4.01	0.39	0.25***		
Morphological complexity	–0.61	14.81	–0.01	–0.00		
HAL log-frequency	0.72	1.78	0.03	0.02		
Frequency ratings	–46.23	7.37	–0.47	–0.37***		
Number of meanings	2.55	45.84	0.00	0.00		
Offensiveness ratings	–13.35	8.48	–0.10	–0.09		
Imageability ratings	24.07	7.68	0.22	0.18**		
Bodily experience ratings	–12.73	5.20	–0.17	–0.14*		

Note. HAL, hyperspace analog to language. The B, SEB, β, and sr values are for the final step in the analysis, where all the predictor variables were included in the equation.
**p < 0.05, **p < 0.01, ***p < 0.001*

Table 5 | Results of hierarchical multiple regression analysis for insult detection task response error data for Experiment 4.

Variable	*B*	SEB	β	sr	R^2	ΔR^2
Step 1 (control variables)					0.271***	
Step 2					0.292***	0.022*
Control variables						
Printed length	0.66	0.26	0.27	0.17*		
Morphological complexity	–0.41	0.97	–0.04	–0.03		
HAL log-frequency	0.10	0.12	0.07	0.06		
Frequency ratings	–0.53	0.48	–0.10	–0.08		
Number of meanings	–3.24	3.00	–0.09	–0.07		
Offensiveness ratings	–2.89	0.56	–0.41	–0.36***		
Imageability ratings	1.68	0.50	0.28	0.23**		
Bodily experience ratings	–0.73	0.34	–0.18	–0.15*		

Note. HAL, hyperspace analog to language. The B, SEB, β, and sr values are for the final step in the analysis, where all the predictor variables were included in the equation.
**p < 0.05, **p < 0.01, ***p < 0.001*

Table 6 | Results of hierarchical multiple regression analysis for insult detection task word recall data for Experiment 4.

Variable	*B*	SEB	β	sr	R^2	ΔR^2
Step 1 (control variables)					0.174**	
Step 2					0.220**	0.046*
Control variables						
Printed length	–0.54	0.48	–0.13	–0.08		
Morphological complexity	1.55	1.77	0.10	0.06		
HAL log-frequency	0.06	0.21	0.02	0.02		
Frequency ratings	3.56	0.88	0.38	0.29**		
Number of meanings	–2.05	5.49	–0.03	–0.03		
Offensiveness ratings	0.73	1.02	0.06	0.05		
Imageability ratings	1.31	0.92	0.13	0.10		
Bodily experience ratings	1.86	0.62	0.26	0.22*		

Note. HAL, hyperspace analog to language. The B, SEB, β, and sr values are for the final step in the analysis, where all the predictor variables were included in the equation.
**p < 0.01, **p < 0.001*

GENERAL DISCUSSION

In the present research we used a task called an insult detection task (a type of semantic categorization task) to examine the influence of bodily knowledge on insult processing. In previous research it has been reported that richer semantic representations lead to faster responding in semantic categorization tasks (e.g., Siakaluk et al., 2003, 2008b; Pexman et al., 2008), and by this logic, we predicted that because more embodied insults should elicit richer sensorimotor simulations (i.e., representations), responses would be faster to more embodied insults as compared to less embodied insults. As predicted, a facilitatory effect for more embodied insults was observed in the response latency data for each of the insult detection tasks reported in the present research (as well as in the response error data for Experiment 4). As such, the results from the present research support the inference that knowledge gained through bodily experience plays an important role in the detection of insults.

After each experiment we conducted a surprise recall task. We predicted that participants would recall a greater percentage of more embodied insults than less embodied insults, because the former type of insult would elicit richer sensorimotor simulations, which should aid in their subsequent recall. As predicted, a greater percentage of more embodied insults were correctly recalled in each of the four experiments. The recall results demonstrate that there are facilitatory effects of bodily experience that are still present after completion of the insult detection task. As such, the embodied effects we observed in the insult detection task are not fleeting but rather influence behavior subsequent to the detection task.

As noted in the Introduction, a theory that can be used to explain the facilitatory effects of bodily experience in insult detection and recall performance is Barsalou's (1999) perceptual symbol systems theory. By this account, bodily interactions with the environment are crucial to many cognitive processes. Barsalou (1999) suggested that the modality-specific neural systems used for perception, action, and introspection are also used to represent concepts in the brain, through the process of simulation. According to Barsalou (1999, 2008), simulation is the partial neural re-enactment of perceptual, motor, and introspective states acquired during bodily experience with the environment. When we conceptualize an object or event, what is occurring is that we are neurally re-experiencing the sensory, motor, and introspective components associated with that object or event from previous bodily experience.

The results of the present study are consistent with the idea that more embodied insults are at least partly understood by creating simulations of bodily experience. As noted above, while simulations for emotional knowledge (e.g., negative emotions evoked by insults) and introspective knowledge (e.g., contextual or situational knowledge gained from previous experiences using or being called a particular insult) would likely be equally rich for more embodied and less embodied insults, the two types of insults differ in the amount of knowledge that is gained through bodily (i.e., sensory, motor, kinesthetic, and proprioceptive) experience. For example, we may simulate the sensation of numbness when we comprehend the insult *numbskull*. Therefore, because of knowledge based on bodily experience, more embodied insults elicit richer sensorimotor simulations (i.e., greater neural re-enactment of perceptual, motor, kinesthetic, and proprioceptive states; Barsalou, 1999, 2008) than less embodied insults. As a result, as demonstrated in the present research, richer sensorimotor simulations lead to faster insult detection latencies. Barsalou's (1999, 2008) notion of simulation can also be used to explain the recall findings. The richer simulations based on knowledge gained through previous bodily experience that were elicited for the more embodied insults during encoding (i.e., when they were being processed during the insult detection task) were more easily reactivated during retrieval (i.e., during the time period allotted to participants to recall as many stimuli as they could), resulting in a greater percentage of more embodied insults being recalled.

In Experiments 2 and 3 of the present research we were able to discount hypotheses suggesting that simulations may not be playing a crucial role in insult processing. One potential issue with Experiment 1 was that 12 of the more embodied insults mentioned a body part, whereas only four of the non-insults mentioned a body part. We therefore hypothesized that perhaps the mention of a body part in a stimulus item could have been used as a cue by participants that the stimulus item was an insult. That is, it may not have been the case that participants were creating richer sensorimotor simulations for the more embodied insults, which led to faster response latencies to these items. Instead, the participants may have noticed that any stimulus with mention of a body part was an insult, and partly based their responses on that factor. To address this issue, Experiment 2 was conducted, in which each more embodied insult that mentioned a body part (e.g., *asswipe*) was matched with a non-insult that also mentioned a body part (e.g., *shinpad*). The results indicated that participants again responded faster to the more embodied insults and recalled a greater percentage of more embodied insults, indicating that the hypothesis that the mention of a body part in a stimulus item was underlying the observed facilitatory effect in Experiment 1 could be discounted. Therefore, the results from Experiment 2 again provided support for the notion that the facilitatory effects of bodily experience are due to richer elicited sensorimotor simulations for the more embodied insults.

In Experiment 3 we examined the possibility that with a more coherent set of non-insults participants could perform the insult detection task with a relatively more cursory or shallow analysis of a small set of target features, thus rendering simulations unnecessary. To examine this possibility, all the non-insults used in Experiment 3 belonged to the same category (i.e., compliments) as opposed to the more general sets of non-insults used in the first two experiments. The results indicated that under conditions in which a well-defined decision category was used (i.e., insults versus compliments) participants still responded faster to and recalled more of the more embodied insults than the less embodied insults, which replicated what was observed in Experiments 1 and 2. Effect sizes (presented in the Results section of Experiment 3) indicated that there was no difference in the magnitude of the bodily experience effect between Experiments 1, 2, and 3. Therefore, it seems likely that simulations were underlying responding in all three experiments, and were not modulated by the type of non-insult used.

The present finding of a facilitatory effect of bodily experience on insult detection (and recall) contrasts with the findings of Siakaluk et al. (in press), who examined the effects of bodily experience in a version of the Stroop task. As noted, Siakaluk et al. (in press) reported an inhibitory bodily experience effect in the Stroop

task such that color naming latencies were slower to relatively more embodied insults. As noted above, the nature of the effect that bodily experience has on performance (i.e., either a facilitatory or an inhibitory effect) appears to be a function of task demands. In the present study, richer sensorimotor simulations provide more evidence that an item is a positive instance of the category "insult." In contrast, in Siakaluk et al.'s (in press) insult Stroop task, task demands required that the participants pronounce the font color. Thus, in the Stroop task, richer sensorimotor simulations hinder participants' ability to pronounce the font color, which leads to slower color naming latencies for the more embodied insults.

The results of the present study extend the findings of Siakaluk et al. (in press) in two important ways. First, as noted above, although Siakaluk et al. (in press) assumed that it was the richer meanings of the more embodied insults that inhibited Stroop task performance, the Stroop task is not a direct measure of the influence of word meaning. The insult detection tasks used in the present study address and resolve this concern because the task more directly assesses the processing of insult meanings. Since the predicted facilitatory effect of bodily experience was observed in a task requiring the processing of insults, we can be more confident that it was indeed the processing of insult meaning that lead to the slowing of Stroop task performance in Siakaluk et al.'s (in press) study. Second, the fact that bodily experience has been observed to both facilitate and inhibit responding (depending on the task and its demands) suggests that simulation is not a process that is simply invoked when convenient, or when it might help participants to be more efficient in a particular task. Rather, the fact that the bodily experience dimension can have both positive and negative consequences for performance suggests that it is an important, non-optional aspect of processing insults.

CONCLUSION

One of the aims of this special issue is to provide an account for the way abstract concepts are represented. The present research has taken a step toward this goal by providing new understanding about how insults are processed. One can easily imagine a situation in which someone did not act with intelligence and is called a *numbskull*. In this situation, one immediately understands two things. First, the term *numbskull* is meant as an insult, and, second, the insult is not based on the individual's skull literally being in a state of numbness. How, then, is it possible to understand the seemingly abstract meaning of *numbskull*?

We proposed above that there are at least two ways an insult can be thought of as abstract. First, many insults (e.g., *stupid*) express a quality or attribute that is distinct from any particular person. This conceptualization of abstractness seems appropriate for many of the insults in Experiment 4, which fit the sentence, "You are _____." However, many other insults do not seem appropriate for this conceptualization of abstractness. A second way of conceptualizing the abstractness of insults is that they require metaphorical understanding. That is, they require knowledge gained from one context to be applied to a different context. In the specific case of the insult *numbskull*, comprehension requires that knowledge gained from the context of bodily experience be applied to a different context, such as one's intellectual ability (e.g., *You are a numbskull*). In this, and many other cases, it is likely that the intended insult meanings are understood metaphorically, because knowledge used from bodily experience cannot *literally* be true in the context of one's intellectual ability (i.e., it is not the case that the person to whom the insult is directed literally has a skull that is numb). This conceptualization of abstractness is likely to be appropriate for many of the insults in Experiments 1–3, of which the majority fit the sentence, "You are a(n) _____," although it is likely to be relevant for many of the insults used in Experiment 4 as well. Finally, we have proposed that Barsalou's (1999, 2008) perceptual symbols systems theory provides an elegant account of how simulation, or neural re-enactment, grounds abstract meanings of insults in modality-specific neural systems used for perception, action, and introspection. In this way, we have extended grounded cognition to a new context; that is, comprehension of insults.

REFERENCES

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Cohen, J. D., Dunbar, K., and McClelland, J. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332–361.

Cowart, M. (2004). Embodied cognition. *Internet Encyclopedia of Philosophy*, http://wwww.iep.utm.edu/.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61, 905–919.

Hino, Y., Lupker, S., and Pexman, P. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: interactions between orthography, phonology, and semantics. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 686–713.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., and Dyer, J. R. (2004). Relations between emotion, memory, and attention: evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Mem. Cognit.* 32, 474–488.

Myung, J.-Y., Blumstein, S. E., and Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition* 98, 223–243.

Neisser, J. (2003). The swaying form: imagination, metaphor, embodiment. *Phenomenol. Cogn. Sci.* 2, 27–53.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Siakaluk, P. D., Buchanan, L., and Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Mem. Cognit.* 31, 100–113.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008a). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., and Owen, W. J. (2008b). The benefits of sensorimotor knowledge: body–object interaction facilitates semantic processing. *Cogn. Sci.* 32, 591–605.

Siakaluk, P. D., Pexman, P. M., Dalrymple, H.-A. R., Stearns, J., and Owen, W. J. (in press). Some insults are more difficult to ignore: the embodied insult Stroop effect. *Lang. Cogn. Process.*

Siakaluk, P. D., Pexman, P. M., Sears, C., and Owen, W. (2007). Multiple mean-

ings are not necessarily a disadvantage in semantic processing: evidence from homophone effects in semantic categorisation. *Lang. Cogn. Process.* 22, 453–467.

Tillotson, S. M., Siakaluk, P. D., and Pexman, P. M. (2008). Body-object interaction ratings for 1,618 monosyllabic nouns. *Behav. Res. Methods* 40, 1075–1078.

Wellsby, M., Siakaluk, P. D., Owen, W. J., and Pexman, P. M. (in press). Embodied semantic processing: the body-object interaction effect in a non-manual task. *Lang. Cogn.*

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Wilson, N. L., and Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cogn. Sci.* 31, 721–731.

Yaxley, R. H., and Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition* 150, 229–236.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 July 2010; accepted: 22 October 2010; published online: 16 November 2010.

Citation: Wellsby M, Siakaluk PD, Pexman PM and Owen WJ (2010) Some insults are easier to detect: the embodied insult detection effect. Front. Psychology ***1****:198. doi: 10.3389/fpsyg.2010.00198*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Wellsby, Siakaluk, Pexman and Owen. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

APPENDIX

STIMULI USED IN EXPERIMENTS 1–3

More embodied insults

airhead, asshole, asswipe, brown noser, butt head, crybaby, dick head, dink, douchebag, dumbass, jackass, jerk off, loud mouth, pinhead, potty mouth, prick, ugly, weak

Less embodied insults

bastard, bugger, cheapskate, cheater, crazy, disgrace, dork, drama queen, ignorant, immature, lunatic, moron, nerd, pansy, retard, stupid, twit, wannabe

Non-insult stimuli used in Experiment 1

airbase, armband, ashtray, ball cap, bandage, burlap, capsize, carve, chandelier, climate, clock work, darn, dashboard, day dream, denture, diameter, disc, gas station, immunize, intrigue, jawbone, lottery, mocha, node, plant pot, polka, pullout, purse, ribbed, simple life, sizzle, summer love, tale, undo, weekend, wife

Non-insult stimuli used in Experiment 2

armband, ball cap, bandage, blue eyes, burlap, capsize, carve, chandelier, climate, daisy, darn, dashboard, diameter, foot, fun, haircut, immunize, intrigue, jawbone, kidneys, knee brace, lottery, mocha, node, nose hair, polka, purse, ribbed, rough hands, shinpad, sizzle, streaky, summer love, tale, toenail, undo, utensil, weekend, white teeth, wife

Compliments used in Experiment 3

agile, amazing, angel, awesome, ballsy, brainy, brave, bright eyed, clear skinned, cute, dependable, dreamy, eyeful, flexible, fragrant, friendly, funny, generous, genius, gutsy, handy, hardcore, hard worker, helpful, honest, level headed, limber, lovable, mindful, muscular, patient, sexy, sharp eyed, slender, smooth, sociable, soulful, strong, stylish, wise

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 18 March 2011
doi: 10.3389/fpsyg.2011.00045

Switching modalities in a sentence verification task: ERP evidence for embodied language processing

Lea A. Hald[1]*, Julie-Ann Marshall[1,2], Dirk P. Janssen[3] and Alan Garnham[4]

[1] Department of Applied Social Sciences, Canterbury Christ Church University, Canterbury, UK
[2] Health Care of Older People, William Harvey Hospital, Ashford, UK
[3] Academy for Digital Entertainment, NHTV Breda University of Applied Sciences, Breda, Netherlands
[4] School of Psychology, University of Sussex, Brighton, UK

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Jennifer Collins, University of California San Diego, USA
Louise Connell, University of Manchester, UK

***Correspondence:**
Lea A. Hald, Department of Applied Social Sciences, Canterbury Christ Church University, North Holmes Road, Canterbury CT1 1QU, UK.
e-mail: lea.hald@gmail.com

In an event related potential (ERP) experiment using written language materials only, we investigated a potential modulation of the N400 by the modality switch effect. The modality switch effect occurs when a first sentence, describing a fact grounded in one modality, is followed by a second sentence describing a second fact grounded in a different modality. For example, "A cellar is *dark*" (visual), was preceded by either another visual property "Ham is pink" or by a tactile property "A mitten is soft." We also investigated whether the modality switch effect occurs for false sentences ("A cellar is *light*"). We found that, for true sentences, the ERP at the critical word "*dark*" elicited a significantly greater frontal, early N400-like effect (270–370 ms) when there was a modality mismatch than when there was a modality-match. This pattern was not found for the critical word "*light*" in false sentences. Results similar to the frontal negativity were obtained in a late time window (500–700 ms). The obtained ERP effect is similar to one previously obtained for pictures. We conclude that in this paradigm we obtained fast access to conceptual properties for modality-matched pairs, which leads to embodiment effects similar to those previously obtained with pictorial stimuli.

Keywords: ERP, N400, embodiment, language processing, veracity, modality, modality switch effect

INTRODUCTION

The idea that our conceptual system is grounded in modality-specific or embodied simulations has received support from many different areas of research including psychology, neuroscience, cognitive modeling, and philosophy (Lakoff and Johnson, 1999; Gibbs, 2005; Pecher and Zwaan, 2005, for reviews). The suggestion that modality-specific simulation also affects language processing has been put forward by a number of authors (Glenberg, 1997; Barsalou, 1999; Glenberg and Robertson, 1999, 2000; Zwaan, 2004; Zwaan and Madden, 2005). For example, Barsalou's (1999) theory of perceptual symbol systems suggests that modality-specific simulations arise from perceptual states and that these (simulated) states underlie the representation of concepts. Hence, all conceptual symbols are grounded in modality-specific states. Linguistic symbols develop alongside the perceptual symbols that they are linked to so that when we use or encounter words, we simulate the perceptual states that are linked to the linguistic information. Such a source of perceptual state simulations is called a simulator by Barsalou (1999).

Zwaan and Madden (2005) similarly assume language is grounded in perception and action via something akin to Barsalou's (1999) perceptual symbols. However, they focus specifically on how language guides the simulators. They assume that what we simulate is based on attentional frames (Langacker, 2001). In particular, within one attentional frame we construct a "construal": a simulation that includes time, spatial information, perspective, and a focal and background entity (for details see Zwaan and Madden, 2005). Furthermore, during construal, information from previous construals forms the context with which we integrate the information from the current construal. This is how we understand connected discourse (Zwaan, 2004). For Glenberg (1997; Glenberg and Robertson, 1999, 2000) the key issue is that we use the perceptual symbols to derive affordances, in the sense of Gibson (1986), for the specific situation. Understanding a sentence is a result of meshing the affordances, which is guided by the syntax of the sentence.

Evidence for the modal grounding of conceptual and linguistic representations has been found using a variety of techniques and tasks. Only a few key findings relative to the current experiment will be reviewed here. Goldberg et al. (2006) measured fMRI BOLD responses while participants did a blocked property verification task. Participants had to press the button for each word that had the property "*green*" (visual), "*soft*" (tactile), "*loud*" (auditory), or "*sweet*" (gustatory). The results for visual and tactile decisions showed increased activation in visual and somatosensory cortex when compared to control, which supports the notion of modal grounding.

Using a behavioral measure and the same paradigm, Pecher et al. (2003) established that there is a cost to switching modalities. They presented participants with short sentences that consisted of a concept followed by a modal property (they used audition, vision, taste, smell, touch, and action). For example, after reading "*blender can be loud*," participants were asked to decide whether "*loud*" is a typical property of "*blender.*" Crucially, half of the experimental trials were preceded by a trial of the same modality (matched modality, "*leaves can be rustling*" – "*blender can be loud*") while the other half were preceded by a trial of a different (mismatched) modality (e.g., "*cranberries can be tart*" – "*blender can be loud*"). Participants were able to verify the property of the concepts faster

and more accurately in matched modality trials than in mismatched modality trials. Similar modality switch effects have been found in other studies across both conceptual and perceptual processing tasks (e.g., Spence et al., 2001; Marques, 2006; Vermeulen et al., 2007; Van Dantzig et al., 2008).

If the mental simulations that are required for understanding involve the premotor areas, keeping these areas otherwise involved should interfere with language comprehension. This has indeed been demonstrated, for example by Zwaan and Taylor (2006), who found that reading about an action which involves clockwise turning (e.g., increasing the volume on a radio) interfered with the action of turning a knob counterclockwise. More abstractly, Glenberg and Kaschak (2002) showed that reading a sentence which involves transferring an object or information away from the participant ("You told Liz the story") interfered with that participant pressing a response button which was located toward their body as compared to a button which was located away from their body.

Evidence from tasks not involving large physical movements comes from sentence–picture verification tasks. For example, a sentence such as "*John pounded the nail into the floor*" was followed by a picture of a nail (Stanfield and Zwaan, 2001). Response times were faster when the picture matched the orientation implied in the sentence (vertical) compared to when there was a mismatch in orientation (see also Zwaan et al., 2002).

What is striking about the behavioral studies described here is the number of innovative tasks and procedures that were created in order to show that concepts and language are grounded in bodily states. While these and other sets of studies form a convincing body of literature, one might question how effects related to embodied cognition might be evident in other tasks that are more standardly used within language comprehension research. If simulating linguistic or conceptual material in terms of our bodily states is the norm, we should see evidence of it in any standard task if properly designed and analyzed. This is important as it could be argued that the use of tasks involving movement and pictures encouraged participants to use an imagery based strategy (e.g., Glucksberg et al., 1973), which would make the embodiment results specific to the tasks that are used in this field. Of course, exceptions already exist: Results from neuroimaging studies where participants read either single words or sentences referring to bodily actions support the embodied view by showing increased activation in the premotor and sometimes the primary motor areas of the cortex (for example, Hauk et al., 2004; Boulenger et al., 2009). Recent findings using the sentence–picture verification task also suggest that the results are not due to the use of imagery as a strategy (Pecher et al., 2009) and the study by Pecher et al. (2003) did use solely linguistic stimuli, albeit in a slightly unnatural task. If embodied simulation is a part of everyday language comprehension, we should be able to find evidence for it using the standard language comprehension techniques that do not involve pictures or movements. In the current study we will therefore use the well-studied paradigm of the sentence verification task (Meyer, 1970). Before discussing results related to this task, we will quickly outline our experiment to frame the discussion below.

The materials of the current study were adapted from the design used by Pecher et al. (2003) to a sentence verification task. We drew our materials from items that have previously been rated as having either a salient visual or a salient tactile property (Pecher et al., 2003; Van Dantzig et al., 2008; Lynott and Connell, 2009; Van Dantzig and Pecher, submitted; see Materials and Methods for details). From this set of concepts with salient modality features, we created true statements "*A cellar is dark*" (visual). As in Pecher et al. (2003), sentences were presented one by one and for the participants, appeared to be unrelated. However, the critical manipulation was that sentences that followed each other were either matched in the salient modality (e.g., visual–visual, "*Ham is pink*" – "*A cellar is dark*") or mismatched in the salient modality (e.g., tactile–visual, "*A mitten is soft*" – "*A cellar is dark*"). We crossed modality with veracity by making half of the experimental target sentences false, while maintaining the same modality information ("*A cellar is light*").

We will now first review the links between property verification and sentence verification and then discuss the previous findings on veracity. As mentioned above, Pecher et al. (2003) asked participants to perform a conceptual-property verification task for statements such as "*blender/can be/loud*" (slashes indicate line breaks on the computer monitor). The participants were asked to verify that the property (always shown on the third line) is "usually true" of the concept (always shown on the first line) and had to respond with a true or false response. In the current study, it was decided to change the task to sentence verification. Sentence verification is a similar but more general task and has been used extensively in early sentence processing literature (for a review see Carpenter and Just, 1975) and in event related potential (ERP) experiments (e.g., Fischler et al., 1983). In this task, sentences are presented and subjects respond with a true or false judgment at the end of the sentence. Some items are almost identical between tasks ("*A blender can be loud*"), others can only be used in the sentence verification ("*A baby drinks milk*"). In our version of sentence verification, the words are presented one by one in the middle of the screen, which leads to a relatively natural reading experience, while avoiding eye movements. Using the sentence verification task, the typical finding is that false sentences take longer to verify than true sentences (e.g., Fischler et al., 1983).

The majority of the response time literature on veracity investigates the time to decide whether a sentence is a true or false representation of a corresponding picture ("*The dots are red*" with a picture showing either red or blue dots). In this situation, true sentences have been consistently shown to be verified faster than false sentences (for example, Trabasso et al., 1971; Clark and Chase, 1972; Wason, 1980). The primary explanation for this is that readers match the color red to the color of the dots. When this is congruent, readers are facilitated; when the colors are incongruent there is a slow down (Carpenter and Just, 1975; see also Fischler et al., 1983).

In this paper, we will try to obtain further empirical evidence for an embodied approach and we will discuss how an embodied language comprehension system can explain the current and past findings through the process of simulation. In an embodied view, determining the veracity of a statement depends on the outcome of a simulation and the comprehension process should be modulated by direct or indirect effects of simulation.

The reaction time effects for modality switch are quite subtle so we decided to use a more sensitive technique for this study: To explore the processing dynamics of modality switching, veracity,

and their interaction, we will look at the presence and significance of modulations in the ERP. If embodied simulation is an automatic process that occurs when we understand language, evidence of modality switching, veracity effects, and their interaction should be evident in ERPs. Predictions relative to modality switching are discussed below followed by veracity predictions.

One possible prediction of the effect of modality switching would be a modulation of the N400 effect. Although often incorrectly thought of as an increased negativity that occurs only to semantic anomalies (e.g., "*He spread the warm bread with socks/butter*"; Kutas and Hillyard, 1980), a large body of research suggests that semantic anomalies are nether necessary or sufficient to elicit an N400 effect (see Kutas et al., 2006, for review). Instead results show that a (small) N400 occurs as a response to each meaningful word as part of normal processing (Van Petten, 1995). The amplitude of the N400 is sensitive to many different semantic and linguistic factors [for example, Cloze probability (Taylor, 1953), word frequency, word class, and discourse context]. Furthermore, relative to the veracity, a consistently larger amplitude N400 is seen for words that change the veracity of a single sentence (at the critical word, here shown in bold for "a ham is **blue**" versus "a ham is **pink**"; Fischler et al., 1983; Hagoort et al., 2004).

Given this range of meaning effects modulating the N400 and the behavioral findings that switching modalities leads to a processing cost (Pecher et al., 2003), a reasonable expectation is that the N400 effect could be modulated by modality switching. We think it is a priori unlikely that a modality switch would trigger a sizeable N400 by itself, as "*a cellar is dark*" is usually a true, semantically coherent statement, even after a tactile context like "*a mitten is soft.*" However, the N400 is sensitive to the *integration* of incoming semantic information into the ongoing representation: Assuming that the ongoing representation is indeed embodied, a switch in the modality may lead to an earlier effect than the N400 (because modality switching should occur before integration), and the modality switch may modulate (enhance or suppress) the N400 itself. The effect of modality switch on the N400 may not be linear, as is known to be the case for word frequency and context (Van Petten and Kutas, 1990). Specifically, one may predict that a match in modality may lead to easier simulation and therefore a reduction or absence of the N400 for integration. Alternatively, there could be an ERP effect that occurs earlier than the N400, which is specifically indicative of the simulation itself.

A second question addressed in our study is what happens when the target sentence is false or commonly false ("*a cellar is light*"). We know that the veracity of the sentence can modulate the N400 (Fischler et al., 1983, with sentence verification task; Hagoort et al., 2004, with no task given). Similar N400 modulation results were found using a task where participants were required to determine whether a probe word was related conceptually to the precious sentence (for example, "flute" following "Mozart was a musical child prodigy"; Nieuwland and Kuperberg, 2008). To better understand the effects of modality switching, we investigate whether there is an interaction between effects for the veracity of the sentence and effects of the modality switch. Barsalou (1999) suggests that when a false sentence is read the simulation fails, which means that the meaning of the sentence cannot be successfully mapped onto reality. After a simulation fails presumably a new simulation is carried out that is grounded in the failed simulation (Barsalou, 1999, p. 601). However, as in the response time literature on veracity discussed above, Barsalou (1999) discusses false sentences in a context where one compares the sentence to a situation (or picture) immediately in front of people, not what would happen when something is false based on background knowledge. Nonetheless, if false sentences lead to a failure of simulation, this may lead to a different ERP modulation based on the point at which the simulation fails. Considering false sentences take longer to verify than true sentences, one might expect the ERP modulation relative to modality switching to occur later in the time course of processing.

MATERIALS AND METHODS

PARTICIPANTS

Sixteen native speakers of English recruited from Canterbury Christ Church University and the University of Kent participated in this experiment, 10 of whom were included in the final analysis (eight females; aged 18–22, mean = 19.7). They were paid a small fee for their participation. All participants had normal or corrected-to-normal vision, and normal hearing and all were right handed. None of the participants had any neurological impairment and none of them had participated in the pretests (see below). The six participants (37.5%) who were excluded from the final analysis were rejected for the following reasons: excessive artifacts (eye-movements, excessive noise from muscle tension, two participants, see EEG Recording and Analysis below for details), technical problems with recording (one participant), reaction time errors over 25% (two participants), and non-native English speaker (one participant). Ethical approval for the ERP study and the pretest was obtained from the Canterbury Christ Church University Faculty Research Ethics Committee, which follows the British Psychological Society guidelines for ethics on human subject testing. All participants signed a consent form prior to participating in the ERP experiment and the pretests.

STIMULUS MATERIAL AND DESIGN

The experimental materials comprised 160 pairs of sentences. Each pair consisted of a first sentence, which we will call the modality context sentence followed by a second sentence, the target sentence. The modality context sentences were always semantically correct, true statements which either described a salient tactile property (tactile context) of an object or a salient visual property (visual context) of an object. We selected a subset of the items that have been previously rated as having one modality that was clearly dominant in people's perception of that item (ratings from Pecher et al., 2003; Van Dantzig et al., 2008; Lynott and Connell, 2009; Van Dantzig and Pecher, submitted). The target sentence either matched the modality of the modality context sentence or mismatched. Additionally, the target sentence could either be true or false. False versions of the target sentences were created by using a word that was rated in a pretest to be the opposite of the salient feature of the object. For example, for "*A cellar is dark*" the word "*light*" was independently rated as the opposite of dark and it was used to create the false version. The false target sentences always contained a property in the same modality as the true target sentences. By using opposites we can keep the format of the true and false sentences

identical, but one may wonder what the effect of an opposite is. Furthermore, many of the studies looking at veracity have used opposites to create false sentences. For example, Nieuwland and Kuperberg (2008) looked at true and false sentences where all false sentences were created by using opposites (for example, "*With proper equipment, scuba-diving is very safe/dangerous…*") and found a typical N400 effect for false sentences compared to true (see also Hald et al., 2005 for similar use of opposites for creating false sentences.

The conditions modality-match and veracity of the target sentence were fully crossed, with 40 pairs in each of the four cells. Half of these 40 target sentences were visual, the other half tactile (see **Table 1** for example materials). Eighty false–false filler pairs were added to balance the number of true and false targets. The filler pairs also contained strongly modality related properties in half of the sentences, using tactile, visual, auditory, and gustatory modalities. The other half of the fillers were not based on modality-specific information but instead contained highly related words, while conveying false information (e.g., *A ball is refereed*; see Pecher et al., 2003, for similar use of semantically related filler items).

The critical words were matched across conditions on the following criteria: (i) word log (lemma) frequency (true–matched modality: 2.32; true–mismatched modality: 2.32; false–matched modality: 2.37; false–mismatched modality: 2.37; from Baayen et al., 1993); (ii) word length (true–matched modality: 4.8 letters; true–mismatched modality: 4.8 letters; false–matched modality: 4.5 letters; false–mismatched modality: 4.5 letters), (iii) word class (all adjectives). None of the critical words was over 12 letters in length.

The 240 pairs of sentences were presented in a pseudorandomized order specific to each participant (created using the program Mix, Van Casteren and Davis, 2006) using a fully within-participants design. The use of within-participants manipulation kept the design similar to that of Pecher et al. (2003), where matched versus mismatched modality was manipulated within-participants. Furthermore, in previous ERP sentence verification experiments a within-participants design was also utilized (Fischler et al., 1983; see also Hald et al., 2005, for a direct comparison of within and between-participants design using a sentence verification task).

Table 1 | Example materials for tactile and visual modality.

Veracity	Modality-match	Modality context	Target sentence
TACTILE TARGET SENTENCE EXAMPLE			
True	Mismatched	A leopard is spotted.	A peach is **soft**.
	Matched	An iron is hot.	A peach is **soft**.
False	Mismatched	A leopard is spotted.	A peach is **hard**.
	Matched	An iron is hot.	A peach is **hard**.
VISUAL TARGET SENTENCE EXAMPLE			
True	Mismatched	A mitten is soft.	A cellar is **dark**.
	Matched	Ham is pink.	A cellar is **dark**.
False	Mismatched	A mitten is soft.	A cellar is **light**.
	Matched	Ham is pink.	A cellar is **light**.

Critical words are shown here in bold for clarification.

PRETEST

In order to create false versions of the target sentences that maintained the same modality information, we decided to replace the adjectives with their opposites. For example, if the target sentence used the critical word "*dark*," we had people rate on a 7-point scale (7 = strongly agree): "*The opposite of dark is light.*"

These opposites do not form anomalous sentences of the kind that were used for the first N400 experiments (Kutas and Hillyard, 1980; Kutas et al., 2006). However, previous research has shown that people can show N400 effects to sentences that are at odds with their basic world knowledge (Hagoort et al., 2004; Hald et al., 2007). By using opposites, we were able to construct an experiment without anomalous sentences and we were able to use properties from tactile and visual modalities for all experimental items.

We tested a total of 52 different candidate opposites to arrive at the final set of 40. In addition to looking at possible opposites of all critical words, we also included fillers of two different types in this pretest to make sure participants were using the full scale in their ratings. Twenty fillers had properties that are difficult to assign an opposite to, for example "*The opposite of checkered is striped.*" The other 20 fillers were based on words that are related but clearly not opposites "*The opposite of clean is polished.*"

For the visual modality, target words were often color terms (37.5% of the time). Although in one technical sense colors do have opposites (complementary colors), these opposites may not be conceived as such by ordinary language users in the same way as terms such as "*dark*" and "*light.*" For that reason we tested all color word opposites (such as "*Black is the opposite of white*") separately, in a list with fillers that were also all color words. We encouraged the participants to use the full scale by including fillers that were related but clearly not opposites ("*Magenta is the opposite of violet*") and fillers that were difficult to judge ("*Black is the opposite of fuchsia*"). For the non-color pretest 27 native English speakers (eight males; mean age = 31) and for the color terms 37 participants (11 males; mean age = 31) completed the ratings online using SurveyMonkey[1]. We selected the words that were rated most highly as opposite as the adjectives for the false condition. The mean rating for the non-color list was 5.75 (SD = 0.52) and for the color list 4.61 (SD = 0.74). Although the color words were rated lower (less opposite) than the non-color words, the key issue for the sentence verification task is that using these words makes the sentences false. Thus, we had a set of clearly false statements that retained the same modality as the true statements, and had very similar content.

PROCEDURE FOR THE ERP STUDY

Participants were asked to fill out a questionnaire about their language and basic health background. Additionally, participants filled out a handedness questionnaire (Oldfield, 1971) and signed a consent form. Participants were tested individually in quiet room, seated in a comfortable chair approximately 70 cm away from a computer monitor. Participants were asked to read the sentences for comprehension and decide whether each sentence was true or false. They were also asked to try not to move or blink during the presentation of the sentences on the computer screen. No other tasks were imposed.

[1]www.surveymonkey.com

The experimental stimuli were presented using E-Prime 2.0 (Schneider et al., 2002). The experimental session began with a practice block of 10 sentences, which were similar in nature to the experimental items. At the end of the practice block the participant had a chance to ask any questions they had about the task. The remaining sentences were split into six blocks lasting approximately 12 min each. A short break followed each block. Each block began with two filler items, which were similar in nature to the experimental items. These filler items were included to minimize loss of data due to artifacts after beginning a new block.

Each trial began with a fixation ("+++") displayed for 1 s in the middle of the computer screen. The participants were told they could blink their eyes during the fixation display, but to be prepared for the next sentence. After a variable time delay (randomly varying across trials from 300 to 450 ms), the sentence was presented word by word in white lowercase letters (Courier New, 18-point font) against a black background. The first word and any proper noun were capitalized and the final word of each sentence was followed by a period. Words were presented for 200 ms with a stimulus-onset asynchrony of 500 ms. Following the final word, the screen remained blank for 1 s, after which three question marks appeared, along with the text "1:true" and "5:false." Participants needed to press either "1" or "5" on the number keypad of a keyboard to indicate whether the sentence was true or false (half of the time, the numbers were reversed). If they responded incorrectly, "Wrong Answer" was displayed and if they took more than 3000 ms, "Too slow" was shown. Exactly the same presentation was used for context and target sentences, so that participants were not aware that sentences were presented in pairs.

Following the experiment, the participants were debriefed and a short questionnaire was given to determine if they were at all aware of the purpose of the experiment.

EEG RECORDING AND ANALYSIS

The EEG was recorded from a 64-channel WaveGuard Cap using small sintered Ag/AgCl electrodes connected to an ANT amplifier (ANT, Enschede, Netherlands). An average reference was used. Electrodes were placed according to the 10–20 standard system of the American Electroencephalographic Society over midline sites at Fpz, Fz, FCz, Cz, CPz, Pz, POz, and Oz locations, along with lateral pairs of electrodes over standard sites on frontal (Fp1, Fp2, AF3, AF4, AF7, AF8, F1, F2, F3, F4, F5, F6, F7, and F8), fronto-central (FC1, FC2, FC3, FC4, FC5, and FC6), central (C1, C2, C3, C4, C5, and C6), temporal (FT7, FT8, T7, T8, TP7, and TP8), centro-parietal (CP1, CP2, CP3, CP4, CP5, and CP6), parietal (P1, P2, P3, P4, P5, P6, and P7), and occipital (PO3, PO4, PO5, PO6, PO7, PO8, O1, and O2) positions. Vertical eye movements were monitored via a supra- to sub-orbital bipolar montage. A right to left canthal bipolar montage was used to monitor for horizontal eye movements. The signals were digitized online with a sampling frequency of 512 Hz, with a 0.01–100 Hz band-pass filter. Electrode impedance was maintained below 10 kΩ, mostly under 5 kΩ. The software package ASA was used to analyze the waveforms[2]. The EEG data were screened for eye movements, electrode drifting, amplifier blocking, and EMG artifacts in a critical window ranging from 100 ms before to 800 ms after the onset of the critical word. Trials containing such artifacts were excluded from further analysis, with 88.25% of the epochs being included. Below, we will carry out region-specific analyses of the predicted effects, comparing anterior regions (frontal and fronto-central electrodes, also including midline electrodes Fpz, Fz and temporal electrodes FT7, FT8) versus posterior regions (centro-parietal, parietal, and occipital electrodes, also including midline electrodes CPz, Pz, and Oz). Electrodes TP7, TP8, and POz were not included in the region analyses to balance the number of electrodes in each region.

[2]www.ant-neuro.com

RESULTS

An overview of nine representative electrodes (out of 64 total electrodes) is shown in **Figures 1 and 2**. It is apparent that, in the true sentences (**Figure 1**), the Modality-Match conditions (abbreviated to ModMatch, levels match and mismatch) clearly differ from each other whereas they are visually almost identical for the false sentences (**Figure 2**). For the true sentences, there are clear difference between the magnitude and direction of the effects across the scalp that leads us to include an additional factor Region (levels Anterior, Posterior) in the analyses.

Based on established effects that have been found in the literature and visual inspection of the peaks of the ERP waveforms, we divided the analysis into four time windows: First, a very early window (160–215 ms) to capture the N1–P2 complex. Second, an early window (270–370 ms), which is positioned just before the classic N400 window. Third, a standard N400 window (350–550 ms). Fourth, a late window (500–700 ms) which should capture any late positive shift effects.

A three-way analysis of Modality-Match, Veracity, and Region (anterior, posterior) was carried out for all time windows. This analysis was followed by additional analyses split by Veracity, exploring the existence of a ModMatch effect and/or a Region effect for the subsets of true and false sentences.

FIRST TIME WINDOW: N1–P2 COMPLEX, 160–215 MS

An N1–P2 complex is seen, which is typical for visual word presentation at this rate. We explored whether there was a difference between conditions in this very early time window. In the $2 \times 2 \times 2$ analysis, a significant effect of ModMatch was found and a significant interaction between Veracity, ModMatch, and Region (*F*-values and significance levels are reported in **Table 2** for easy reference; full details are in **Table A1** in Appendix). We explored this interaction by computing simple effects analysis for both levels of the Veracity condition: In the first follow-up analysis (for true sentences only), the factor ModMatch was again significant in this very early window (True-Match mean = 0.145 μV, True-Mismatch mean = 0.063 μV, difference = 0.082, see **Table 2** for significance levels). No significant effects were found in the second analysis (for false sentences).

SECOND TIME WINDOW: EARLY N400-LIKE EFFECTS, 270–370 MS

This time window was chosen after visual inspection of the ERP waveforms to capture the majority of differences that occur over the scalp, in all conditions. Given the theoretical and observed difference between our true and false sentences (see **Figure 3**), separate windows for true and false sentences could have been justified but we felt this would unnecessarily complicate the analysis (we carried out *post hoc* analyses on a number of other time windows but these analyses did not result in a different pattern of significance).

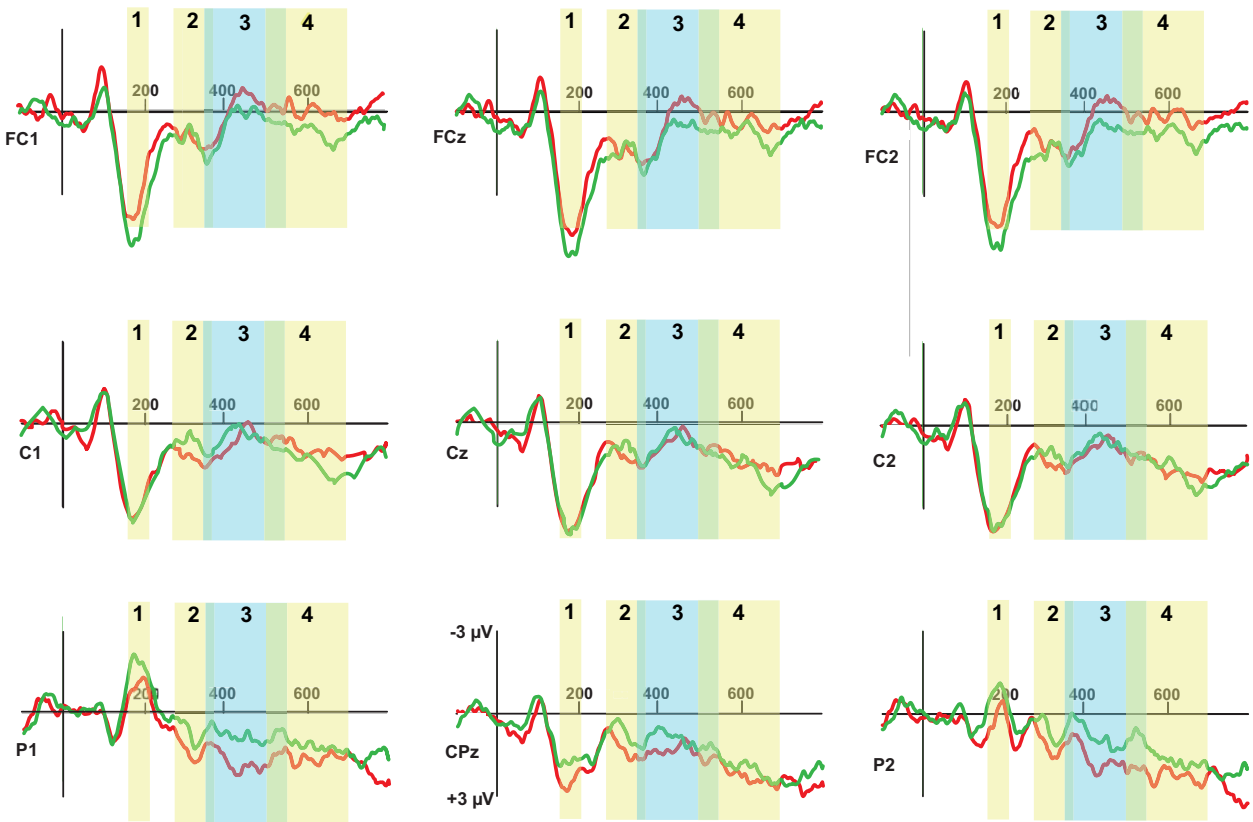


FIGURE 1 | Event related potential traces for true sentences for nine selected sites across the scalp, time locked to onset of the critical word (presented at 0 ms). Negative activation is plotted up. The red lines show the True-Mismatched condition, the green line shows the True-Matched condition. The limits of the early (270–370) and late (500–700) time windows for analysis are indicated.

In the overall 2 × 2 × 2 analysis, we found that there was a significant difference between anterior and posterior electrodes and a significant three-way interaction between Veracity, ModMatch, and Region. We explored this interaction by computing simple effects analysis for both levels of the Veracity condition. In the first follow-up analysis (for true sentences only), the main effect of ModMatch was not significant, but the main effect of Region and the interaction ModMatch × Region were significant (see **Table 2**).

We further explored this two-way interaction for true sentences in a second follow-up and found that for true sentences, a significant ModMatch effect was found both on anterior electrodes [Anterior-True-Match mean = 2.35 μV, Anterior-True-Mismatch mean = 1.52 μV, difference = 0.83, $F(1,9) = 19.615$, MSE = 4.396, $p = 0.002$] and posterior electrodes [Posterior-True-Match mean = −2.00 μV, Posterior-True-Mismatch mean = −1.28 μV, difference = −0.72, $F(1,9) = 19.221$, MSE = 3.498, $p = 0.002$; see **Table 3**; and **Table A2** in Appendix]. Because the ModMatch effect for anterior electrodes has a different polarity than the effect for posterior electrodes, the effects cancel out in the first follow-up analysis, but they are significant in the second follow-up. In the third follow-up analysis (for false sentences only), the factors Region and ModMatch were included but only Region was significant; this effect is not of substantive interest.

THIRD TIME WINDOW: N400 EFFECTS, 350–550 MS

In this classic N400 window, we are probing the tail end of the modality effects reported above for the 270–370 ms time window. A first question is whether the modality-match effects persist; a second question is whether a classic N400 for false versus true sentences will be obtained.

In the initial 2 × 2 × 2 analysis, we obtained a significant three-way interaction between Veracity, ModMatch, and Region. The main effect of Veracity was borderline significant ($F = 4.939$, $p = 0.053$); no other effects reached significance. We explored the three-way interaction by computing simple effects analysis for both levels of the Veracity condition. In the first follow-up analysis (for true sentences only), we found no significant effects; in the second follow-up analysis (for false sentences only), we also found no significant effects.

So although a significant three-way interaction (Veracity, ModMatch, and Region) was found, no significant effects of ModMatch and Region are found when the data are split by Veracity. However, one can also split the data by Modality-Match and look for effects of Veracity and Region. This analysis corresponds to looking for a veracity N400; the results are reported in **Table 3** and **Figure 4**. For modality-matched sentences, no effect of Veracity or Region were found (all $p > 0.25$). However, for modality-mismatched sentences, a significant effect of Veracity and a significant interaction of Veracity and Region were found [Anterior-True-Mismatch

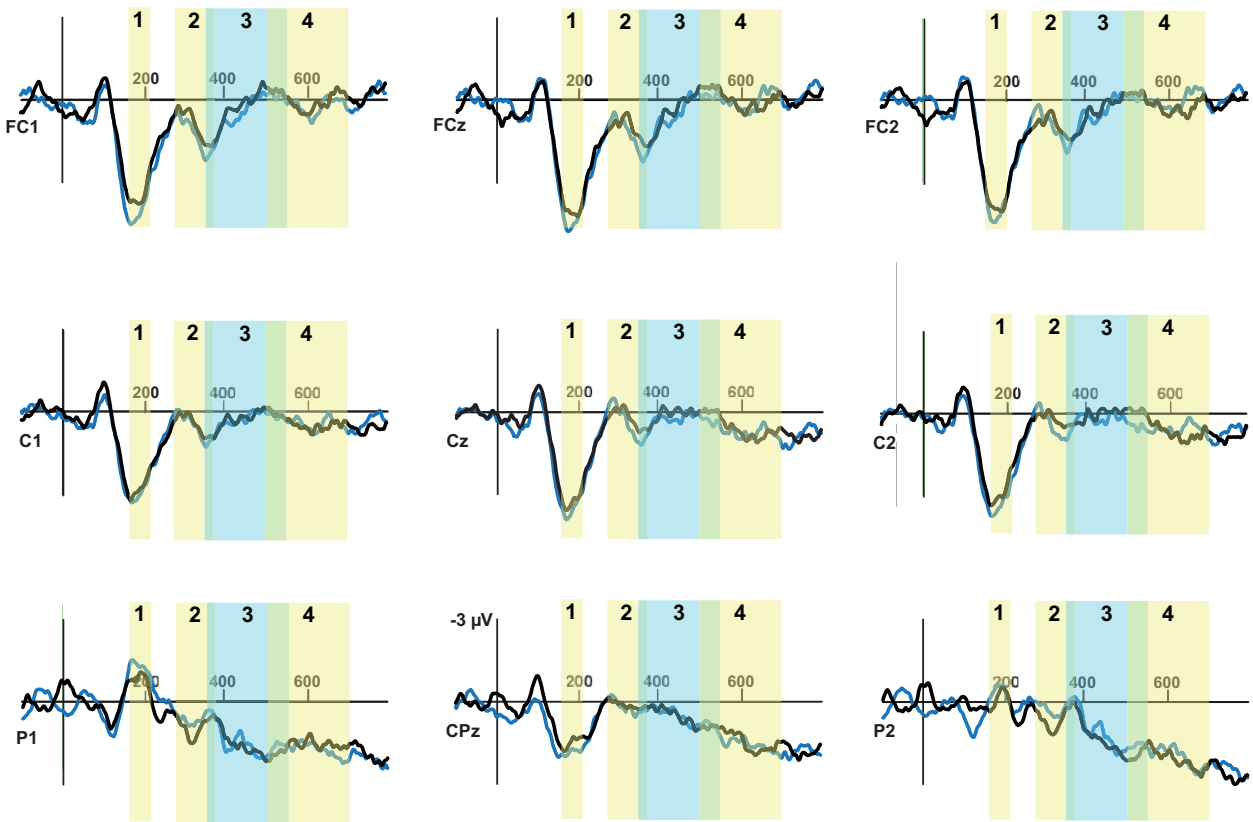


FIGURE 2 | Event related potential traces for false sentences for nine selected sites across the scalp, time locked to onset of the critical word (presented at 0 ms). Negative activation is plotted up. The blue lines show the False-Mismatched condition, the black line shows the False-Matched condition. The limits of the early (270–370) and late (500–700) time windows for analysis are indicated.

Table 2 | Results of the Veracity × ModMatch × Region analysis for each window, with follow-up analyses for the results of ModMatch × Region within levels of Veracity (Vera).

Analysis	Effect	Window 1 (160–215 ms)	Window 2 (270–370 ms)	Window 3 (350–550 ms)	Window 4 (500–700 ms)
		F-value	*F*-value	*F*-value	*F*-value
2 × 2 × 2	Vera	0.126	2.133	4.939	4.066
	ModMatch	14.009**	2.059	0.003	0.057
	Region	0.010	18.376**	0.070	4.784
	Vera × ModMatch	3.300	1.517	1.892	1.986
	Vera × Region	0.091	1.040	1.419	0.305
	ModMatch × Region	0.440	0.487	0.424	1.798
	Vera × ModMatch × Region	8.550*	13.715**	5.399* (c)	3.496
For true	ModMatch	9.932*	2.488	0.755	2.017
	Region	0.042	22.042**	0.454	2.627
	ModMatch × Region	3.921	19.965** (a)	3.805	7.271* (b)
For false	ModMatch	0.301	3.353	0.698	0.230
	Region	0.000	15.176**	0.027	6.875*
	ModMatch × Region	0.421	1.714	2.094	0.491

For each analysis, the significance of the effects of interest are reported by the F-values, with asterisks indicating levels of significance. Full details (MSE, df, p-values) are supplied in the Appendix. Notes: (a–c) refer to simple effects follow-up analyses, see text and ***Table 3****.*
**p < 0.05, **p < 0.01, ***p < 0.001.*

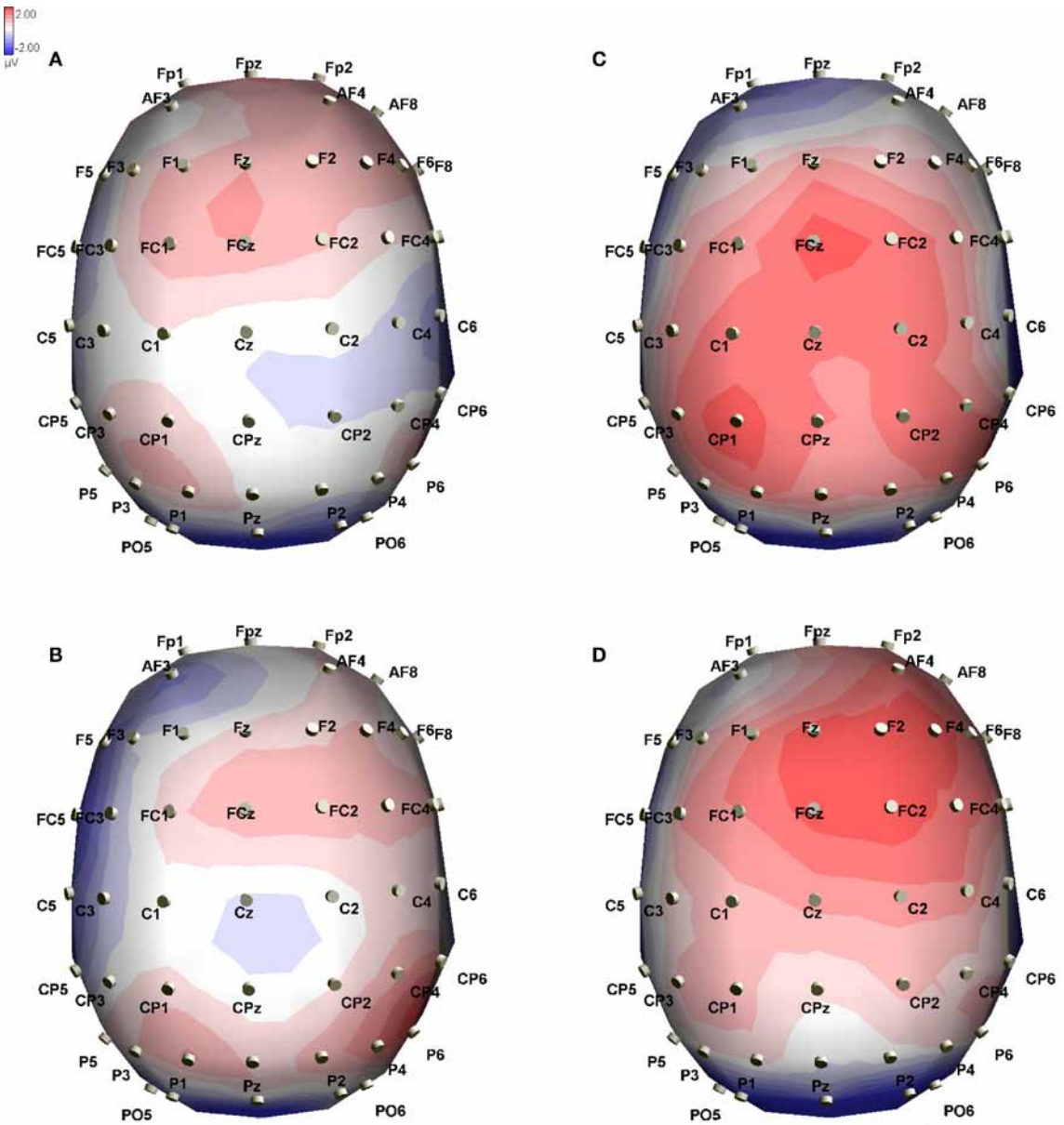


FIGURE 3 | Event related potentials in microvolts across the scalp at 300 ms post onset of the critical word. Blue hues indicate negative potentials, red hues positive potentials. The four conditions shown are False-Mismatch **(A)**, False-Match **(B)**, True-Mismatch **(C)**, and True-Match **(D)**.

mean = −0.48 μV, Anterior-False-Mismatch mean = 0.38 μV, Posterior-True-Mismatch mean = 0.51 μV, Posterior-False-Mismatch mean = −0.17, $F(1,9) = 8.519$, MSE = 0.201, $p = 0.017$; $F(1,9) = 6.358$, MSE = 23.706, $p = 0.033$]. This points at the presence of a classic Veracity N400 effect which is stronger in the anterior region and which is present in the classic N400 window.

FOURTH TIME WINDOW: LATE EFFECTS (500–700 MS)

This late time window was chosen to analyze the late negativity that is apparent for true sentences on the anterior electrodes (see **Figure 1**). In line with the analyses above, the 2 × 2 × 2 was followed by two separate statistical analyses for true and false sentences. No significant effects of interest were found in the 2 × 2 × 2 analysis, although three effects (Veracity; Region; three-way interaction) were marginally significant, $p < 0.10$. For true sentences, a significant interaction between ModMatch and Region was again found, follow-up analyses showed a ModMatch effect in both anterior and posterior regions [Anterior-True-Match mean = −0.28 μV, Anterior-True-Mismatch mean = −1.14 μV, difference = 0.86, Anterior $F(1,9) = 7.628$, MSE = 12.131, $p = 0.022$; Posterior-True-Match mean = 0.23 μV, Posterior-True-Mismatch mean = 0.99 μV, difference = −0.77, Posterior $F(1,9) = 6.803$, MSE = 11.290, $p = 0.028$]. The effect of ModMatch was numerically of the opposite sign in the regions so that the ModMatch main effect was non-significant.

Table 3 | Results of the simple effects follow-up analyses.

Analysis	Subset	Effect	*F*-value
(a) Simple effect of ModMatch within True sentences for Anterior/Posterior regions in window 2	For Anterior	ModMatch	19.615**
	For Posterior	ModMatch	19.221**
(b) Simple effects of ModMatch within True sentences for Anterior/Posterior regions in window 4	For Anterior	ModMatch	7.628*
	For Posterior	ModMatch	6.803*
(c) Effects of Veracity and Region for Matched/Mismatched sentences in window 3	For Mismatched	Veracity	8.519*
		Region	0.379
		Veracity × Region	6.358*
	For Matched	Veracity	0.561
		Region	0.000
		Veracity × Region	1.303

For each analysis, the significance of the effects of interest are reported by the F-values, with asterisks indicating levels of significance. Full details (MSE, df, p-values) are supplied in the Appendix.
**p < 0.05, **p < 0.01, ***p < 0.001.*

This corroborates the findings in the 270–370 ms time window. Also similar to those earlier findings, no substantive significant effects were obtained for false sentences.

REACTION TIME DATA

Participants made a true/false judgment after each sentence was presented. Although there are enough participants for EEG analysis, the analysis of reaction times may lack the power to detect all differences. Note that Pecher et al. (2003) included 32 participants per between subject experimental condition, three times the number of participants in this study. The average reaction time and standard deviations are given in **Table 4**. One should keep in mind that, to keep the task as natural as possible, participants were not required to give a speeded response and this generally leads to large standard deviations. Additionally, to avoid movement artifacts we used a delay response (see Materials and Methods), which may also contribute to more variation. The means for the four conditions are very close to each other and do not differ significantly: In a ModMatch × Veracity ANOVA, we found no significant effects [Veracity $F(1,9) < 1$; ModMatch $F(1,9) < 1$; Veracity × ModMatch $F(1,9) < 1$]. For this analysis, we included all correct responses to target sentences and removed responses faster than 200 ms and slower than 2500 ms. Similarly, accuracy for the conditions was very high and not significantly different: True-Match 94.25% accurate; True-Mismatch 95.25% accurate; False-Match 90% accurate; and False-Mismatch 90% accurate.

DISCUSSION

We conducted an ERP study where participants were exposed to written sentence pairs that either matched or mismatched in modality. We looked for an effect of modality-match in true and in false sentences. Previous research suggests that true and false sentences are processed differently (Fischler et al., 1983). Our results indicate very different effects of modality for true and false sentences, see for example the scalp distributions in **Figure 3**. For true target sentences, we found a large early frontal N400-like effect for true, modality-mismatched pairs ("*A mitten is soft*" – "*A cellar is dark*") compared to true, modality-matched pairs ("*Ham is pink*" – "*A cellar is dark*") in our time window 1 (160–215 ms) and 2 (270–370 ms). In time window 1 (160–215) the anterior negativity effect did not significantly interact with region. However, in time window 2 (270–370 ms) this effect interacted with region such that true, mismatched sentences elicited a larger anterior negativity than true, matched sentences. True, mismatched sentences also elicited a larger positivity on posterior sites compared to true, matched sentences. The effects of modality on the true statements were replicated in a late time window (500–700 ms). As with the early time windows, more negativity is seen for the true, mismatched condition as compared to the true, matched condition across the frontal electrodes. Across the posterior electrodes, more positivity is seen for the true, mismatched condition compared to the true, matched condition. For false target sentences ("*A cellar is light*"), no significant effects of modality were seen at any time window. This is unlikely to be due to a lack of sensitivity, as the pattern for false sentences was numerically reversed compared to the true sentences (false, matched pairs eliciting a non-significant but larger anterior negativity in the waveforms than false, mismatched pairs).

We obtained one additional finding: False, mismatched sentences elicited a classical Veracity N400 in the 350–500 ms window when compared to true, mismatched sentences. This negativity interacted with region, such that it was strongest centro-posteriorly. No effect of veracity was found for the modality-matched sentences.

We will first discuss the relatively early time course of the effect and its distribution. The effect of modality-match on the frontal ERP sites begins in the first time window, as early as 160 ms, and is clearest in the second time window, around 300 ms. The presence of a modality-match effect in our earliest window (165–215 ms) indicates that modality switching is a precursor to and likely to be necessary for meaning integration. The modality-match effect develops further and becomes easily discernible across the scalp in the second time window (270–370 ms). This is the main effect of interest as the polarity of the effect reverses across the scalp, with mismatched pairs eliciting a larger anterior negativity and a larger posterior positivity. Both effects are much earlier than a standard N400 effect which typically begins around 250 ms and peaks around 400 ms (Kutas et al., 2006). In addition, the N400 typically has the strongest negativity on occipital and posterior sites. The distribution of the negativity in window 1 is mostly anterior. The distribution of the negativity in window 2 is also mostly anterior, but in window 2 we see an additional posterior positive distribution, which does not resemble the standard N400 at all.

We are not the first to find an anterior N400-like effect in an embodied context. For example, Van Elk et al. (2010) found an anterior N400 for the preparation of meaningful actions compared to meaningless actions, in a task that required participants to grasp objects. Interestingly, their N400 was largest for the preparation of meaningful actions. Holcomb et al. (1999) found that concrete words elicit a stronger anterior N400 than abstract words, an effect

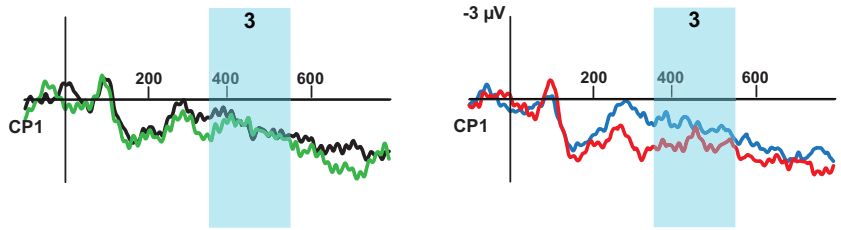


FIGURE 4 | N400 effect for Veracity shown for a representative electrode (CP1). The left panel shows the false (black) and true (green) sentences in matched modality contexts; the right panel shows the false (blue) and true (red) sentences in a mismatched modality context. The standard N400 window from 350 to 550 ms is indicated. This is also our analysis window 3. ERPs are time locked to onset of the critical word (0 ms) and negative activation is plotted up.

Table 4 | Average reaction time (milliseconds) and standard deviation for the true/false judgments on target sentences.

Modality-match	Veracity	
	True	False
Matched	941 (445)	935 (471)
Mismatched	931 (444)	944 (507)

which they coined the concreteness-N400. However, we did not use abstract words in the current study so this particular N400 variant cannot explain our results.

The effect in windows 1 and 2 and 4 are quite similar to the ERP modulation that has been found for pictures and combined sentence–picture stimuli (Barrett and Rugg, 1990; Ganis et al., 1996). In the Ganis et al. (1996) study, the relevant experimental stimuli were sentence fragments that were followed by a picture. The picture was either semantically congruent or incongruent with the sentence semantics up to that point. It was found that, on the frontal electrodes only, incongruent pictures elicited a large negative deflection between 150 and 275 ms compared to the congruent pictures. Barrett and Rugg (1990) found a similar effect, which they called the N300. This effect is similar in time course and distribution to the window 2 effect we report. Ganis et al. (1996) also found that there was a larger anterior N400-like effect for pictures than for control words, and that this effect was reversed on the posterior sites. We found that our window 2 early anterior negativity also reverses on the posterior sites. Lastly, Ganis et al. (1996) report a late congruency effect from 575 to 800 ms whereby the incongruent pictures elicit a negativity at anterior sites and a positivity at posterior sites, which is similar to the findings in our fourth window (500–700 ms).

Ganis et al. (1996) suggest that their findings are specific to pictorial stimuli (see also Barrett and Rugg, 1990). However, we found a very similar effect using only language stimuli. We argue that our specific design, in which all the experimental stimuli refer to a highly salient modal (physical) aspect of an object, induces effects that are comparable in distribution and time course to those that have been obtained with pictures.

This explanation is somewhat consistent with the explanation of the so-called N700 effect proposed by West and Holcomb (2000). The N700 is very similar in time course and scalp distribution to the late effect that was obtained here in window 4. However, the N700 is sensitive to abstractness and shows a stronger anterior negativity and a stronger posterior positivity for concrete words than for abstract words. We do not want to argue that our mismatched stimuli were somehow more concrete but the interesting parallel is that the N700 is stronger in a mental imagery task than in two other tasks (lexical decision, letter spotting). This lead West and Holcomb (2000) to propose that the N700 reflects some image-based type of processing for purely linguistic stimuli.

The similarity between our results using sentences and those from previous work with pictures can best be explained in an embodied view of conceptual representation that uses simulation to arrive at semantic interpretation (Barsalou, 1999). It has been shown that reading action verbs can activate motor cortex (Hauk et al., 2004; Boulenger et al., 2009), presumably because participants were simulating the action. If our participants generated a mental simulation of the properties of the object ("*A cellar is dark*"), this could have produced activation that is very similar to actually seeing the object. Hence, we found effects that are very similar to those that so far have been exclusively found with picture presentation.

An embodied view of concepts would predict that there are no fundamental differences between representations derived from words and those derived from pictures, because each type of stimulus connects to underlying concepts that are grounded in modality-specific representations (in contrast to a dual coding view, such as Paivio's, 1986). Normally, access to concepts happens earlier for pictures (a long-held assumption, e.g., Caramazza et al., 1990; Schriefers et al., 1990) and effects of embodiment and modality are therefore commonly obtained with pictorial stimuli (Stanfield and Zwaan, 2001). In the paradigm used here, modality is primed and access to concepts and modality information is very fast (see also below), which leads to ERP effects that are comparable to those obtained with pictures.

The above is indirect evidence for an embodied view. We also found direct evidence for such a view in the clear ERP differences between true sentences with matched and mismatched modalities. This is also consistent with results from Collins et al. (2011): Using a concept property verification task, modality switching lead to increased amplitude N400 for visual property verifications and a larger late positive complex for auditory verifications. These embodiment effects would not be predicted by models that assume that an abstract propositional representation is necessary

for language comprehension in general and for the sentence verification task specifically (e.g., the Constituent Comparison model; Carpenter and Just, 1975).

The proposed similarity with pictorial stimuli makes it likely (but not necessary) that the modality mismatch effects are stronger for the visual than for the tactile dimension. The idea that different modalities may lead to different modality switch effects in the ERP is supported by Collins et al. (2011), where their results indicate different ERP effects for visual and auditory verifications. Qualitative inspection of the frontal waveforms broadly supports this view, but unlike the Collins et al. (2011) study, the current design does not have the statistical power to investigate this matter quantitatively as there are only 20 items per cell.

MODALITY-MATCH FINDINGS ON TRUE SENTENCES

We offer the following, tentative, explanation for the findings on true sentences. Although the full range of mechanisms underlying the generation of an N400 is still not fully understood, integration processes is one possibility (Brown et al., 2000). Increasing the difficulty of integration will produce a greater (more negative) modulation of the N400. Additionally or alternatively the amplitude may serve as an indicator of the ease or difficulty of retrieving stored conceptual knowledge related to a word. The modulation may be dependent on the stored conceptual representation as well as the preceding contextual information (Kutas et al., 2006). One way to integrate a word with the current discourse is to have a set of possible continuations at hand, which requires some type of prediction. In highly constraining contexts, strong predictive N400 effects have indeed been demonstrated (Van Berkum et al., 2005; see also DeLong et al., 2005). The experiment by Van Berkum et al. (2005) was conducted in Dutch, where adjectives must linguistically agree with nouns. The results showed an N400 effect to adjectives that did not agree with a strongly predicted noun.

In the current experiment, all experimental sentences speak about the visual or tactile modality and a half of the experimental sentences are in the same modality as the preceding sentence. Hence, when a visual context is followed by the target sentence "*the cellar is…*," participants are likely to have "*dark*" as the highly activated top candidate in the set of possible continuations. This prediction is derived from being in a visual context and simulating the visual experience of "*cellar.*" When, in the true, matched condition, the word "*dark*" is read it is immediately integrated in the simulation.

At the onset of the critical word in a true, mismatched sentence, the most highly activated candidate is "*moist,*" because the participant's simulation of the concept cellar is in the tactile modality. When the critical word "*dark,*" comes in, the modality of the simulation has to be changed which leads to a modality switch effect and the observed anterior negativity and posterior positivity in windows 1 and 2 (160–215, 270–370 ms). This switch takes time, as was evidenced by the behavioral results of Pecher et al. (2003).

MODALITY-MATCH FINDINGS ON FALSE SENTENCES

As is clear from the scalp distribution shown in **Figure 3**, a very different pattern of activation was obtained for false sentences than for true sentences. In the false conditions, the target sentence is "*the cellar is light,*" preceded by either a tactile or a visual context sentence. We obtained no differences in windows 1 or 2 between the modality-matched and mismatched conditions because "*light*" is never an expected word: Similar to the explanation above, a visual context would raise an expectation for "*dark*" and the tactile context for "*moist.*" Therefore "*light*" is equally unexpected for both modalities and no difference between modality-match conditions is found.

How does this activation explanation fit with embodied theories of language? Barsalou (1999) discusses falsity only with regards to comparing a sentence to a given situation. For that case, Barsalou (1999) essentially suggests that a simulation of the sentence is made and compared to the scene at hand. If there is a mismatch, then the simulation fails. In our experiment, participants presumably compare the information from the simulation of the false sentence to background knowledge, as there is no scene to compare to. Following Barsalou's (1999) line of reasoning, we would conclude that simulation of the sentence fails.

However this is an incomplete explanation of falsity since it seems that making the simulation of the false sentence should still show a benefit of modality-match. We can explain our results more completely if it is assumed that simulation is based on our prior recent experiences (for example, Glenberg et al., 2009) and that it never fails, but simply takes longer to complete. When trying to simulate "*a cellar is light*" out of context, we are unable to immediately activate the relevant perceptual/action/emotion information because we have limited experience with this. This is not to say we cannot simulate things we have no experience with, but this account would predict that such simulations take longer out of context. The modality switch effect is a small and subtle effect that cannot be observed in this case. In our experiment, that would mean that a false sentence cannot benefit significantly from the preceding modality-match at time windows 1 and 2, but we believe participants still arrive at a simulation of "*the cellar is light.*" After all, this is what is required to understand larger discourse.

The inclusion of the false condition and the findings we obtained for it rule out a semantic relatedness explanation for our true sentence pairs. Under a semantic relatedness explanation, the results we find for true sentences are not due to embodiment but to simple semantic field priming. If, for example, a visual context used a color term and the following target sentence also used a color term, facilitation could be expected. There are independent arguments against this explanation: Pecher et al. (2003) provide empirical evidence against it and semantic priming does not usually last long enough to produce such an effect (see for example McQueen and Cutler, 1998). However, we can also rule out this explanation from our data: The same semantic priming should have occurred in our false sentences as word priming is not sensitive to veracity, but we found no effect for false sentences.

VERACITY FINDINGS: MODALITY-MISMATCHED SENTENCES

Overall, no effect of veracity was found. However, when splitting the data by modality-match versus modality-mismatch, an effect of veracity (greater amplitude N400 for false sentences) was seen in the modality-mismatched conditions. As already suggested, at the onset of the critical word in a false, mismatched sentence, the participant has simulated the concept cellar in the tactile modality and the most highly activated candidate is "*moist.*" When the critical word "*light*" comes in, the modality of the simulation changes and this

causes a delay as outlined above. By the time of the N400 window (350–500 ms), the modality of the simulation may have switched to visual, but the simulation of "*light*" is minimal (assuming that simulation is based on our prior recent experiences) therefore a standard veracity N400 is observed. We tentatively conclude that a delayed minimal simulation leads to the difficulty in integrating "*light*" in the N400 time window.

VERACITY FINDINGS: MODALITY-MATCHED SENTENCES

The situation in the false, matched sentences is slightly more complex. At the onset of the critical word, the participant has simulated the concept Cellar in the visual modality and the most highly activated candidate is "*dark*." When the critical word "*light*" comes in, the modality of the simulation does not need to be changed and a wider simulation can be done, which will arrive at "*light*" as a possible property of cellars: Hence, no Veracity N400 is observed. In other words, although simulation is delayed due to falseness, some benefit occurs from the modality-match that occurs too late to show an effect in time windows 1 and 2, but by the N400 time window, the simulation is rich enough to provide support to the processing of the critical word "*light,*" making it less difficult to integrate. This means the modality context modulates the N400 observed for veracity. We have previously provided evidence showing that the Veracity N400 can be modulated. In Hald et al. (2007), a three sentence context introducing new (supposed) facts about the world significantly reduced the N400 effect to objectively false sentences ("*Venice has many roundabouts*").

CONCLUSION

Our results fit well with the ideas of Zwaan and Madden (2005) and Glenberg and Robertson (1999) in that both sets of authors assume that, during comprehension, we build upon simulations constructed from the previous part of the discourse to integrate the ongoing information with the current simulation (Zwaan calls this process construal). This idea applies most naturally to the comprehension of coherent discourse, but it should also apply to pairs of sentences such as our stimuli. It appears that the construction of a simulation in one modality for the context sentence can aid the simulation of the target sentence if it is in the same modality. A modality switch cost is incurred if the target sentence is of another modality, which leads to larger early anterior ERP effects.

Because the modality of previous sentences helps guide prediction, "*the cellar is…*" proceeded by a tactile context leads to a weaker activation of "*dark*" than when the preceding context is visual. Guided by the tactile context, the system is looking for a tactile property of "*cellar*" and this will lead to a modality switch negativity in our analysis windows 1 and 2 (160–370 ms) for true sentences. We think that the mismatch effect is not observed for false sentences because the comprehension system is engaged in efforts to integrate the false information (see above). Our finding suggests that the simulation process, which is central to embodied language processing, can be predictive (in line with Barsalou, 2009) and that that process will make stronger predictions when there is no modality switch.

AUTHORS' CONTRIBUTION

All authors contributed to the research reported here: Lea A. Hald came up with the idea, contributed to all stages of the research, and wrote and revised the paper; Julie-Ann Marshall prepared the stimuli, tested all participants, and did ERP data preparation; Dirk P. Janssen prepared E-Prime files, assisted in the analysis, and revised the paper; Alan Garnham provided essential guidance throughout.

ACKNOWLEDGMENTS

The authors would like to thank Diane Pecher and her co-authors for making the materials of their study available to us and for helpful discussion about the design of this experiment.

REFERENCES

Baayen, R. H., Piepenbrock, R., and Van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM) [Computer Software]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Barrett, S. E., and Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain Cogn.* 14, 201–212.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1281–1289.

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Brown, C. M., Hagoort, P., and Kutas, M. (2000). "Postlexical integration processes during language comprehension: evidence from brain-imaging research," in *The New Cognitive Neurosciences*, 2nd Edn., ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 881–895.

Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: multiple confusions? *Cogn. Neuropsychol.* 7, 161–189.

Carpenter, P. A., and Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychol. Rev.* 82, 45–73.

Clark, H. H., and Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cogn. Psychol.* 3, 472–517.

Collins, J., Pecher, D., Zeelenberg, R., and Coulson, S. (2011). Modality switching in a property verification task: An ERP study of what happens when candles flicker after high heels click. *Front. Psychol.* 2, 1–10.

DeLong, K., Urbach, T., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.

Fischler, I., Bloom, P., Childers, D. G., Roucos, S. E., and Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology* 20, 400–409.

Ganis, G., Kutas, M., and Sereno, M. I. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *J. Cogn. Neurosci.* 8, 89–106.

Gibbs, R. W. (2005). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. London: Laurence Erlbaum.

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Glenberg, A. M., Becker, R., Klötzer, S., Kolanko, L, Müller, S., and Rinck, M. (2009). Episodic affordances contribute to language comprehension. *Lang. Cogn.* 1, 113–135.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glenberg, A. M., and Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Process.* 28, 1–26.

Glenberg, A. M., and Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *J. Mem. Lang.* 43, 379–401.

Glucksberg, S., Trabasso, T., and Wald, J. (1973). Linguistic structures and mental operations. *Cogn. Psychol.* 5, 338–370.

Goldberg, R., Perfetti, C., and Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* 26, 4917–4921.

Hagoort, P., Hald, L., Bastiaansen, M., and Peterson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441.

Hald, L., Kutas, M., Urbach, T. P., and Parhizkari, B. (2005). "The N400 is not a brainwave. Negation and N400 effects for true and false sentences," in *12th Annual Meeting of the Cognitive Neuroscience Society*, New York, April 2005.

Hald, L., Steenbeek-Planting, E., and Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Res.* 1146, 210–218.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.

Kutas, M., Van Petten, C., and Kluender, R. (2006). "Psycholinguistics electrified II: 1994–2005," in *Handbook of Psycholinguistics*, 2nd Edn, eds M. Traxler and M. A. Gernsbacher (New York: Elsevier), 659–724.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Langacker, R. W. (2001). Discourse in cognitive grammar. *Cogn. Linguist.* 12, 143–188.

Lynott, D., and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* 41, 558–564.

Marques, J. M. (2006). Specialization and semantic organization: evidence from multiple-semantics linked to sensory modalities. *Mem. Cogn.* 34, 60–67.

McQueen, J. M., and Cutler, A. (1998). "Morphology in word recognition," in *Handbook of Morphology*, eds A. Spencer and A. M. Zwicky (Oxford: Blackwell), 406–427.

Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cogn. Psychol.* 1, 242–300.

Nieuwland, M. S., and Kuperberg, G. R. (2008). When the truth isn't too hard to handle: an event-related potential study on the pragmatics of negation. *Psychol. Sci.* 19, 1213–1218.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford: Oxford University Press.

Pecher, D., Van Dantzig, S., Zwaan, R. A., and Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Q. J. Exp. Psychol.* 62, 1108–1114.

Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychol. Sci.* 14, 119–124.

Pecher, D., and Zwaan, R. A. (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge: Cambridge University Press.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Schriefers, H., Meyer, A. S., and Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture–word interference studies. *J. Mem. Lang.* 29, 86–102.

Spence, C., Nicholls, M. E. R., and Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Percept. Psychophys.* 63, 330–336.

Stanfield, R. A., and Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychol. Sci.* 12, 153–156.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journal. Q.* 30, 415–433.

Trabasso, T., Rollins, H., and Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cogn. Psychol.* 2, 239–289.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 443–467.

Van Casteren, M., and Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behav. Res. Methods* 38, 584–589.

Van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.

Van Elk, M., Van Schie, H. T., and Bekkering, H. (2010). The N400-concreteness effect reflects the retrieval of semantic information during the preparation of meaningful actions. *Biol. Psychol.* 85, 134–142.

Van Petten, C. (1995). Words and sentences: event-related brain potential measures. *Psychophysiology* 32, 511–525.

Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cogn.* 18, 380–393.

Vermeulen, N., Niedenthal, P. M., and Luminet, O. (2007). Switching between sensory and affective systems incurs processing costs. *Cogn. Sci.* 31, 183–192.

Wason, P. C. (1980). "The verification task and beyond," in *The Social Foundations of Language and Thought*, ed. D. R. Olson (New York: Norton), 28–46.

West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037.

Zwaan, R. A. (2004). "The immersed experiencer: toward an embodied theory of language comprehension," in *The Psychology of Learning and Motivation*, ed. B. H. Ross (New York: Academic Press), 35–62.

Zwaan, R. A., and Madden, C. J. (2005). "Embodied sentence comprehension," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 224–245.

Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychol. Sci.* 13, 168–171.

Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *J. Exp. Psychol. Gen.* 135, 1–11.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 July 2010; accepted: 03 March 2011; published online: 18 March 2011.

Citation: Hald LA, Marshall J-A, Janssen DP and Garnham A (2011) Switching modalities in a sentence verification task: ERP evidence for embodied language processing. Front. Psychology ***2****:45. doi: 10.3389/fpsyg.2011.00045*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Hald, Marshall, Janssen and Garnham. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

APPENDIX: FULL STATISTICAL REPORTS ON THE ANOVAS

Table A1 | Full results of the Veracity × ModMatch × Region analysis for each window, with follow-up analyses for the results of ModMatch × Region within levels of Veracity.

Analysis	Effect	Window 1 (160–215 ms)	Window 2 (270–370 ms)	Window 3 (350–550 ms)	Window 4 (500–700 ms)
2 × 2 × 2	Vera	$F(1,9) = 0.126$, MSE = 0.544, $p = 0.731$	$F(1,9) = 2.133$, MSE = 0.431, $p = 0.178$	$F(1,9) = 4.939$, MSE = 0.301, $p = 0.053$	$F(1,9) = 4.066$, MSE = 0.592, $p = 0.075$
	ModMatch	$F(1,9) = 14.009$, MSE = 0.075, $p = 0.005$	$F(1,9) = 2.059$, MSE = 0.180, $p = 0.185$	$F(1,9) = 0.003$, MSE = 0.303, $p = 0.955$	$F(1,9) = 0.057$, MSE = 0.403, $p = 0.817$
	Region	$F(1,9) = 0.010$, MSE = 218.497, $p = 0.923$	$F(1,9) = 18.376$, MSE = 369.775, $p = 0.002$	$F(1,9) = 0.070$, MSE = 124.872, $p = 0.797$	$F(1,9) = 4.784$, MSE = 228.636, $p = 0.057$
	Vera × ModMatch	$F(1,9) = 3.300$, MSE = 0.204, $p = 0.103$	$F(1,9) = 1.517$, MSE = 0.150, $p = 0.249$	$F(1,9) = 1.892$, MSE = 0.210, $p = 0.202$	$F(1,9) = 1.986$, MSE = 0.250, $p = 0.192$
	Vera × Region	$F(1,9) = 0.091$, MSE = 22.114, $p = 0.770$	$F(1,9) = 1.040$, MSE = 15.214, $p = 0.335$	$F(1,9) = 1.419$, MSE = 16.905, $p = 0.264$	$F(1,9) = .305$, MSE = 35.701, $p = 0.594$
	ModMatch × Region	$F(1,9) = 0.440$, MSE = 31.207, $p = 0.524$	$F(1,9) = 0.487$, MSE = 27.203, $p = 0.503$	$F(1,9) = 0.424$, MSE = 24.636, $p = 0.531$	$F(1,9) = 1.798$, MSE = 20.344, $p = 0.213$
	Vera × ModMatch × Region	$F(1,9) = 8.550$, MSE = 7.446, $p = 0.017$	$F(1,9) = 13.715$, MSE = 13.975, $p = 0.005$	$F(1,9) = 5.399$, MSE = 28.781, $p = 0.045$ (c)	$F(1,9) = 3.496$, MSE = 43.514, $p = 0.094$
For True	ModMatch	$F(1,9) = 9.932$, MSE = 0.172, $p = 0.012$	$F(1,9) = 2.488$, MSE = 0.237, $p = 0.149$	$F(1,9) = 0.755$, MSE = 0.290, $p = 0.408$	$F(1,9) = 2.017$, MSE = 0.182, $p = 0.189$
	Region	$F(1,9) = 0.042$, MSE = 97.613, $p = 0.841$	$F(1,9) = 22.042$, MSE = 139.620, $p = 0.001$	$F(1,9) = 0.454$, MSE = 67.963, $p = 0.517$	$F(1,9) = 2.627$, MSE = 168.700, $p = 0.140$
	ModMatch × Region	$F(1,9) = 3.921$, MSE = 17.413, $p = 0.079$	$F(1,9) = 19.965$, MSE = 7.657, $p = 0.002$ (a)	$F(1,9) = 3.805$, MSE = 32.378, $p = 0.083$	$F(1,9) = 7.271$, MSE = 23.239, $p = 0.025$ (b)
For False	ModMatch	$F(1,9) = 0.301$, MSE = 0.079, $p = 0.596$	$F(1,9) = 3.353$, MSE = 0.056, $p = 0.100$	$F(1,9) = 0.698$, MSE = 0.178, $p = 0.425$	$F(1,9) = 0.230$, MSE = 0.529, $p = 0.643$
	Region	$F(1,9) = 0.000$, MSE = 146.185, $p = 0.999$	$F(1,9) = 15.176$, MSE = 248.720, $p = 0.004$	$F(1,9) = 0.027$, MSE = 75.609, $p = 0.872$	$F(1,9) = 6.875$, MSE = 95.893, $p = 0.028$
	ModMatch × Region	$F(1,9) = 0.421$, MSE = 21.812, $p = 0.533$	$F(1,9) = 1.714$, MSE = 33.282, $p = 0.223$	$F(1,9) = 2.094$, MSE = 20.794, $p = 0.182$	$F(1,9) = 0.491$, MSE = 40.976, $p = 0.501$

Table A2 | Full results of the simple effects follow-up analyses.

Analysis	Effect	Statistics
(a) Simple effect of ModMatch within True sentences for Anterior/Posterior regions in window 2		
Anterior	ModMatch	$F(1,9) = 19.615$, MSE = 4.396, $p = 0.002$
Posterior	ModMatch	$F(1,9) = 19.221$, MSE = 3.498, $p = 0.002$
(b) Simple effects of ModMatch within True sentences for Anterior/Posterior regions in window 4		
Anterior	ModMatch	$F(1,9) = 7.628$, MSE = 12.131, $p = 0.022$
Posterior	ModMatch	$F(1,9) = 6.803$, MSE = 11.290, $p = 0.028$
(c) Effects of Veracity and Region for Matched/Mismatched sentences in window 3		
Mismatched	Veracity	$F(1,9) = 8.519$, MSE = 0.201, $p = 0.017$
	Region	$F(1,9) = 0.379$, MSE = 50.505, $p = 0.553$
	Vera × Region	$F(1,9) = 6.358$, MSE = 23.706, $p = 0.033$
Matched	Veracity	$F(1,9) = 0.561$, MSE = 0.310, $p = 0.473$
	Region	$F(1,9) = 0.000$, MSE = 99.002, $p = 0.985$
	Vera × Region	$F(1,9) = 1.303$, MSE = 21.979, $p = 0.283$

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 08 February 2011
doi: 10.3389/fpsyg.2011.00010

Modality switching in a property verification task: an ERP study of what happens when candles flicker after high heels click

Jennifer Collins[1]*, Diane Pecher[2], René Zeelenberg[2] and Seana Coulson[1]

[1] *Brain and Cognition Lab, Department of Cognitive Science, University of California San Diego, San Diego, CA, USA*
[2] *Memory Lab, Department of Psychology, Faculty of Social Sciences, Erasmus University Rotterdam, Rotterdam, Netherlands*

Edited by:
Anna M. Borghi, University of Bologna and Institute of Cognitive Sciences and Technologies, Italy

Reviewed by:
Lea Hald, Canterbury Christ Church University, UK
Frederico Marques, Lisbon University, Portugal

***Correspondence:**
Jennifer Collins, Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA.
e-mail: jcollins@cogsci.ucsd.edu

The perceptual modalities associated with property words, such as *flicker* or *click*, have previously been demonstrated to affect subsequent property verification judgments (Pecher et al., 2003). Known as the *conceptual modality switch effect*, this finding supports the claim that brain systems for perception and action help subserve the representation of concepts. The present study addressed the cognitive and neural substrate of this effect by recording event-related potentials (ERPs) as participants performed a property verification task with visual or auditory properties in key trials. We found that for visual property verifications, modality switching was associated with an increased amplitude N400. For auditory verifications, switching led to a larger late positive complex. Observed ERP effects of modality switching suggest property words access perceptual brain systems. Moreover, the timing and pattern of the effects suggest perceptual systems impact the decision-making stage in the verification of auditory properties, and the semantic stage in the verification of visual properties.

Keywords: property verification, modality switch, ERPs, grounded cognition

INTRODUCTION

Over the past decade, cognitive scientists have gradually moved away from the assumption that concepts are *symbolic*, that is, arbitrarily related to the things they represent, and *amodal*, or independent of any sensory modality (see Murphy, 2002 for a review of traditional models), and have increasingly come to embrace an embodied or grounded approach. These more recent accounts have focused on how concepts are grounded in our perception of, and interaction with, the physical and social world, and stressed their modal characteristics (see Barsalou, 2008 for a review). The perceptual symbol system hypothesis, for example, is that conceptual knowledge involves schematized perceptual and motor representations involved in one's prior experience with the concept's referent (Barsalou, 1999). On this account, a concept is a sensorimotor simulation involving the partial reactivation of brain regions that participated in the acquisition of that concept. For example, the concept of a dog is a simulation involving brain areas that represent one's visual, auditory, tactile, olfactory, affective, and motoric experiences with dogs. Importantly, simulations are not holistic records of experience, but can be flexibly adapted to the current context and task (Barsalou et al., 2003).

The use of visual mental images for ostensibly conceptual tasks has been demonstrated with the property verification task, in which participants are asked whether or not a particular property (e.g., *has-a-head*) is true for a given concept (e.g., *CAT*). The perceptual symbol system hypothesis suggests that accessing conceptual knowledge involves the activation of associated visual images, and thus predicts a systematic relationship between the difficulty of property verification and that of activating the relevant visual image. Consistent with this prediction, Solomon and Barsalou (2004) found that participants took less time to verify visually large properties of a concept (e.g., that a *CAT* has a *head*) than visually smaller properties of the same concept (e.g., that a *CAT* has a *paw*). The fact that performance on this conceptual task was modulated in a similar way as performance on a visual imagery task was argued to implicate the importance of visual processes in conceptual representations.

Moreover, a functional magnetic resonance imaging (fMRI) study in which participants performed the property verification task employed by Solomon and Barsalou (2004) revealed activation in the left fusiform gyrus, an area important for object recognition and visual imagery (Kan et al., 2003). The recruitment of perceptual brain areas for the conceptual task of property verification is consistent with the perceptual symbol system hypothesis, and is also in keeping with other fMRI studies in which conceptual tasks have activated brain regions used to perceive the concepts' referents (Goldberg et al., 2006; Martin, 2007; Simmons et al., 2007).

MODALITY SWITCH EFFECTS

Although the bulk of empirical support for the perceptual symbol system hypothesis concerns the involvement of specifically visual representations, the hypothesis is, in fact, farther ranging, extending to the full multimodal characteristic of human experience. The concept of a lemon, for example, should not only represent its color, but also its taste, its smell, and its texture. Moreover, because simulations involve the coordination of information from multiple perceptual modalities, the perceptual symbol system hypothesis predicts that conceptual operations will display many of the same properties as complex perceptual operations, and be subject to similar constraints. Accordingly, Pecher et al. (2003) tested whether a property verification task using properties

from several modalities, including vision, audition, and touch, was modulated by factors known to affect perceptual detection tasks with stimuli from multiple modalities.

In particular, Pecher et al. (2003) focused on the modality switch effect, a phenomenon observed in the literature on perceptual processing. In a study designed to assess cross-modal effects of spatial attention, Spence et al. (2001) asked participants to detect brief auditory, visual, or tactile targets at peripheral locations. The *modality switch effect* is the finding that reaction times were longer for all stimulus types when they were preceded by a stimulus from a different modality than from the same modality, and has been interpreted as an exogenously driven attentional cost for the switch trials (Spence et al., 2001; Rodway, 2005).

Pecher et al. (2003) reasoned that if conceptual processing relies on perceptual systems, the well-known cost for successive trials from different modalities in perceptual tasks might also be expected to occur on a property verification task employing properties from multiple modalities. In their conceptual analog to the modality switch studies, Pecher et al. (2003) asked participants to determine whether a property (e.g., *yellow* or *sour*) applied to the preceding concept (e.g., *LEMON* or *MOUSE*). The manipulation of interest was whether a pair of trials was from the same modality (*LEAVES–rustling* followed by *BLENDER–loud*) or different modalities (*CRANBERRIES–tart* followed by *BLENDER–loud*). As predicted by the perceptual symbol system hypothesis, Pecher et al. (2003) found longer reaction times for the second trial in a pair of different modality (switch) trials than for the second trial in a pair of same modality (no-switch) trials, the conceptual modality switch effect.

Variations on the conceptual modality switch paradigm have shown that results cannot be attributed to alternative explanations, such as word association (Pecher et al., 2003), or category overlap (Marques, 2006). The generality of the effect is supported by the demonstration of a similar switch effect on a property verification task using perceptual and emotional attributes (Vermeulen et al., 2007). Importantly, property verification has also been shown to be speeded by the presentation of a perceptual stimulus from the same modality relative to one from a different modality (van Dantzig et al., 2008). The finding that the verification of visual features of a concept is faster after the perceptual detection of visual than auditory or tactile stimuli provides strong support for the suggestion that the conceptual task of property verification recruits perceptual processing resources, as opposed to an amodal re-representation of perceptual information.

Another direction this research has taken has been to investigate the neural substrate of modality specific concepts using cognitive neuroscience methods. Goldberg et al. (2006) recorded participants' brain activity using fMRI while they engaged in a property verification task. The experiment used a design in which different blocks required participants to make decisions about properties referring to different modalities – visual, auditory, tactile, and gustatory. The brain regions uniquely activated for each property category were regions related to the perception of stimuli in the different domains. These results are particularly important given that reaction time results for similar conceptual tasks have not distinguished between responses to properties of different modalities (Pecher et al., 2009).

Neuroimaging data thus provide compelling evidence that conceptual tasks are associated with the activation of perceptual brain regions. At issue, however, is whether perceptual systems play a central or a peripheral role in cognition (Barsalou, 2008). Perceptual activations might, for example, be an artifact of the blocked design used by Goldberg et al. (2006). Alternatively, perceptual activations might reflect top-down processing initiated only after the meaning of the property words has been activated.

THE PRESENT STUDY

The present study addressed the cognitive and neural basis of the conceptual modality switch effect by recording event-related potentials (ERPs) as participants made property verification judgments about the visual and auditory properties of objects. ERPs are patterned voltage changes in the on-going electroencephalogram (EEG) that are time-locked to classes of specific processing events. As a continuous, real-time measure of brain activity, ERPs are well-suited for investigating the neural processes relevant to the conceptual modality switch effect allowing us to better understand when a perceptual system is accessed by a related concept. In particular, the present study was designed to address whether the modality manipulation affected ERP components associated with the visual processing of property terms, such as the N1 and P2, semantic processing of property terms, such as the N400, or their task-relevant categorization as typical properties of the relevant concept, indexed by the P3 or late positive complex (LPC).

We used stimuli similar to those employed by Pecher et al. (2003), but included only visual (*CANDLES–flicker*) and auditory (*NEWSPAPERS–rustle*) trials in our critical conditions. This reduction in variation was important in order to have enough trials in critical conditions for averaging ERPs. Participants' task was to determine whether or not the property applied to the concept. The correct response on all experimental trials was "true," and a large number of filler trials requiring a "false" response (e.g., *COCKROACHES–ablaze*) were included to discourage the development of a particular response bias. A subset of false filler trials included properties and concepts that were lexically associated (e.g., *STRAWBERRIES–cream*) and were intended to discourage the use of word association strategies (Solomon and Barsalou, 2004). The critical manipulation concerned whether the target concept–property trial (e.g., *NEWSPAPERS–rustle*) was preceded by a prime concept–property trial from the same modality (e.g., *HIGH HEELS–click*), or a different modality (e.g., *CHERRIES–ruby*). Half of the experimental trials involved visual and half auditory properties, and were equally likely to follow a concept–property trial from the same modality (a visual property following a visual property, or an auditory property following an auditory property, viz. no-switch trials) as one from a different modality (visual–auditory or auditory–visual, viz. switch trials).

The primary goal of the study was thus to identify electrophysiological correlates of the conceptual modality switch effect in order to determine which stage or stages of processing the switch manipulation would modulate. If concepts automatically engage early sensory processing, then the mention of a visual property such as "*flicker*" could modulate the actual perception of visual word forms presented shortly afterward. The converse of this type of effect was found behaviorally by van Dantzig et al. (2008).

Low-level perceptual engagement of this sort would be indexed by modulation of visual ERP components to the word form such as the N1, and P2.

Alternatively, perceptual access might be part of an extended, standard semantic network that subserves the representation of concepts. The N400, a negative-going wave evident between 200 and 700 ms after the visual presentation of a word, was of particular interest due to its association with the processing of meaningful events. The N400 is elicited by words in all modalities, whether written, spoken, or signed (Holcomb and Neville, 1990). Words preceded by semantically related words elicit smaller amplitude N400 than do words preceded by unrelated words, the N400 priming effect (Bentin, 1987; Holcomb, 1988; Smith and Halgren, 1989). The N400 is also sensitive to contextual factors related to meaning at the sentence and text level. In general, N400 amplitude varies inversely with the predictability of the target word: N400s are large for unexpected items, smaller for words of intermediate predictability, and are barely detectable for highly predictable words (Kutas and Hillyard, 1984; see Kutas and Federmeier, 2011 for a review).

Yet another possibility is that the conceptual modality switch effect is attributable to decision processes specifically induced by the property verification task. If this is the case, we would expect the conceptual modality switch paradigm to modulate later, decision-related components such as the P3, or LPC. This family of ERP components is generally thought to index the updating of mental representations modulated by processes such as allocation of attention and task-dependent target classification (Polich, 2007).

A secondary goal of the study was to test whether property terms from different modalities (viz. visual versus auditory) would activate different modality specific brain areas as found in related fMRI studies (e.g., Goldberg et al., 2006). Although the spatial resolution of the EEG is limited, such differences might be detectable as subtle differences in the scalp topography of ERPs to visual versus auditory properties. An interaction between the modality factor in our analysis and electrode site would suggest that non-overlapping neural generators underlie the brain response to auditory and visual properties, viz. that the exact same brain regions do not subserve the processing of visual and auditory properties (Urbach and Kutas, 2002). More generally, differences between the modality switch process in the visual and auditory domains would connect this paradigm with Pecher et al.'s (2003) claim that the conceptual modality switch effect results from switching between different perceptual networks.

As a time-sensitive measure of online cognitive processing, ERPs can provide more information about whether the real-time processing of property terms involves the recruitment of perceptual brain areas during early perceptual processing, during semantic processing, or whether the switch effect would be evident only later, during decision-related stimulus processing. Given Barsalou's (1999) claim that sensorimotor simulations comprise an intrinsic component of concept meaning, we hypothesized that the facilitative impact of a same modality prime would involve the semantic processing of the target trial, and thus would modulate the amplitude of the N400 component of the ERP. In particular, we predicted that no-switch trials would elicit reduced amplitude N400 relative to switch trials.

MATERIALS AND METHODS

The protocol for this study was approved by the University of California, San Diego Social and Behavioral Science Institutional Review Board. As such, informed consent was obtained from all participants prior to their enrollment.

PARTICIPANTS

Twenty undergraduates from the UCSD community (13 women) participated as part of a course requirement. Data from six additional participants were not included in the analysis due to the presence of an excessive number of artifacts (greater than 30% of trials in a critical condition). All participants were between the ages of 18 and 40 years old. As reported in a screening questionnaire, all participants had normal vision, and none had any history of neurological or psychiatric disorders within the previous 10 years.

MATERIALS

Each trial in the study consisted of a concept–property combination such as *HIGH HEELS* (concept) and *click* (property). Experimental trials involved 48 visual properties (such as *flicker*), and 48 auditory properties (such as *click*). Each property was presented with two different concepts for a total of 192 experimental trials; all properties were repeated once over the course of the experiment, while all concepts were unique. Half (96) of the concept–property combinations served as prime trials (48 involving auditory properties, and 48 involving visual properties), and half (96) served as target trials (48 involving auditory properties, and 48 involving visual ones). Experimental trials were presented in pairs, so that a prime trial was immediately followed by a target trial that was either from the same modality (no-switch condition), or the other modality (switch condition). Materials were thus comprised of 96 trial pairs in which the modality of the probe property was crossed with the modality switch dimension (24 auditory prime/auditory target, 24 visual prime/auditory target, 24 visual prime/visual target, and 24 auditory prime/visual target pairs). Apart from the modality manipulation the prime–target pairs were unrelated. All properties in experimental trials were valid for their concept so that the correct response on the property verification task was always "true."

Materials also included 384 filler trials, 96 of which involved auditory properties that did not pertain to their concept (e.g., *LOBSTERS–bark*) and 96 of which involved visual properties also eliciting *false* responses (e.g., *LAWNS–scarlet*). These two sets were included so that participants could not strategically respond *true* to any trial involving an auditory or visual property. Another 96 filler trials involved tactile properties, half of which were valid for their concept (e.g., *CAVES–damp*), and half of which were not (e.g., *TOASTERS–damp*; one response for each property repetition). The final 96 filler trials were lexical associates (e.g., *BUFFALOS–winged*), included to discourage participants from shallow processing strategies relying on word association (as in Solomon and Barsalou, 2004). Half of the associated trials were *true* trials, and half were *false* trials. Of the 384 filler trials, the correct response on the property verification task was *true* for 96, and *false* for 288. When including the 192 experimental trials as well, the correct response on the task was thus *true* for half of the total trials, and *false* for the other. Moreover, even though the experimental trials always involved two *true* responses in a row (viz. one for the prime, and

one for the probe), the inclusion of filler trials guaranteed that a correct *true* response was equally likely to be followed by a correct *false* response as by another correct *true*.

Two lists were employed so that any given target property occurred once in a switch trial (that is, following a prime from the other modality), and once in a no-switch trial (that is, following a prime from the same modality). Two variants of each were created by swapping the first and second half of each list. In this way, each concept–property combination was presented equally often in the first and second half of the experiment.

PROCEDURE

Participants were seated in a dim, sound attenuating chamber approximately 50 inches from a 17-inch computer monitor. They read a standard set of instructions telling them to "read the entity (such as objects, people, animals, etc.) and property words, … and respond *true* if the property was typical or often possible for the entity, and *false* if the property was highly unusual for the entity." They read several examples and were presented with practice trials on which they received feedback. Participants were told, "after you read the property, decide as quickly and accurately as possible whether the property is true or false," but no explicit feedback was given on either of these dimensions during the course of the experiment.

The timing of events in the experimental paradigm is presented in **Figure 1**. Each trial began with the presentation of a white fixation cross for 250 ms. The inter-stimulus interval (ISI) between the fixation cross and the concept was randomly varied with 50 ms steps between 200 and 400 ms. The concept appeared on the center of the screen in capital letters for 150 ms followed by a 250 ms ISI and the property word in lowercase letters for 200 ms. In order to limit the potential for eye-movement artifacts in our EEG signal we chose to centrally present both concepts and properties and eliminate the phrase "can be" from the original paradigm which is not a necessary aspect of the conceptual modality switching procedure (e.g., Pecher et al., 2004). All type was in white font presented on a black background. Participants had 2600 ms to make their decision and prepare for the next trial. Responses were made via a button press in which a right hand response indicated *true* and a left hand response indicated *false*. Trials were presented in ten blocks, each lasting about 3.5 min with time in between for participants to rest. The first block began with eight practice trials that were not included in the analysis. All blocks had 60 trials except for the last block which had 44 trials.

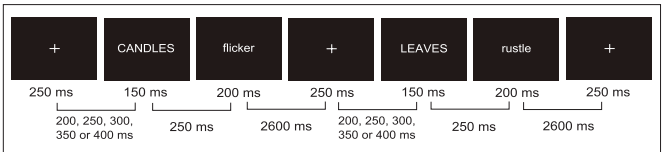


FIGURE 1 | Participants saw pairs of words – a concept (in capitals) followed by a property (in lowercase) – after which they would make a *true/ false* judgment during a 2600-ms blank screen. Both examples shown in the figure should elicit *true* responses because the properties are typical of their respective concept. The critical manipulation in this experiment is the perceptual modalities evoked by subsequent trials. In this example the first is a visual decision, the second is an auditory decision and together they make up an item in the "switch" condition.

EEG RECORDING AND ANALYSIS

Participants' EEG was recorded with tin electrodes mounted in an electrode cap with 29 scalp sites (see **Figure 2**). Scalp electrodes were referenced online to the left mastoid, and subsequently re-referenced to the average of the left and right mastoid electrodes. Blinks were monitored with an electrode below the right eye. Horizontal eye movements were monitored via a bipolar derivation of electrodes placed over the outer canthi. EEG was recorded and amplified with an SA Instruments isolated bioelectric amplifier at a bandpass of 0.1 and 100 Hz, digitized online at 250 Hz, and stored on a hard drive for subsequent averaging. The EEG was later monitored offline for blinks and eye movements which were rejected manually. ERPs were time-locked to the onset of property words on probe trials.

For each time interval of interest we performed a $2 \times 2 \times 29$ repeated measures ANOVA with the factors switch (switch/no-switch), target property modality (visual/auditory), and electrode site (29 levels). The dependent measure was the mean amplitude within the time intervals of interest. In cases where the overall analysis revealed a significant interaction between modality switch and property modality, follow-up analyses were conducted separately for the visual and auditory properties. Follow-up analyses thus involved factors switch (switch/no-switch) and electrode site (29 levels). The Huynh–Feldt correction was applied where relevant. For clarity, however, we report the original degrees of freedom.

RESULTS

BEHAVIORAL RESULTS

Analysis of reaction times failed to reveal any statistically significant effects in a 2×2 ANOVA testing switch (switch/no-switch) and modality (visual/auditory; all Fs < 2). Given that behavioral studies of this phenomenon typically do not test the modalities separately and employ data from at least 60 participants (cf. the 20 employed in the present study), these null results are likely due to a lack of power. The pattern of reaction times was in the expected direction for the visual properties (switch = 902 ms, SD = 152 ms;

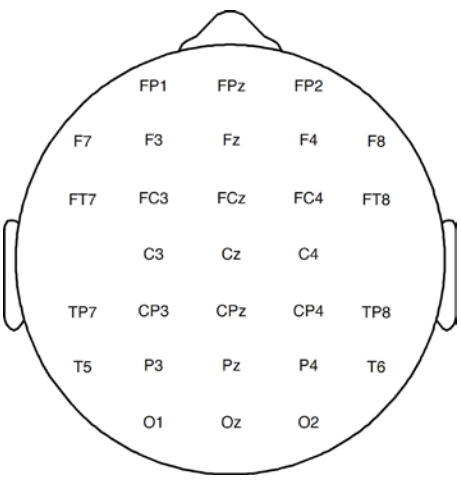


FIGURE 2 | Relative placement of 29 scalp electrodes at which EEG was recorded.

no-switch = 891 ms, SD = 155 ms) but not for the auditory properties (switch = 908 ms, SD = 148 ms; no-switch = 917 ms, SD = 163 ms).

Analysis of accuracy rates revealed a main effect of modality type with auditory properties showing worse accuracy than visual properties [$F(1,19) = 13.81$, $p < 0.01$]. There were no significant effects of switch condition for the visual (switch = 0.92, SD = 0.07; no-switch = 0.94, SD = 0.05) nor auditory properties (switch = 0.86, SD = 0.09; no-switch = 0.87, SD = 0.09; Fs < 1) but both modality types showed slightly worse performance for the switch condition.

ERP RESULTS

Probe properties elicited ERPs typical of visually presented words, an N1–P2 complex followed by the N400 and a LPC. The switch manipulation did not affect ERP waveforms in the early 100–200 ms interval. The switch manipulation modulated the amplitude of the N400 (measured 200–500 ms post-stimulus) and the LPC (measured 500–800 ms), but did so differently for visual and auditory properties. Whereas visual properties elicited a larger N400 for switch than no-switch trials, auditory properties elicited a larger LPC for the same comparison.

100–200 ms

Analysis of ERPs measured 100–200 ms after stimulus onset did not show any differences for analyses of switch effects (all Fs < 1). Nor did it reveal differences based on the modality elicited by the properties (all Fs < 1.4).

200–500 ms

Overall analysis of ERPs measured 200–500 ms after stimulus onset revealed a significant interaction between the switch and the modality factors [$F(1,19) = 4.61$, $p < 0.05$, MSE = 147.25]. Follow-up analyses of each individual modality revealed no effects in the auditory modality (Fs < 1; auditory switch = 5.08 μV, auditory no-switch = 4.76 μV), but a reliable switch effect in the visual one [$F(1,19) = 4.93$, $p < 0.05$, MSE = 135.52]. The latter reflects the slightly more negative (0.7 μV) ERPs elicited in the visual switch (4.53 μV) than the visual no-switch (5.21 μV) condition (**Figure 3**). Although this difference showed up as a main effect in the analysis, visual inspection suggests it was largest over centro-parietal sites characteristic of the classic N400 effect (**Figure 4**).

500–800 ms

Overall analysis of ERPs measured 500–800 ms after stimulus onset revealed a significant interaction between the modality and the switch factors [$F(1,19) = 5.27$, $p < 0.05$, MSE = 162.78], as well as a marginal interaction between modality and electrode site [$F(28,532) = 1.81$, $p = 0.10$, ε = 0.20, MSE = 3.49]. Follow-up analyses suggested the interaction between modality and switch results from a positive-going switch effect evident only for auditory properties. Separate analysis of the visual modality revealed no effect of the switch factor, either as a main effect ($F < 1$; visual switch = 6.00 μV; visual no-switch = 6.22 μV), or in interaction with electrode site ($F < 1$). Separate analysis of the auditory modality suggested a trend for switch trials to elicit a slightly larger positivity (switch = 6.70 μV) than did no-switch trials [5.86 μV; $F(1,19) = 3.02$, $p = 0.098$, MSE = 201.31; see **Figures 5 and 6**].

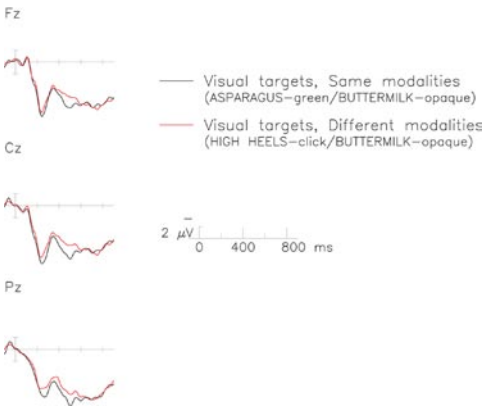


FIGURE 3 | The N400 elicited by visual property verification targets in the switch (red) and no-switch (black) conditions. Each graph represents data recorded from a midline electrode over frontal (top), central (middle), and parietal (bottom) regions of the scalp. Time is plotted on the *x*-axis against voltage on the *y*-axis. By convention, negative polarity is plotted upward.

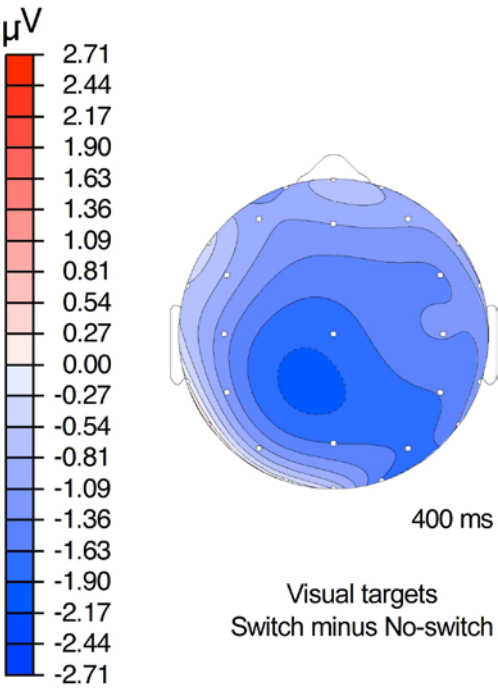


FIGURE 4 | Topography of the switch effect for visual property verification targets.

We also followed up on the marginal interaction between modality and electrode site as the possible topographic differences were of interest to our question of access to underlying perceptual modalities by property words. We tested midline, medial, and lateral sites separately. Our midline test included factors of modality (visual, auditory) and anteriority (seven midline electrodes, see **Figure 2**). This test revealed a marginal interaction between modality and anteriority [$F(6,114) = 2.67$, $p = 0.057$, ε = 0.45 MSE = 1.35]. Our test of the medial sites



FIGURE 5 | The late positive complex (LPC) to auditory targets in the switch (red) relative to the no-switch (black) conditions. Each graph represents data recorded from a midline electrode over frontal (top), central (middle), and parietal (bottom) regions of the scalp. Time is plotted on the *x*-axis against voltage on the *y*-axis and negative polarity is plotted upward.

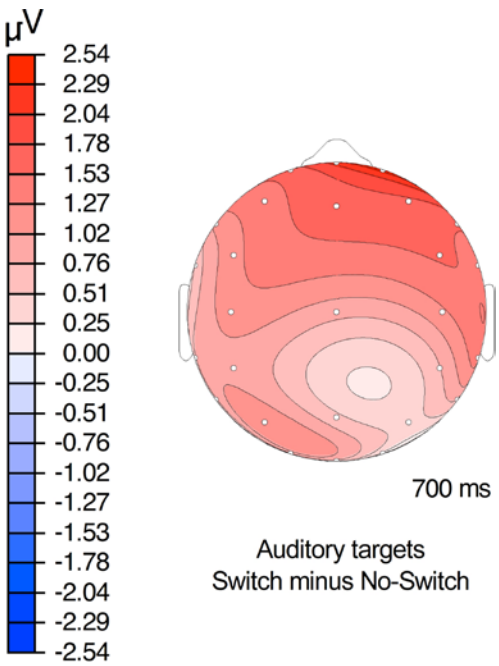


FIGURE 6 | Topography of the switch effect for auditory targets.

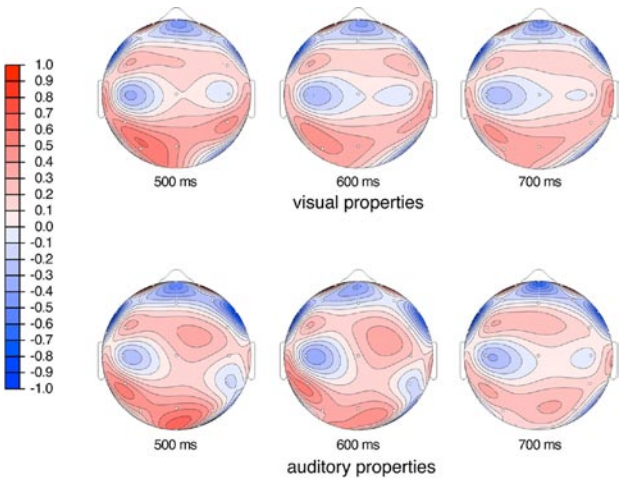


FIGURE 7 | Current source density (CSD) maps of responses to visual and auditory properties including both switch and no-switch conditions. The units are normalized values of micro amps per square meter. CSD maps highlight local differences between electrode sites likely to reflect nearby neural generators. These maps suggest a subtle difference in the configuration of neural generators and timing of activation for the visual versus auditory property stimuli during the 500–800 ms interval, particularly at 600 ms.

was similar and also included a factor of hemisphere (right, left). This test also revealed a difference between modalities that interacted with anteriority [$F(6,114) = 3.55$, $p < 0.05$, $\varepsilon = 0.41$ MSE = 3.70], but no hemispheric differences were significant (Fs < 1.8). No differences at the lateral sites were observed (Fs < 2). The interaction effects between modality and scalp location can be seen in **Figure 7** with the current source density (CSD) plots. These figures show that the visual and auditory properties result in different patterns of voltage change during this time interval.

DISCUSSION

The present study investigated the electrophysiological correlates of the conceptual modality switch effect, an effect used to argue that conceptual tasks recruit perceptual processing systems. We predicted that the sequencing of property verification trials in same modality versus different modality pairs would be reflected in semantic processing of target properties, and thus would modulate the amplitude of the N400 component of the ERP. While this was indeed the case for the visual properties we tested, it was not the case for the auditory properties. Relative to the no-switch trials, visual properties in the switch condition elicited a larger negativity in the N400 time window; by contrast, auditory properties elicited a larger positivity 500–800 ms after stimulus onset in the switch condition. No early differences emerged for N1–P2 components arguing against the suggestion that the switch effect involves low-level visual processing.

N400 EFFECT

The first effect of interest was the negativity observed 200–500 ms after the onset of visual property terms. As predicted, no-switch trials elicited a smaller negativity than did the switch trials during a time interval typically associated with the semantic processing of words and the elicitation of the N400 component. Experts differ on the exact functional significance of this component, with some arguing it indexes lexical access (Kutas and Federmeier, 2000; Lau et al., 2008), and others contextual integration processes (e.g., Hagoort, 2008). There is widespread agreement, however, that the N400 indexes processing events associated with the construction of meaning, and, further, that its amplitude is related to processing difficulty (see Wu and Coulson, 2005 for a review). In general, contextual factors that facilitate processing lead to reduced amplitude N400; for example, words elicit smaller N400 when preceded by related than unrelated words, and smaller N400 when preceded by supportive than unsupportive sentence and paragraph contexts (see Kutas and Federmeier, 2011 for extensive review).

Results of the present study suggest that the perceptual modality of the property term on a previous trial can comprise a supportive semantic context, and that N400 priming effects can be observed between subsequent decisions disguised to participants as completely independent trials. The smaller negativity observed here for the no-switch trials thus suggests that semantic processing of visual target properties was facilitated by processing a visual prime property relative to an auditory prime property. We attribute this facilitated processing to the use of modality specific sensory simulations to mentally represent objects. While perceptual modalities are recruited automatically during concept processing in general, attention can focus more or less on specific modalities. In the property verification task, the presentation of a modality specific property can direct attention to the relevant modality. If the next trial has a property from a different modality (as in the switch condition) the focus shifts to a simulation in the newly relevant modality in order to represent the property. This shift incurs a processing cost which is evident in the ERP differences observed in the present study and reaction time differences of previous studies (Pecher et al., 2003).

Our results are consistent with those of a recent study by Hald et al. (submitted). Hald et al. (submitted) also used a modality switch paradigm in which they presented visual and tactile properties and obtained N400 differences between switch and no-switch trials. Thus, it seems that the N400 effect for modality switching is robust. The identification of the N400 as an ERP index of the conceptual modality switch effect suggests that the cost of shifting between modalities, in this case driven by visual property words, is reflected in semantic processes. This further implies that the semantic activation indexed by this ERP component includes the activation of perceptual features. Results of the present study are thus consistent with ERP studies that have demonstrated modulation of the N400 based on categorical relations that imply similar visual features (Federmeier and Kutas, 1999), and so-called perceptual priming between items such as *pizza* and *coin* that share a salient visual feature (Kellenbach et al., 2000). In sum, results of the present study are in keeping with an account of concepts as involving sensorimotor simulations (e.g., Barsalou, 1999) and suggest that the access of visual features occurs during meaning processing.

LPC EFFECTS

Two effects of interest were observed in the interval 500–800 ms after the onset of property terms. First, visual versus auditory properties elicited ERPs with subtle topographic differences (modality effects). Second, the switch manipulation modulated the ERPs to auditory but not visual property terms (modality switch effects).

Modality effects

Between 500 and 800 ms ERP patterns differed across midline and medial electrode sites for auditory versus visual property decisions. The positivity elicited by auditory properties was more frontocentrally focused than that elicited by visual properties. **Figure 7** illustrates this relatively subtle difference in the scalp topography particularly visible at 500 and 600 ms after stimulus onset. The CSD maps plot the second spatial derivative of the ERP waveforms, and as such highlight differences in the voltage recorded at adjacent electrode sites. The electrode montage used in the present study was too sparse to allow localization but the observed scalp topography differences imply differences in the neural generators underlying the brain response to visual versus auditory property terms. These differences observed between visual and auditory processing are compatible with related fMRI studies that show areas of unique brain activity for properties describing different modalities (Goldberg et al., 2006). The timing of observed topographic differences is later than initial semantic activation implicated in the generation of the N400 component. Semantic and pragmatic manipulations have, however, been observed to modulate the amplitude of the ERP in this interval (see e.g., Regel et al., 2010 for a review). Differences in the brain response to visual and auditory properties are consistent with the hypothesis that perceptual networks help subserve the neural representations of concepts, and the corollary that such networks would be different for concepts that predominantly activate one perceptual modality over another.

Modality switch effects

The other effect of interest in the present study was a positive deflection of the LPC for auditory switch trials relative to the auditory no-switch trials between 500 and 800 ms, primarily at anterior electrode sites (see **Figure 6**). This effect is likely related to the P3, a family of ERP components that index memory processing, whose amplitude reflects the allocation of attention, and whose latency is proportional to the task-relevant stimulus evaluation process (see Polich, 2007 for a review). In view of the relatively long reaction times on the property verification task (>900 ms), the timing of the late positivity observed in the present study (500–800 ms after the onset of the auditory property term) is consistent with its interpretation as an index of the property verification decision. In studies of the P3, the same target stimulus has been shown to elicit a larger positivity in the ERP in difficult than in easy discrimination tasks (Comerchero and Polich, 1999). On this interpretation, the larger late positivity on the switch trials suggests the auditory property verification judgments were more difficult when preceded by a visual prime trial than another auditory one. Alternatively, the anterior distribution of the LPC switch effect suggests the predominance of the P3a sub-component associated with attentional orienting to novel stimuli (see Polich, 2007 for review). On this interpretation, the larger late positivity we observed need not imply greater processing difficulty, but rather an appreciation of the switch trials as involving more novelty than the no-switch trials – presumably because the switch trials required participants to activate semantic features from a different modality.

Hald et al. (submitted) also found a positivity for switch items elicited by a conceptual modality switch task but only over posterior electrodes, differing from the distribution described here (**Figure 6**). Their finding of a posterior positivity co-occurred with a larger negativity for switch trials over anterior electrodes in the same time intervals. The timing and scalp distribution of these effects were interpreted as a unified frontal N400 effect similar to that elicited by pictures. The different ERP patterns found by Hald et al. (submitted) at anterior and posterior electrode sites were revealed as a topography difference but this scalp difference cannot be compared to that reported in the current study because the topographic differences reported here were driven by different modalities, a dimension not tested by Hald et al. (submitted).

DIFFERENCES BETWEEN VISUAL AND AUDITORY PROPERTY VERIFICATION

The most surprising result of the present study was the observed difference in the conceptual modality switch effect for visual versus auditory properties. As noted above, visual properties elicited reduced N400 in no-switch relative to switch trials, suggesting our experimental manipulation affected semantic processing of the targets. Auditory properties, however, elicited an enhanced LPC, suggesting the manipulation impacted neural processes occurring later than those indexed by the N400, and were more likely related to making the decision about whether the property was typical of the concept.

Whereas neither finding is surprising alone – that is, a conceptual modality switch might reasonably be predicted to impact either the semantic processing of the stimuli, or the difficulty of decisions regarding property verification, or, indeed, both sets of processes – our finding of semantic effects for visual properties and decision-related effects for auditory properties was unexpected. Prior reports of the conceptual modality switch effect using reaction time measures have found similar sized switch effects for properties from different modalities (Pecher et al., 2009). Similarly, studies of the perceptual modality switch effect also report similar sized switch effects for visual, auditory, and tactile stimuli, with the only difference being a trend for tactile primes to yield longer reaction times for subsequent visual and auditory probes (Spence et al., 2001). However, reaction times measure only the end point of a property verification process, while ERPs provide an index of brain activity from the onset of the stimulus until the generation of the behavioral response on the task. ERP data in the present study suggest the switch manipulation affects different aspects of processing in the verification of visual versus auditory properties.

Our observed differences between auditory and visual switch effects are consistent with a prior ERP study of the perceptual modality shift effect by Gondan et al. (2007) in which stimuli involved either LED flashes (visual targets) or bursts of white noise (auditory targets). They found that visual targets following visual primes compared to when they followed auditory primes elicited ERP effects similar to those found for increased visual attention – namely, an amplified N1 component. In contrast, auditory targets elicited smaller N1 and P2 components when they followed auditory primes than when they followed visual primes. The fact that ERP differences for the switch effect were opposite in the visual and auditory domains was an unexpected asymmetry. The authors explain this asymmetry by suggesting different mechanisms driving the switch effects in the two perceptual domains. They suggest a "neural trace" explanation for the auditory domain in which residual activity from an auditory prime speeds the response and processing for a subsequent auditory stimulus. The result of this priming is a smaller ERP component for the target auditory stimulus. For the visual targets, ERP amplification for the same modality condition is explained through attentional mechanisms because increased attention tends to result in amplified perceptual ERP components. These different patterns suggest that different mechanisms might be driving the modality switch effect in the visual and auditory domains. Likewise, results of the present study suggest that different mechanisms were involved for the conceptual modality switch in the case of visual versus auditory property terms.

One account for why different mechanisms would drive the conceptual switch effects in the present study is that the particular visual and auditory property words we used access the perceptual domains differently. In particular, the visual property words may refer to relatively pure visual experiences, whereas auditory properties may refer to mixed visual and auditory experiences. For example, green (as for asparagus) might refer to a purely visual perception while clicking (as for high heels) might refer to a combined auditory and visual experience. We examined this possibility using the Lynott and Connell (2009) norms. Lynott and Connell (2009) asked participants to what extent each of 423 property words were experienced via each of the five sensory modalities. Of the 48 property words used in each modality category of our study, 37 visual properties and 27 auditory properties were represented in their list. Our subset of visual property words had an average visual ranking of 4.65 (out of 5.0 possible) and the subset of auditory words had an average auditory ranking of 4.60 (two-tailed t-test, $t < 1$), verifying the experimental conditions used in our study.

However, when considering both visual and auditory rankings for each of these sets our auditory properties appear more multimodal than our visual properties as indicated by a smaller difference between their auditory and visual rankings (auditory property difference = 2.44, visual property difference = 4.18; $t(41) = 8.99$, $p < 0.01$). This classification is also consistent with a *modality exclusivity* score available in Lynott and Connell's (2009) norms. For each property word the modality exclusivity score factors the strength of the rating for an individual modality relative to ratings for all five sensory modalities. The visual properties used in our study had a higher modality exclusivity ranking (0.73, of possible values between 0.0 and 1.0) than our auditory properties [0.58; $t(59) = 4.31$, $p < 0.01$].

While it is clear that our auditory properties are characterized as typically experienced via hearing [as indicated by values derived from the Lynott and Connell (2009) norming study], their greater multimodal characteristic might have led to a weaker switch effect than that seen for the visual properties. In perceptual studies of the modality switch effect, a bimodal target stimulus (e.g., simultaneous beep and flash) following a unimodal stimulus (e.g., a flash) produces a smaller switch cost than unimodal targets following unimodal primes (e.g., a beep following a flash; Gondan et al., 2004). The reduction in the switch effect is presumed to result because only one of the two modalities making up the bimodal target stimulus requires a switch from the modality of the previous stimulus; the other, in fact, is primed. The absence of an observed N400 effect in our ERP results for auditory properties could reflect a lack of power to see an attenuated modality switch for these auditory properties that are more multimodal than the visual properties for which we did find an N400 effect. The decision-related LPC effect on the other hand would thus index more effort required to attribute the multimodal (auditory) property to a concept in the context of a visual prime.

Using a combination of published norms and dictionary definitions, we identified four of the visual target properties and eight of the auditory target properties employed in our study as being multimodal, that is, either having a modality exclusivity score (as defined by Lynott and Connell, 2009) of less than 0.51, or a dictionary definition that mentioned more than one modality. We elimi-

nated multimodal items from ERP waveforms and conducted a *post hoc* analysis of ERPs elicited by the remaining unimodal stimuli. In the 200–500 ms window the same pattern of significant effects was obtained as for the complete dataset. Reanalysis restricted to unimodal items thus suggested the N400 switch effect for visual properties was robust and slightly larger than that measured for the full set of experimental items, but still failed to reveal an N400 switch effect for the auditory properties.

A similar reanalysis of the LPC interval failed to reveal either the modality by switch interaction ($F < 1$) or the auditory switch effect ($F < 1$) observed in the original analyses. This raises the possibility that the LPC switch effect observed in the present study primarily reflects the brain response to the multimodal items. Consistent with this, further analysis also suggested a trend for multimodal visual and multimodal auditory properties to elicit slightly more positive ERPs in the 500–800 ms interval than unimodal visual [1.13 μV, $F(1,19) = 3.57$, $p = 0.074$, MSE = 370.59] and unimodal auditory [0.84 μV, $F(1,19) = 2.86$, $p = 0.107$, MSE = 205.51] properties, respectively. According to this interpretation, the auditory switch effect observed in our original analysis reflects the greater difficulty of responding to multimodal auditory properties following (more likely unimodal) visual than (more likely multimodal) auditory primes. The greater multimodality of the auditory properties also suggests an alternative explanation for the different topography of ERPs elicited by auditory and visual properties measured 500–800 ms post-stimulus onset (discussed in "Modality Effects"). These reanalyses must, however, be interpreted with caution since the comparison of unimodal versus multimodal stimuli involve ERPs derived from different numbers of trials, and the number of visual multimodal trials was particularly low. More definitive conclusions regarding brain activity elicited by multimodal versus unimodal items in a property verification task would require a stimulus set specifically designed for this purpose.

CONCLUSION

The present study contributes to evidence demonstrating that concepts referring to perceptual properties are recruiting perceptual processing resources. Whereas previous studies have shown similar modality switch effects in conceptual processing, the present study informs us in a more detailed way on the locus of this switch effect. ERP measures showed that the elicitation of perceptual meaning, as typically demonstrated by switching costs, is evident at the semantic level or at later decision-making stages of processing. The switch effect for visual properties was different from the switch effect for auditory properties due to either different underlying mechanisms driving the processes or different modal representations of these properties. Both explanations support a theory of concepts as a reactivation of brain areas important for the perception of the world. Just as seeing candles flicker generates different neural activity from hearing high heels click, we expect the concepts representing these events to differ as well.

ACKNOWLEDGMENTS

This research was supported by grants from the Netherlands Organization for Scientific Research (NWO) to Diane Pecher and from the Erasmus University Trustfonds to Diane Pecher and René Zeelenberg. We would like to thank Lauren Cardoso, Nafees Hamid, and Rubén Moreno for their assistance collecting data. The authors would also like to thank Lea Hald and Frederico Marques for their feedback and reviews that allowed us to improve this paper.

REFERENCES

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain Lang.* 31, 308–327.

Comerchero, M. D., and Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clin. Neuropsychol.* 110, 24–30.

Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495.

Goldberg, R. F., Perfetti, C. A., and Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* 26, 4917–4921.

Gondan, M., Lange, K., Rösler, F., and Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychon. Bull. Rev.* 11, 307–313.

Gondan, M., Vorberg, D., and Greenlee, M. W. (2007). Modality shift effects mimic multisensory interactions: an event-related potential study. *Exp. Brain Res.* 182, 199–214.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363, 1055–1069.

Holcomb, P. (1988). Automatic and attentional processing: an event-related brain potential analysis of semantic priming. *Brain Lang.* 35, 66–85.

Holcomb, P., and Neville, H. (1990). Semantic priming in visual and auditory lexical processing. *Lang. Cogn. Process.* 5, 281–312.

Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., and Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cogn. Neuropsychol.* 20, 525–540.

Kellenbach, M. L., Wijers, A. A., and Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cogn. Brain Res.* 10, 67–75.

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.

Lau, E. F., Phillips, C., and Poeppel, C. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.

Lynott, D., and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* 41, 558–564.

Marques, J. F. (2006). Specialization and semantic organization: evidence for multiple semantics linked to sensory modalities. *Mem. Cogn.* 34, 60–67.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Pecher, D., Van Dantzig, S., and Schifferstein, H. N. J. (2009). Concepts are not represented by imagery. *Psychon. Bull. Rev.* 16, 914–919.

Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychol. Sci.* 14, 119–124.

Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: modality-specific effects of prior activation. *Psychon. Bull. Rev.* 11, 164–167.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.

Regel, S., Coulson, S., and Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the

comprehension of irony. *Brain Res.* 1311, 121–135.
Rodway, P. (2005). The modality shift effect and the effectiveness of warning signals in different modalities. *Acta Psychol.* 120, 199–226.
Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., and Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45, 2802–2810.
Smith, M. E., and Halgren, E. (1989). Dissociation of recognition memory components following temporal lobe lesions. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 50–60.
Solomon, K., and Barsalou, L. W. (2004). Perceptual simulation in property verification. *Mem. Cogn.* 32, 244–259.
Spence, C., Nicholls, M. E. R., and Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Percept. Psychophys.* 63, 330–336.
Urbach, T. P., and Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology* 39, 791–808.
van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.
Vermeulen, N., Niedenthal, P. M., and Luminet, O. (2007). Switching between sensory and affective systems incurs processing costs. *Cogn. Sci.* 31, 183–192.
Wu, Y. C., and Coulson, S. (2005). Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42, 654–667.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2010; paper pending published: 29 July 2010; accepted: 09 January 2011; published online: 08 February 2011.

Citation: Collins J, Pecher D, Zeelenberg R and Coulson S (2011) Modality switching in a property verification task: an ERP study of what happens when candles flicker after high heels click. Front. Psychology ***2****:10. doi: 10.3389/fpsyg.2011.00010*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Collins, Pecher, Zeelenberg and Coulson. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 05 November 2010
doi: 10.3389/fpsyg.2010.00183

Top-down and bottom-up contributions to understanding sentences describing objects in motion

Shirley-Ann Rueschemeyer[1,2]*, Arthur M. Glenberg[3], Michael P. Kaschak[4], Karsten Mueller[1] and Angela D. Friederici[1,5]

[1] *Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2] *Donders Centre for Cognition, Radboud University, Nijmegen, Netherlands*
[3] *Laboratory for Embodied Cognition, School of Biological and Health Systems Engineering, Arizona State University, School of Life Sciences, Tempe, AZ, USA*
[4] *Departments of Neuroscience and Communicative Disorders, Florida State University, Gainesville, FL, USA*
[5] *Center for Advanced Study in the Behavioral Sciences at Stanford University, Stanford, CA, USA*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Michael E. J. Masson, University of Victoria, Canada
Christine Wilson, Emory University, USA

****Correspondence:***
Shirley-Ann Rueschemeyer, Donders Centre for Cognition, Radboud University, Montessorilaan 3, 6525 HR, Nijmegen, Netherlands.
e-mail: s.rueschemeyer@donders.ru.nl

Theories of embodied language comprehension propose that the neural systems used for perception, action, and emotion are also engaged during language comprehension. Consistent with these theories, behavioral studies have shown that the comprehension of language that describes motion is affected by simultaneously perceiving a moving stimulus (Kaschak et al., 2005). In two neuroimaging studies, we investigate whether comprehension of sentences describing moving objects activates brain areas known to support the visual perception of moving objects (i.e., area MT/V5). Our data indicate that MT/V5 is indeed selectively engaged by sentences describing objects in motion toward the comprehender compared to sentences describing visual scenes without motion. Moreover, these sentences activate areas along the cortical midline of the brain, known to be engaged when participants process self-referential information. The current data thus suggest that sentences describing situations with potential relevance to one's own actions activate both higher-order visual cortex as well brain areas involved in processing information about the self. The data have consequences for embodied theories of language comprehension: first, they show that perceptual brain areas support sentential-semantic processing. Second the data indicate that sensory-motor simulation of events described through language are susceptible to top-down modulation of factors such as relevance of the described situation to the self.

Keywords: embodiment, sentence comprehension, self-referentiality, visual motion

According to embodiment theory, language comprehension requires neural resources ordinarily used for perception, action, and emotion. Thus, for example, understanding language about actions (e.g., "Open the drawer.") requires simulation using neural systems involved in action (i.e., ventral premotor and inferior parietal cortex), whereas comprehension of language about visual motion (e.g., "The car approached you.") requires access to neural systems involved in motion perception (i.e., posterior lateral temporal cortex). The predictions regarding action-related language have been confirmed behaviorally (e.g., Glenberg and Kaschak, 2002), using imaging techniques (e.g., Hauk et al., 2004; Rueschemeyer et al., 2007), and using TMS (Buccino et al., 2005; Glenberg et al., 2008). The case for language denoting moving objects appears less clear cut.

Behavioral evidence has been reported supporting a link between neural systems involved in perception of visual motion and comprehension of language describing objects in motion. For example, Kaschak et al. (2005) showed that participants watching a video clip depicting motion are slower to respond to an acoustically presented sentence describing an object in motion, when the trajectories of the visually perceived and linguistically described objects are congruent in contrast to if the trajectories are incongruent. The interpretation offered is that in the visual-linguistic match condition, the neural systems used to analyze motion are engaged by the visual stimulus, and are thus less available for use in sentence comprehension. This leads to interference between the processing of the visual and the language stimuli. While this study may indeed show the behavioral consequences of interference on the neural level, Kaschak et al. (2005) cannot provide conclusive neuro-functional evidence on the basis of their reaction time study.

The visual perception of moving stimuli is known to rely on area MT/V5 in posterior middle temporal cortex (Tootell et al., 1995a; Smith et al., 1998). Importantly, this area has been shown to be responsive both to the actual perception of moving stimuli, as well as to imagery of motion (Goebel et al., 1998). Thus, according to the above mentioned theory it is MT/V5 which is expected to respond both to visually perceived objects in motion and to sentences describing objects in motion.

Those imaging studies investigating visuo-motor language representations to date have focused on the processing of words in isolation (Kable et al., 2002, 2005; Noppeney et al., 2005; Pirog Revill et al., 2008). Typically motion verbs are compared with non-motion verbs (Noppeney et al., 2005; Pirog Revill et al., 2008) or with words describing various types of objects or states (Kable et al., 2002, 2005). For example, Kable and colleagues (2002, 2005) presented participants with a target word and two alternative words, one of which was relatively more semantically related to the target than

the other (e.g., target word: *skipping*: alternatives: *rolling*, *bouncing*). When making decisions about motion verbs, participants showed greater levels of activation in brain areas very proximal to MT/V5 compared to when making the same type of decision about static object words. Typically (as in the case of the studies by Kable and colleagues), greater activation for motion verbs is observed in posterior lateral temporal cortex (PLTC), somewhat anterior to and distinct from MT/V5. Despite the lack of a clean overlap with MT/V5 these results are generally taken to support the notion that retrieving conceptual information about motion (i.e., action) verbs activates visual-motion representations.

One shortcoming in these previous studies is the fact that most have focused on the processing of words in isolation. Specifically, words in isolation may be indeed be associated with motions (or have motion components in their semantic composition), but they are not directly comparable with the visual stimuli which have been used to elicit activation in MT/V5 in the visual domain (see also Wallentin et al., 2005). For example, a typical stimulus used to activate MT/V5 is a display of expanding dots (Tootell et al., 1995a), however neither the word "expand" nor the word "dot" describes the stimulus as well as the combination of "expanding dot display" or better yet the sentence "The dots moved away from the center of the screen." Thus previous studies using single word stimuli may simply fail to specify enough information about an event to make visual processing possible. In the current two experiments, we therefore presented participants with *sentences* describing objects in motion or objects at rest, and compared the neural correlates of comprehending each sentence type. This design is comparable to what was used in the behavioral study of Kaschak et al. (2005), however it provides direct evidence as to the substrate underlying the previously observed behavioral effect.

A second reason to use sentence stimuli is to create a better match between language stimuli with and without motion content. Specifically, previous studies have frequently contrasted action verbs (i.e., language with motion content) with concrete nouns (i.e., language without motion content). This has led some researchers to question whether observed differences in PLTC might actually reflect differences in grammatical category rather than semantic information (Bedny et al., 2008). On the other hand, while the focus of the aforementioned studies has indeed been on action verbs, Kable et al. (2005) did include a contrast between nouns with implicit motion content (e.g., words denoting manipulable objects) and nouns without motion content (e.g., animals). The results showed that manipulable object nouns elicited greater levels of activation within PLTC (i.e., middle temporal gyrus) than animal nouns. This indicates that differences in PLTC can be found within word categories. Nevertheless, we avoid grammatically dependent activation differences by presenting sentences with similar syntactic structure but different propositional content.

EXPERIMENT 1: SENTENCES DESCRIBING MOVING VS. STATIC SCENES

MATERIALS AND METHODS

Participants

Twelve participants (six male) aged 22–30 years (mean = 26) participated in this experiment after giving informed written consent.

Materials

Seventy-two sentences were constructed for the experiment and recorded by a native German speaker (examples can be seen in **Table 1**). Twenty-four of the 72 sentences described items moving toward the participant ($SENT_T$); 24 sentences described items moving away from the participant ($SENT_A$) and 24 sentences described items which were not in motion ($SENT_S$).

In addition three visual stimuli were created. The first of these was a black and white spiral rotated in a clockwise direction to give the illusion of motion toward the participant (VIS_T); secondly a black and white spiral rotated in a counter-clockwise direction to give the illusion of motion away from the participant (VIS_A); and thirdly a static image, which was created by scrambling the visual input of the spiral stimuli, which did not create any illusion of motion (VIS_S). All images were 580×580 pixels.

Each of the 72 constructed sentences was presented in conjunction with all three visual stimulus options (VIS_T, VIS_A, VIS_S) yielding 216 critical trials and nine potential trial conditions.

In addition to the critical experimental stimuli, participants received catch trials designed to test their engagement (see below), as well as null events in which no stimulus was presented for the duration of a normal trial.

Stimulus presentation

Participants lay supine in the scanner. Visual stimuli were presented via 3D glasses onto a virtual computer monitor, and spoken sentence stimuli were presented via headphones. Participants could control a small response box with their right hand (i.e., by pressing one of two buttons). Responses to catch trials were recorded via the response box (see below for more details).

A single critical trial constituted presentation of a sentence in conjunction with visual stimulation. Trial length was 4 s, and the duration of each sentence was approximately 2 s. To enhance the temporal resolution of the acquired signal, trial onset was jittered by 250 ms with respect to the first scan in each trial. In other words, while the first trial was initiated co-incidentally with the beginning of a scan, the second trial and third trials were initiated 250 and 500 ms. after initiation of a scan, respectively. This was done by setting the intertrial interval (ITI) to 4250 ms. Each trial began with presentation of the visual stimulus (VIS_T, VIS_A, VIS_S), which remained visible for the duration of the trial (4 s). An auditory sentence ($SENT_T$, $SENT_A$, $SENT_S$) was presented to participants over headphones 500 ms post onset of visual stimulation.

Following 8% of the trials ($N = 18$) a catch trial was introduced, which did not enter into the functional data analysis. In catch trials participants were asked to indicate whether motion in the preceding

Table 1 | Examples of the sentences used in Experiment 1 with English translations.

Condition	Example sentence	English translation
Toward	Das Auto fährt auf Dich zu.	The car drives toward you.
Away	Das Auto fährt von Dir weg.	The car drives away from you.
Static	Das Auto sieht groß aus.	The car looks big.

visual stimulation had been in a clockwise or a counter-clockwise direction. In this manner we ensured that participants processed the visual stimuli.

Experimental stimuli were presented in three blocks of approximately 10 min each with a 1-min pause between blocks (i.e., the purpose of the blocks was only to give participants a rest). Each of the 72 constructed sentences was presented in conjunction with all three visual stimuli options (VIS_T, VIS_A, VIS_S) yielding 216 critical trials. Although sentences were presented three times in the course of the entire experiment, each sentence was presented only once within any given block. The order of visual stimulus presentation in conjunction with a single sentence was balanced across the experiment. The 216 critical trials plus 24 null events (low level baseline condition consisting of a blank screen presented for 4 seconds) were presented in a balanced, pseudorandomized order, such that a single condition was not repeated on more than three consecutive trials, and the probability of each condition following any other condition was matched.

FMRI data acquisition

Scanning was performed using a 3-T MedSpec 30/100 scanner (Bruker, Ettlingen, Germany) and a birdcage head coil. Twenty axial slices (4-mm thickness, 1 mm inter-slice distance, FOV 19.2 cm, data matrix of 64 × 64 voxels, in-plane resolution of 3 × 3 mm) were acquired every 2 s during functional measurements (BOLD sensitive gradient EPI sequence, TR = 2 s, TE = 30 ms, flip angle = 90°, acquisition bandwidth = 100 Hz) with a 3-Tesla Bruker Medspec 30/100 system. Prior to functional imaging T1-weighted MDEFT images (data matrix 256 × 256, TR = 1.3 s, TE = 10 ms) were obtained with a non-slice selective inversion pulse followed by a single excitation of each slice (Norris, 2000). These images were used to co-register functional scans with previously obtained high-resolution whole head 3D brain scans (128 sagittal slices, 1.5 mm thickness, FOV 25 × 25 × 19.2 cm, data matrix of 256 × 256 voxels) (Lee et al., 1995).

FMRI data analysis

The functional imaging data was processed using the software package LIPSIA (Lohmann et al., 2001). Functional data were corrected first for motion artifacts using a matching metric based on linear correlation. Data were subsequently corrected for the temporal offset between slices acquired in one scan using a cubic–spline interpolation based on the Nyquist–Shannon Theorem. Low-frequency signal changes and baseline drifts were removed by applying a temporal highpass filter to remove frequencies lower than 1/80 Hz and a spatial Gaussian filter with 10 mm FWHM was applied.

Functional slices were aligned with a 3D stereotactic coordinate reference system using linear registration. The registration parameters were acquired on the basis of the MDEFT and EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference data set which was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The registration parameters were further used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system. This linear normalization process was improved by a subsequent processing step that performed an additional non-linear normalization (Thiron, 1998). The transformation parameters obtained from both normalization steps were subsequently applied to the functional data. Voxel size was interpolated during co-registration from 3 × 3 × 4 mm to 3 × 3 × 3 mm.

The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Worsley and Friston, 1995). The design matrix was generated with a synthetic hemodynamic response function and its first derivative (Josephs et al., 1997; Friston et al., 1998). The model equation, made up of the observed data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM.

For each participant three critical contrasts were computed. The first of these served to localize the functional region of interest (fROI) in area MT (see section Functional Region of Interest), known to be important in the processing of visual motion (Tootell et al., 1995b). The second two contrasts (see section Whole Brain Analysis) served to locate areas which were involved in the processing of motion content of sentence materials outside of the predefined fROI. Because individual functional datasets had been aligned to the standard stereotactic reference space, a group analysis based on contrast images could be performed. Single-participant contrast images were entered into a second-level random effects analysis for each of the contrasts. The group analysis consisted of a one-sample *t*-test across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from 0. Subsequently, *t*-values were transformed into *Z*-scores.

Functional region of interest

In order to locate the area MT, a direct contrast between brain activation elicited by moving visual stimuli was compared to activation elicited by static images ($VIS_{T,A}$ vs. VIS_S) in a group average across all participants. In all cases, co-occurring sentence stimuli described static scenes ($SENT_S$). The maximally activated voxel in the posterior middle temporal gyrus plus the 26 voxels adjacent to the peak voxel (ca. 700 mm^3) in this contrast were identified as the fROI. For each individual participant, time course data was extracted from the voxels in the predefined MT fROI for each Sentence condition ($SENT_T$, $SENT_A$, $SENT_S$) in conjunction with static images (VIS_S). In this way, trials belonging to the functional localizer and trials included in the contrast of sentence materials were independent from one another (see also Saxe et al., 2006).

For the analysis of the experimental sentence stimuli, the time course of MR signal intensity was extracted for each individual participant from all voxels within the predefined MT region of interest. The average percent signal change was calculated for each subject and stimulus type, using the average signal intensity during null events as a baseline. Because the fMRI response typically peaks 6 s after stimulus onset, mean percent signal change was calculated for each participant between 4 and 8 s post-stimulus onset.

Mean percent signal change for each participant in each sentence condition was entered into a repeated measure ANOVA with the within subject factor Sentence Meaning (Toward, Away, Static). Within each fROI the two critical comparisons concerned (1) sentences describing motion toward the listener vs. sentences

describing stationary scenes, and (2) sentences describing motion away from the listener vs. sentences describing stationary scenes. Bonferroni correction for multiple comparisons was used in determining the statistical significance of differences between sentence conditions within fROIs.

Whole brain analysis

In order to identify brain areas outside of MT selectively activated by the motion content of sentences, we compared brain activation elicited by sentences describing motion toward the participant vs. sentences describing static content ($SENT_T$ vs. $SENT_S$) in conjunction with static images (VIS_S) as well as sentences describing motion away from the participant vs. sentences describing static content ($SENT_A$ vs. $SENT_S$) in conjunction with static images (VIS_S).

To protect against false positive activation a double threshold was applied by which only regions with a *Z*-score exceeding 3.09 ($p < 0.001$, uncorrected) and a volume exceeding 18 voxels (650 mm^3) were considered (Forman et al., 1995). This non-arbitrary activation size was determined using a Monte Carlo simulation and is equivalent to a corrected significance level of $p < 0.05$.

RESULTS

BEHAVIORAL RESULTS

Participants responded successfully to the 18 catch trials (mean = 16.3, SD = 2.7). Eliminating two participants who made more than 60% errors improved average performance (mean = 17.4, SD = 1.1), but did not substantially change the functional results. Therefore functional data from all 12 participants were included.

FUNCTIONAL REGION OF INTEREST (fROI)

In the left hemisphere MT was identified at Talairach coordinate −47, −75, 9. In the right hemisphere the peak activated voxel was located at 43, −67, 9 (see **Figure 1**).

Mean percent signal change for each participant within left and right MT for the critical sentence conditions was entered into a repeated measure ANOVA with the within subjects factor Sentence Meaning (Toward, Away, Static). In the left hemisphere a main effect of Sentence Meaning was observed [$F(2,22) = 4.19$; $p < 0.05$]. This reflected significantly more activation for Toward than Static sentences [$F(1,11 = 9.07$, $p = 0.01$], however no reliable difference in activation level was seen for Away vs. Static sentences [$F(1,11) < 1$]. In the right hemisphere no reliable main effect of Sentence Meaning was observed [$F(2,22) < 1$].

WHOLE BRAIN ANALYSIS

In addition to the fROI analysis described above, we identified additional areas of the brain which responded selectively to sentence meaning in a whole brain analysis. Relevant brain areas were identified by directly comparing activation elicited by (1) sentences describing motion toward the participant vs. sentences describing static scenes, and (2) sentences describing motion away from the participant vs. sentences describing static scenes.

Sentences describing motion toward the participant vs. sentences describing static images elicited activation within the right posterior superior temporal sulcus, directly superior to the region identified as MT in the right hemisphere by the fROI analysis. In the left hemisphere activation in MT did not pass the cluster-size threshold (see **Table 2**) at the whole brain level, however a smaller area (108 mm^3) was activated at the level $p < 0.001$, uncorrected. Additionally activation was seen in several areas

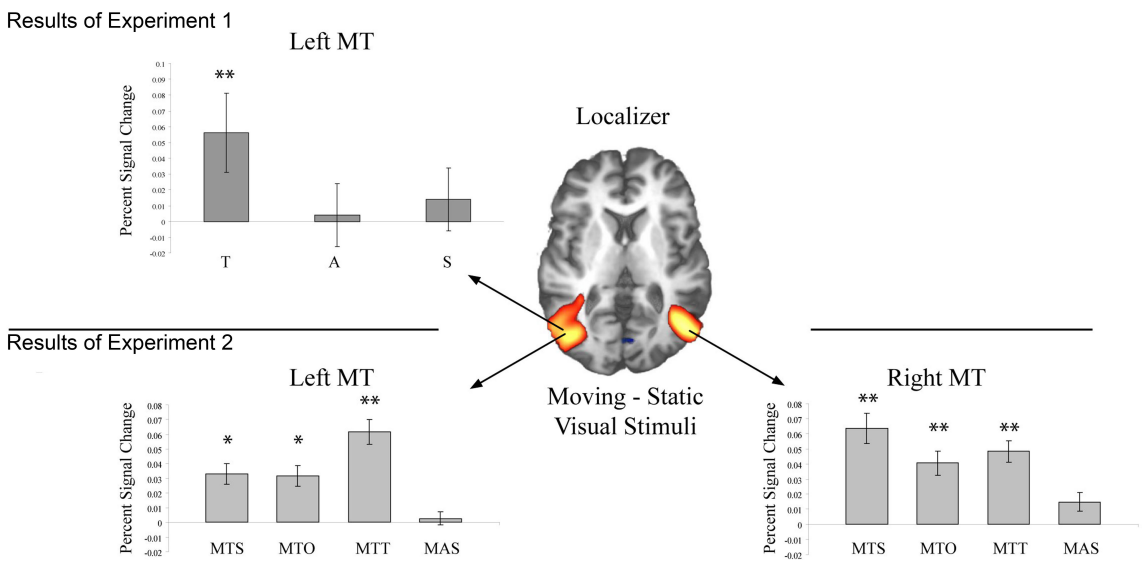


FIGURE 1 | Activation in bilateral MT during each of the two experiments. The *Z*-map in the center shows the pattern of activation elicited by the functional localizer ($Z > 3.01$) in Experiment 1. In the top panel activation within left MT is shown for each sentence condition (T = Toward, A = Away, S = Static). Only activation in conjunction with T sentences reached significance in left MT. In the bottom panel activation within bilateral MT is shown for each sentence condition (MTS = movement toward self, MTO = movement toward other, MTT = movement toward thing, MAS = movement away from self). A single asterisk indicates a difference significant at the 0.05 level, and two asterisks indicate a difference significant at the 0.01 level.

Table 2 | Areas showing significantly different activation (A) within posterior middle temporal gyrus (MT) for the functional localizer and (B) whole brain analysis contrasting different sentence types.

Contrast	Region	Extent (mm³)	Z_{max}	x	y	z
(A) FUNCTIONAL LOCALIZER						
visual motion – static image	Right posterior middle temporal gyrus (RMT)	7965	5.47	43	–66	9
	Left posterior middle temporal gyrus (LMT)	10854	5.11	–44	–73	9
(B) WHOLE BRAIN ANALYSIS						
static image with toward sentences – static sentences	Right posterior superior temporal gyrus (pSTG)	783	3.77	49	–61	21
	Orbitofrontal median wall	648	3.38	1	65	3
	Middle cingulate gyrus	2079	3.93	10	–7	36
	Posterior cingulate gyrus	4968	3.41	7	–49	30
	Posterior median wall (Cuneus)	1080	3.87	10	–85	42
Static image with away sentences – static sentences	No significant differences					

An activation was regarded significant at a voxel threshold of $p < 0.001$ and a cluster size >640 mm³ (equivalent to corrected $p < 0.05$). Region, extent (mm³), Z-values and Talairach coordinates are reported.

along the median wall including medial prefrontal cortex, middle cingulate gyrus, posterior cingulate gyrus, and cuneus (see **Table 2**, **Figure 2**).

Sentences describing motion away from the participant vs. sentences describing static images yielded no significantly activated brain regions in the whole brain analysis.

DISCUSSION

In Experiment 1, we investigated whether comprehending sentences describing an event involving motion reliably activated MT/V5 in contrast to sentences describing a static scene. Previous fMRI research on motion-related language has focused on the processing of action verbs and single nouns (Kable et al., 2002, 2005; Noppeney et al., 2005; Pirog Revill et al., 2008). These studies tend to show activation anterior and dorsal to MT/V5, but not directly within MT/V5. This anterior shift might due to the fact that an isolated verb or noun gives little specific information about motion parameters. Therefore, participants in the current study listened to sentences describing short events about objects in motion and static scenes.

The most important results from Experiment 1 are as follows. First, the coordinates obtained with the functional localizer task are in accordance with those described in previous literature (Tootell et al., 1995a; Kable et al., 2005). Second, within left MT/V5 activation was modulated by sentence meaning, such that sentences describing objects in motion toward the participant (e.g., "The car drives toward you.") elicited reliably more activation than sentences describing static scenes (e.g., "The car looks big."). However, a similar difference was not observed for sentences describing objects in motion away from the participant (e.g., "The car drives away from you."). Additionally, multiple voxels within the right MT/V5 appear to show modulation by sentence meaning (i.e., Toward > Static, see **Figure 1**), this difference, however, did not reach significance when all voxels in the ROI were considered. Thus, we see reliable difference in left MT/V5, but only for those motion sentences presented which describe a "toward" motion rather than an "away" motion.

It is possible that the design used in Experiment 1 was suboptimal for detecting subtle changes within MT (for away motion compared to the static scene sentences). Participants were always presented simultaneously with both a language and a visual stimulus. It is conceivable that subtle language-driven changes in MT were weakened by changes elicited in MT throughout the experiment in response to actual moving stimuli. Additionally, posterior temporal cortex (in close proximity to MT) is known to be a multimodal area that receives input from both auditory and visual channels, and this area may be important in the integration of auditory and visual input (Beauchamp, 2005). Thus, the hypothesized general modulation of MT as a function of movement sentences may have been difficult to detect because participants performed some degree of audio–visual integration in all trials. We therefore conducted a second experiment in which presentations of visual and auditory information were kept separate (see further below).

In addition to the ROI analysis, a whole brain analysis was conducted, which revealed significantly more activation in several brain areas for Toward sentences in contrast to Static sentences, but no reliable differences for Away sentences in contrast to Static sentences. In particular Toward sentences showed higher activation in the orbitomedial prefrontal cortex (OMPFC), posterior cingulate cortex (PC) and cuneus (see **Figure 2**).

The cortical midline structures have been implicated in both processing of self-referential stimuli (review see Northoff and Bermpohl, 2004) as well as modulation of visual attention based on cue-induced anticipation (Small et al., 2003). With respect to the first of these points, the OMPFC together with the PC has been seen to play a role in tasks requiring self-reflection, for example indicating whether or not a given word describes oneself vs. indicating whether or not the same word describes another person (Johnson

et al., 2002; Kelley et al., 2002; Kjaer et al., 2002). In the current data, OMPFC and medial posterior parietal cortex are seen to be more activated for sentences describing objects in motion toward the participant than for sentences with no self-referential content (i.e., Static sentences). Activation in these areas may thus reflect the self-referential content of these sentences for participants. The OMPFC is also one of the primary regions to send output to visceromotor structures in the hypothalamus and brainstem, and is thus well-suited to initiate changes in the body (Ongur and Price, 2000). Therefore activation in this area in conjunction with sentences describing objects in motion toward the participant may reflect awareness of the self entering a situation in which action should be initiated (see further below for related discussion).

With respect to visual attention, the cingulate gyrus along with MPFC has been shown to support monitoring and modulation of visual attention (Small et al., 2003; Taylor et al., 2009). Specifically a distinction has been made between anterior and posterior cingulate regions, with anterior cingulate gyrus (AC) active during visual search, spatial working memory and conflict monitoring and PC together with MPFC responsible for the mediation of visual attention based on cue-induced anticipation. It has been suggested that PC is particularly involved when shifts in visual attention are required to efficiently modify behavior (Small et al., 2003). In contrast to the classic distinction between AC and PC, Vogt (2005) proposes a four-region model of cingulate cortex. Specifically, based on the results of monkey and human neuorophysiological and cytoarchetectonic studies, cingulate cortex is divided into anterior, middle and posterior cingulate as well as retrosplenial cortex. In this model posterior middle cingulate cortex (pMCC) is involved in skeletomotor orientation while PCC is involved in visuospatial orientation that is mediated through connections to the parietal lobe as well as assessment of self-relevant sensations. In a similar vein, Taylor et al. (2009) have shown that connections between insular cortex and MCC support response selection and skeletomotor body control during monitoring of the environment. The results presented here show greater engagement of PC according to traditional views of cingulate cortex, and greater engagement of pMCC and dorsal PPC according to Vogt (2005) for the comprehension of sentences describing objects in motion toward the listener. We suggest that motion of an item toward oneself in many cases requires monitoring of the environment and subsequent initiation of a motor reaction (i.e., to locate the car moving toward one, or ultimately to move out of the way of the approaching car), whereas objects moving away from the participant or static objects may not require an immediate motor response or modification of behavior. Therefore it is possible that activation in the cortical midline structures for Toward sentences reflect (1) self-relevance of sentences describing approaching objects and (2) preparation of neural systems to identify and behaviorally respond to approaching objects. This interpretation is in line with the role of cingulate cortex proposed by Vogt (2005) and Taylor et al. (2009).

In Experiment 2, in addition to separating visual and language stimuli, we test the hypothesis that self-referentiality is critical in eliciting activity in the cortical midline structures. To this end, participants were presented with sentences describing objects in motion toward not only themselves but also toward inanimate objects (e.g., "The car drives toward the bridge"). If our interpretation is correct, sentences describing objects in motion toward inanimate objects should not activate the cortical midline structures in the same manner as seen for sentences describing objects in motion toward oneself. Moreover, we included a further experimental condition in which sentences describing objects in motion toward other people were presented (e.g., "The car drives toward Maria"). If the observed activation of the cortical midline structures are a reflection of self-referentiality we predict these structures to be active only in the Toward Self sentence, but not in the Toward Other sentences. However, there is recent evidence that the self vs. other distinction in action understanding might be less strict than previously thought (Rizolatti et al., 2001; Wilson and Knoblich, 2005). Under this assumption one might expect an overlap of brain activation for the Toward self and Toward other condition.

EXPERIMENT 2: REFERENCE TO SELF AND OTHERS IN THE PROCESSING OF MOTION SENTENCES

MATERIALS AND METHODS

Participants

Sixteen participants (eight male) aged 22–27 years (mean = 24.8) participated in this experiment after giving informed written consent.

Materials

Sentence material consisted of a total of 192 short sentences. All sentences were constructed in the active voice with a subject, a verb, and prepositional phrase (see examples in **Table 3**). Of the 192 sentences, 120 sentences constituted critical items. These belonged to one of five conditions: Movement Toward Self (MTS), Movement Toward Other (MTO), Movement Toward Thing (MTT), Movement Away from Self (MAS), and No Movement (NM). For each condition 24 sentences were constructed. The remaining 72 non-critical items comprised 24 catch trials (see below) and 48 filler sentences, which were of a form comparable to that of the critical items, but always described objects in motion away from various things. These were included in order to keep the number of occurrences of the words "toward" and "away" balanced.

Catch trials were designed in order to ensure that participants were listening carefully to sentence content. Catch trials were semantically anomalous sentences with the same syntactic form as critical items.

Table 3 | Examples of the sentences used in Experiment 2 with English translations.

Condition	Example sentence	English translation
MTS	Das Auto fährt auf Dich zu.	The car drives toward you.
MTA	Das Auto fährt von Dir weg.	The car drives away from you.
MTO	Das Auto fährt auf Maria zu.	The car drives toward Maria.
MTT	Das Auto fährt auf die Brücke zu.	The car drives toward the bridge.
NM	Das Auto sieht groß aus.	The car looks big.

The three visual stimuli created for Experiment 1 were used again. These constituted (1) a black and white spiral rotated in a clockwise direction to give the illusion of motion toward the participant (VIS_T); (2) a black and white spiral rotated in a counter-clockwise direction to give the illusion of motion away from the participant (VIS_A); and (3) a static image, which was created by scrambling the visual input of the spiral stimuli and did not create any illusion of motion (VIS_S). All images were 580 × 580 pixels.

Stimulus presentation

Participants lay in the scanner. Auditory stimuli were presented over headphones and visual stimuli were presented on a virtual monitor seen through 3D glasses. After being instructed and given practice trials, participants listened to all sentence stimuli in a single continuous block (approx. 33 min) and performed a sentence congruency task. Specifically, participants were instructed to respond by pressing a button whenever a semantically incongruent sentence was presented. In this manner, all critical trials remained free of motion artifacts. Following the sentence comprehension block, participants were shown visual stimuli for the purpose of localizing area MT/V5.

In the sentence comprehension block, each trial constituted presentation of a single sentence. The interstimulus interval was approximately 6 s. To enhance the temporal resolution of the acquired signal, a temporal jitter of 500, 1000, or 1500 ms. was inserted into the beginning of each trial.

FMRI data acquisition

Scanning was performed using a 3-T MedSpec 30/100 scanner (Bruker, Ettlingen, Germany) and a birdcage head coil. Twenty axial slices (4-mm thickness, 1 mm inter-slice distance, FOV 19.2 cm, data matrix of 64 × 64 voxels, in-plane resolution of 3 × 3 mm) were acquired every 2 s during functional measurements (BOLD sensitive gradient EPI sequence, TR = 2 s, TE = 30 ms, flip angle = 90°, acquisition bandwidth = 100 Hz) with a 3-Tesla Bruker Medspec 30/100 system. Prior to functional imaging T1-weighted MDEFT images (data matrix 256 × 256, TR = 1.3 s, TE = 10 ms) were obtained with a non-slice selective inversion pulse followed by a single excitation of each slice (Norris, 2000). These images were used to co-register functional scans with previously obtained high-resolution whole head 3D brain scans (128 sagittal slices, 1.5-mm thickness, FOV 25 × 25 × 19.2 cm, data matrix of 256 × 256 voxels) (Lee et al., 1995).

FMRI data analysis

In order to perform a motion correction, functional volumes were realigned and unwarped using SPM5. The further processing steps were performed using the software package LIPSIA (Lohmann et al., 2001). To perform a slicetime correction, a cubic–spline interpolation was applied.

In order to align the functional data with a 3D stereotactic coordinate reference system (Talairach and Tournoux, 1988), a linear registration was performed. The registration parameters were acquired using an anatomical reference brain. Hereafter, the registration parameters were used to transform the functional data set to the stereotactic coordinate system, by using a trilinear interpolation. The resulting voxel size was 3 × 3 × 3 mm.

After registration and normalization, a temporal highpass filter with a cutoff frequency of 1/90 Hz was used for baseline correction of the signal and a spatial Gaussian filter with 6 mm FWHM was applied. The statistical evaluation was based on a general linear regression with pre-whitening (Worsley et al., 2002). Specifically, autocorrelation parameters were estimated from the least-squares residuals using the Yule–Walker equations. These parameters were subsequently used to whiten both data and design matrix. Finally, the linear model was re-estimated using least squares on the whitened data to produce estimates of effects and their standard errors.

Whole brain analysis: localizer block

As noted before, each individual functional dataset was aligned with the standard stereotactic reference space, so that a group analysis based on the contrast-images could be performed. The single-participant contrast-images were entered into a second-level random effects analysis for each of the contrasts. The group analysis consisted of a one-sample *t*-test across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from 0. The results were corrected for multiple comparisons using cluster-size and cluster-value thresholds obtained by Monte–Carlo simulations using a significance level of $p < 0.05$ (clusters in the resulting maps were obtained using a Z-value threshold of 2.576).

Functional region of interest (ROI)

The contrast images were also used for regions of interest (ROI) analysis. We tested whether contrast values in ROIs were significantly different from 0 in each experimental condition. Masks for the left and the right MT/V5 were created using the functional images of the localizer run. Three other ROIs were determined using the results of Experiment 1: the orbitomedial prefrontal cortex (OMPFC), posterior cingulate (PC) and cuneus (CUN).

RESULTS

BEHAVIORAL

Participants showed high performance (93.75%, SD = 7.13) for detection of catch trials (i.e., semantically anomalous sentences), indicating that they were indeed listening to and processing the semantic content of sentences in general.

FUNCTIONAL REGION OF INTEREST (ROI)

Area MT/V5

In the left hemisphere the peak activation in MT/V5 was identified at −43, −68, 10. In the right hemisphere the peak coordinate was at 43, −64, 8 (see **Figure 1**). Planned paired samples *t*-tests showed that in bilateral MT/V5 sentences describing objects in motion toward the participant as well as toward others and toward inanimate objects elicited reliably more activation than sentences describing static scenes (Left MT/V5: MTS-NM: $t(15) = 2.35, p < 0.05$; MTO-NM: $t(15) = 2.18, p < 0.05$); MTT-NM: $t(15) = 3.69$, $p = 0.001$; Right MT/V5: MTS-NM: $t(15) = 3.23$, $p < 0.01$; MTO-NM: $t(15) = 2.47, p < 0.05$); MTT-NM: $t(15) = 3.36$, $p = 0.005$.). Sentences describing objects in motion away from participants showed no reliably different activation than sen-

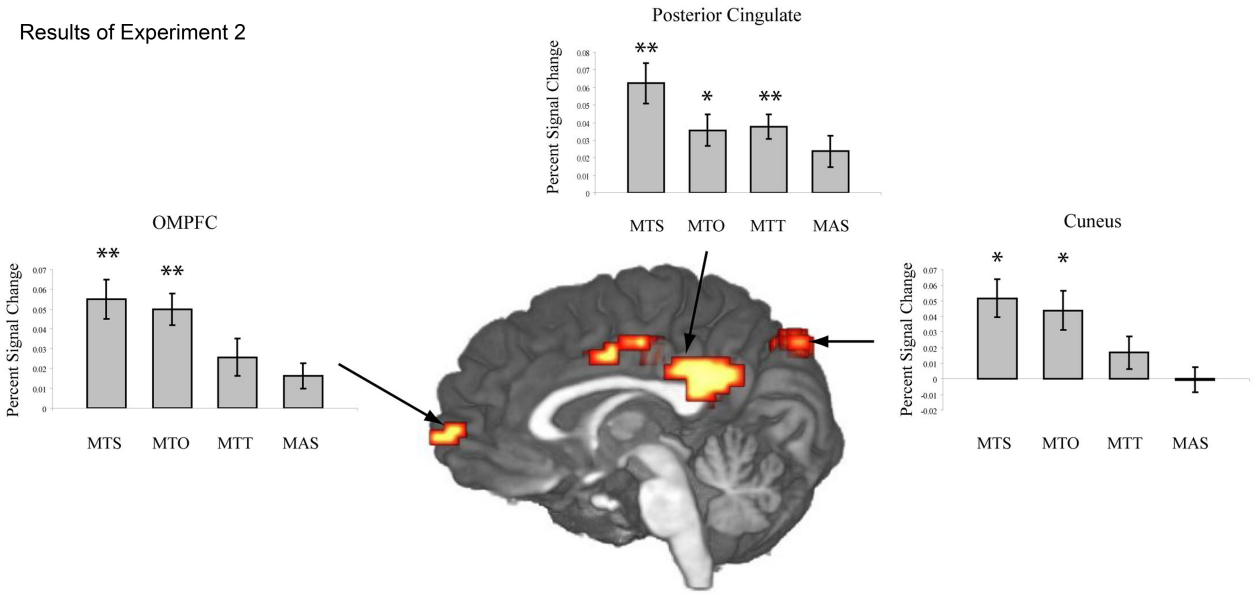


FIGURE 2 | The *Z*-map in the center shows the results of the whole brain analysis from Experiment 1. Greater levels of activation are seen along the cortical midline in the orbitomedial prefrontal cortex (OMPFC), posterior cingulate (PC) and cuneus (CUN) for sentences describing objects in motion toward oneself vs. sentences describing static objects. The histograms show mean activation (percent signal change) within these regions for the experimental sentence conditions presented in Experiment 2: MTS = movement toward self, MTO = movement toward other, MTT = movement toward thing, MAS = movement away from self. Activation in the cortical midline structures is significant for both MTS and MTO sentences, but not for MTT or MAS sentences. A single asterisk indicates a difference significant at the 0.05 level, and two asterisks indicate a difference significant at the 0.01 level.

tences describing static scenes in either hemisphere (Left MT/V5: MAS-NM: $t(15) = 0.28$, $p > 0.1$. Right MT/V5: MAS-NM: $t(15) = 1.23$, $p > 0.1$).

CORTICAL MIDLINE STRUCTURES

In the OMPFC reliably more activation was seen for sentences describing objects moving toward participants and other people than for sentences describing static scenes (MTS-NM: $t(15) = 2.82$, $p < 0.05$; MTO-NM: $t(15) = 3.11$, $p < 0.01$). Sentences describing objects in motion toward inanimate objects and objects in motion away from participants did not activate OMPFC differently than sentences describing static scenes (MTT-NM: $t(15) = 1.37$, $p > 0.1$; MAS-NM: $t(15) = 1.21$, $p > 0.1$).

In PC reliably greater activation was observed for sentences describing objects moving toward the participant, toward another person and toward inanimate objects than for sentences describing static scenes (MTS-NM: $t(15) = 2.7$, $p < 0.01$; MTO-NM: $t(15) = 1.95$, $p < 0.05$; MTT-NM: $t(15) = 2.69$, $p < 0.01$). Sentences describing objects in motion away from participants did not activate PC reliably differently than sentences describing static scenes (MAS-NM: $t(15) = 1.28$, $p > 0.1$).

In the cuneus significantly more activation was recorded for sentences describing objects moving toward the participant and toward another person than for sentences describing static scenes (MTS-NM: $t(15) = 2.15$, $p < 0.05$; MTO-NM: $t(15) = 1.75$, $p = 0.050$). There was no reliably different activation for sentences describing objects in motion toward inanimate objects or objects in motion away from participants when compared with sentences describing static scenes (MTT-NM: $t(15) = 0.79$, $p > 0.1$; MAS-NM: $t(15) = -0.02$, $p > 0.1$).

DISCUSSION

The goal of Experiment 2 was twofold. First, we asked whether more conclusive results regarding the role of MT/V5 in processing sentences describing objects in motion could be obtained with more participants and an experimental design separating visual and language stimuli. Secondly, we tested the hypothesis that sentences describing events necessitating reassignment of visual attention and potential modification of behavior to avoid danger elicit more activation than non-threatening sentences in the cortical midline structures (i.e., as seen in Experiment 1).

Using a functional localizer, MT/V5 was identified bilaterally in the lateral posterior middle temporal gyrus. The co-ordinates are very close to those reported in Experiment 1 and also to those reported in previous literature (Tootell et al., 1995a). Within bilateral MT/V5, activation was modulated by sentence content, such that sentences describing objects in motion toward oneself (MTS: "The car drives toward you") elicited more activation than sentences describing static events (NM "The car looks big"). Activation in MT was not modulated by sentences describing objects in motion away from oneself (MAS: "The car drives away from you"). This suggests that MT/V5 becomes active in comprehending sentences describing approaching but not receding objects. The implications of this are discussed further below.

Cortical midline structures (OMPFC, PC, and cuneus) were reliably more activated for sentences describing objects in motion toward participants than for sentences describing objects in motion away from participants. As discussed previously, the cortical midline structures have been implicated in (1) the processing of self-referential information (particularly OMPFC) and (2) the guiding of visual

attention to support a change in behavior (in particular PC). Thus the results suggest that comprehending self-referential sentences describing an event that would require reallocation of the participant's visual attention (i.e., to detect the approaching object) and a potential change in the participant's subsequent behavior (i.e., to move out of the way of the approaching object) activate those cortical networks known to support execution of these tasks in a natural setting.

In support of this interpretation, sentences describing objects in motion toward inanimate objects (MTT: "The car approaches the bridge.") did not reliably activate the OMPFC. MTT sentences elicited reliable modulation of activation in bilateral MT/V5 and in PC only. These findings are consistent with a network responsible for guiding visual attention toward a moving object (Small et al., 2003), but this network has little overlap with those areas thought to be important in the processing of self-referential information. It is interesting to note here that sentences describing objects in motion toward other things in general (i.e., both toward the self and another object) activate a cortical network involved in the guidance of visual attention and perception of moving stimuli, but sentences describing objects in motion *per se* (e.g., sentences describing objects in motion away from participants) do not. We speculate that in all "toward" conditions, participants may understand a potential need to modify behavior based on the visual scene described. In the case that an object is approaching the participant or an inanimate object, the participant should ultimately change his behavior to avoid the imminent collision. In the case that an object is moving away from the participant, there is no need to modify behavior. Indeed some converging evidence can be found in behavioral studies that argue that looming (i.e., approaching) objects are processed with priority over objects with other motion trajectories (Lin et al., 2008).

Comprehension of sentences describing objects in motion toward other people (MTO: "The car drives toward Mary") elicited activation in those areas also seen to be activated for MTS sentences (i.e., MT/V5, OMPFC, PC, and cuneus). We included this sentence condition in order to test whether activation in the cortical midline structures reflects only self-referential sentence content. The finding that we see more activation in these areas for both self-referential and other-referential sentences suggests that sentences describing objects in motion toward other human beings may be of equal importance to the subsequent behavior of the participant than those sentences describing objects moving toward oneself. The importance may be twofold: the sentence "The car approaches Mary" could either be a warning to a participant that he/she should attempt to avoid hitting Mary with the car or such a sentence could be processed in a manner very similar to those describing objects in motion toward oneself due to the participant's mental representation of the other. In both cases the action plan of the participant must be modified in order to prevent harm to him/herself or the other. Therefore we suggest that activation in OMPFC and cuneus observed during the comprehension of sentences describing objects in motion toward the participant and toward another person reflects awareness that sentence content has potential personal consequences.

GENERAL DISCUSSION

There are several important findings from this work. First, using a functional localizer, MT/V5 was identified bilaterally in the lateral posterior middle temporal gyrus in both experiments. The co-ordinates obtained in both studies are very close to those reported in previous literature (Tootell et al., 1995a). Within bilateral MT/V5, activation was modulated by sentence content, such that sentences describing objects in motion toward oneself elicited more activation than sentences describing static events. Whereas activation difference in right MT/V5 only approached statistical significance in Experiment 1, the results of Experiment 2 showed a statistically reliable difference bilaterally. Activation in MT was not modulated by sentences describing objects in motion away from oneself. Taken together, we provide good evidence that sentences describing objects in motion toward participants activate bilateral MT/V5, while sentences describing objects in motion away from participants do not.

Our data are broadly consistent with those reported by Saygin et al. (2010) who also found greater activation in the MT area for sentences conveying motion compared to non-motion sentences. However, Saygin et al. did not report any difference between sentences describing self-motion and those referring to motion of objects and animals. A notable difference between the methodologies used here and in Saygin et al. is that their stimuli were audio-visual, that is, a video tape of a speaker uttering the sentences. Presenting the stimuli in a format closer to a conversational setting may have encouraged greater simulation (in preparation to respond to the speaker). This speculation is consistent with our claim that degree of simulation is subject to top-down modulation.

Another key finding from this work is that brain areas along the median wall, known to be involved in the processing of self-referential stimuli and the direction of attention, are modulated by the self-referential content of experimental sentences. In both experiments sentences describing objects in motion toward the listener elicited greater activation in OMPFC, PC and cuneus than sentences describing static scenes. This pattern was not seen for sentences describing objects in motion away from the listener (Experiments 1 and 2) or for sentences describing objects in motion toward inanimate things (Experiment 2). To summarize, in all sentence conditions in which the propositional content of the sentence could be construed as relevant to the self, activation was elicited in OMPFC and posterior medial parietal cortex. We suggest that these sentences are perceived as more relevant for the participant than sentences describing static scenes or scenes in which the participant's potential actions play no role.

IMPLICATIONS FOR AN EMBODIED APPROACH TO LANGUAGE COMPREHENSION

The results from the two experiments are partially consistent with the predictions derived from an embodied approach to language comprehension. That is, if language comprehension requires the simulation of sentence content using neural systems also used for perception, action, and emotion, then we should expect activation of MT/V5 during processing of language about visual motion. Indeed we found that activation for sentences describing movement toward oneself, another person, or an object all activated area MT/V5 significantly. However, sentences describing objects in motion away from oneself had no effect on activation level in MT/V5. Thus, although an selective involvement of visual brain areas is seen for the processing of sentences with visual motion content, it cannot be verified that sentences describing objects in motion *per se* activate

MT/V5 in the same manner that actual visual stimuli would. This indicates that higher level visual areas become involved in sentential processing, but only at a late stage and only after sentence meaning has already been at least partially derived.

Our data therefore suggest that some modification of the strong embodiment position that all sentence content is simulated during language comprehension is in order. This notion is in accordance with other research indicating that the simulation view of language comprehension may not be as straightforward as initially thought (e.g., Masson et al., 2008). Specifically, Masson et al. (2008) showed in a behavioral priming study that the type of action-information simulated during comprehension of language denoting manipulable objects differs depending on the sentence context surrounding the manipulable object word. Specifically, they showed that words presented in isolation prime the execution of hand movements related to the functional use of the denoted object (e.g., calculator primes a finger poking action). The same words do not prime execution of hand movements related to moving objects (e.g., calculator does not prime a manual horizontal grasp). However, if manipulable object words are embedded in sentences in which the physical properties of the object become more relevant than the function of the object (e.g., The lawyer kicked the calculator aside.), then the priming effect for functional actions is compromised, while a priming effect for movement actions becomes apparent. This indicates that sensory-motor representations of lexical items are constructed *during* sentence comprehension. Motor simulation is suggested to draw on whatever experience best captures the relevant properties of the denoted object in the situation in which it is denoted (e.g., the physical properties of the calculator become more relevant than functional properties in understanding what it means to kick a calculator). There is thus top-down modulation of what kind of action to simulate in conjunction with a manipulable object word. Research to date leaves open what processes underlie this top-down modulation; however the current data suggest that relevance of the described situation for the self is one possible option.

CONCLUSION

In two fMRI experiments we tested whether area MT/V5, a visual processing area known to be sensitive to moving visual stimuli, is activated by sentences describing objects in motion. The results of our studies show that sentences describing objects in motion toward other entities (people and things) indeed activate higher-order visual brain areas. The results cannot be reconciled fully with a strict embodied interpretation, because sentences describing receding objects did not elicit the same pattern of results in MT/V5. In addition to MT/V5, sentences describing objects in motion toward oneself or another person elicited increased levels of activation along the cortical midline structure. We suggest that this reflects the potential relevance of the information conveyed by the sentence for oneself. The results extend the literature on embodied language processing by showing that higher-order visual areas are also involved in language processing, but they leave open questions concerning the timing with which these areas become activated during sentence comprehension.

REFERENCES

Beauchamp, M. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr. Opin. Neurobiol.* 15, 145–153.

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., and Saxe, R. (2008). Concepts are more than percepts: the case of action verbs. *J. Neurosci*. 28, 11347–11353.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res.* 24, 355–363.

Forman, S., Cohen, J., Fitzgerald, M., Eddy, W., Mintun, M., and Noll, D. (1995). Improved assessment of significant activation in functional magnetic resonance imaging: use of a cluster-size threshold. *MRM* 33, 636–647.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage* 7, 30–40.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61, 905–919.

Goebel, R., Sefat, D., Muckli, L., Hacker, H., and Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur. J. Neurosci.* 10, 1563–1573.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Johnson, C., Baxter, L., Wilder, L., Pipe, J., Heiserman, J., and Prigatano, G. (2002). Neural correlates of self-reflection. *Brain* 125, 1808–1814.

Josephs, O., Turner, R., and Friston, K. (1997). Event-related fMRI. *Hum. Brain Mapp*. 5, 243–248.

Kable, J., Kan, I., Wilson, A., Thompson-Schill, S., and Chatterjee, A. (2005). Conceptual representations of action in the lateral temporal cortex. *J. Cogn. Neurosci.* 17, 1855–1870.

Kable, J., Lease-Spellmeyer, J., and Chatterjee, A. (2002). Neural substrates of action event knowledge. *J. Cogn. Neurosci.* 14, 795–805.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A., and Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition* 94, B79–B89.

Kelley, W., Macrae, C., Wyland, C., Caglar, S., Inati, S., and Heatherton, T. (2002). Finding the self? An event-related fMRI study. *J. Cogn. Neurosci*. 14, 785–794.

Kjaer, T., Nowak, M., and Lou, H. (2002). Reflective self-awareness and conscious states: PET Evidence for a common midline parietofrontal core. *NeuroImage* 17, 1080–1086.

Lee, J., Garwood, M., Menon, R., Adriany, G., Andersen, P., Truwit, C., and Ugurbil, K. (1995). High contrast and fast three dimensional magnetic resonance imaging at high fields. *Magn. Reson. Med*. 34, 308–312.

Lin, J., Franconeri, S., and Enns, J. (2008). Objects on a collision path with the observer demand attention. *Psychol. Sci.* 19, 686–692.

Lohmann, G., Mueller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., and von Cramon, D. Y. (2001). Lipsia – a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comput. Med. Imaging Graph*. 25, 449–457.

Masson, M., Bub, D., and Warren, C. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *J. Mem. Lang*. 59, 256–265.

Noppeney, U., Friston, K., and Price, C. (2003). Effects of visual deprivation on the organization of the semantic system. *Brain* 126, 1620–1627.

Norris, D. (2000). Reduced power multislice MDEFT imaging. *J. Magn. Reson. Imaging*. 11, 445–451.

Northoff, G., and Bermpohl, F. (2004). Cortical midline structures and the self. *Trends Cogn. Sci.* 8, 102–107.

Ongur, D., and Price, J. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206–219.

Pirog Revill, K., Aslin, R., Tanenhaus, M., and Bavelier, D. (2008). Neural correlates of partial lexical activation. *Proc. Natl. Acad. Sci.* 105, 13111–13115.

Rizolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.

Rueschemeyer, S.-A., Brass, M., and Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor stems. *J. Cogn. Neurosci.* 19, 855–865.

Saxe, R., Brett, M., and Kanwisher, N. (2006). Divide and conquer: a defense of functional localizers. *NeuroImage* 30, 1088–1096.

Saygin, A., McCullough, S., Alac, M., and Emmorey, K. (2010). Modulation of BOLD response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *J. Cogn. Neurosci.* 22, 2480–2490.

Small, D., Gitelman, D., Gregory, M., Nobre, A., Parrish, T., and Mesulam, M. (2003). The posterior cingulate and medial prefrontal cortex mediate the anticipatory allocation of spatial attention. *Neuroimage* 18, 633–641.

Smith, A. T., Greenlee, M. W., Singh, K. D., Kraemer, F. M., and Hennig, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *J. Neurosci.* 18, 3816–3830.

Talairach, P., and Tournoux, J. (1988). *A Stereotactic Coplanar Atlas of the Human Brain.* Stuttgart: Thieme.

Taylor, K., Seminowicz, D., and Davis, K. (2009). Two systems of resting state connectivity between the insula and cingulate cortex. *Hum. Brain Mapp.* 30, 2731–2745.

Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.* 2, 243–260.

Tootell, R., Reppas, J., Dale, A., Look, R., Sereno, M., Malach, R., Brady, T., and Rosen, B. (1995a). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* 375, 139–141.

Tootell, R., Reppas, J., Kwong, K., Malach, R., Born, R., Brady, T, Rosen, B., and Belliveau, J. (1995b). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15, 3215–3230.

Vogt, B. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nat. Rev. Neurosci.* 6, 533–544.

Wallentin, M., Østergaarda, S., Lund, T., Østergaard, L., and Roepstorff, A. (2005). Concrete spatial language: see what I mean? *Brain Lang.* 92, 221–233.

Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–473.

Worsley, K., and Friston, K. (1995). Analysis of fMRI time-series revisited – again. *NeuroImage* 2, 173–181.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., and Evans, A. (2002). A general statistical analysis for fMRI data. *Neuroimage* 15, 1–15.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 July 2010; accepted: 09 October 2010; published online: 05 November 2010.
Citation: Rueschemeyer S, Glenberg AM, Kaschak MP, Mueller K and Friederici AD (2010) Top-down and bottom-up contributions to understanding sentences describing objects in motion. Front. Psychology ***1****:183. doi: 10.3389/fpsyg.2010.00183*
This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.
Copyright © 2010 Rueschemeyer, Glenberg, Kaschak, Mueller and Friederici. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 04 October 2010
doi: 10.3389/fpsyg.2010.00150

Context effects in embodied lexical-semantic processing

Wessel O. van Dam*, Shirley-Ann Rueschemeyer, Oliver Lindemann and Harold Bekkering

Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Netherlands

Edited by:
Anna M Borghi, University of Bologna, Italy

Reviewed by:
Veronique Boulenger, Centre National de la Recherche Scientifique, France
Rutvik Desai, Medical College of Wisconsin, USA

****Correspondence:***
Wessel O. van Dam, P.O. Box 9104, 6500 HE Nijmegen, Netherlands.
e-mail: w.vandam@donders.ru.nl

The embodied view of language comprehension proposes that the meaning of words is grounded in perception and action rather than represented in abstract amodal symbols. Support for embodied theories of language processing comes from behavioral studies showing that understanding a sentence about an action can modulate congruent and incongruent physical responses, suggesting motor involvement during comprehension of sentences referring to bodily movement. Additionally, several neuroimaging studies have provided evidence that comprehending single words denoting manipulable objects elicits specific responses in the neural motor system. An interesting question that remains is whether action semantic knowledge is directly activated as motor simulations in the brain, or rather modulated by the semantic context in which action words are encountered. In the current paper we investigated the nature of conceptual representations using a go/no-go lexical decision task. Specifically, target words were either presented in a semantic context that emphasized dominant action features (features related to the functional use of an object) or non-dominant action features. The response latencies in a lexical decision task reveal that participants were faster to respond to words denoting objects for which the functional use was congruent with the prepared movement. This facilitation effect, however, was only apparent when the semantic context emphasized corresponding motor properties. These findings suggest that conceptual processing is a context-dependent process that incorporates motor-related knowledge in a flexible manner.

Keywords: embodiment, semantics, action, conceptual flexibility

INTRODUCTION

According to an embodied view of language comprehension, language concepts are grounded in motor and perceptual systems (Glenberg, 1997; Barsalou, 1999; Pulvermueller, 1999; Barsalou et al., 2003). Lexical-semantic representations are postulated to rely on sensori-motor brain areas and to reflect real-world experience with words' referents. For example an object such as a hammer is experienced by most people visually (i.e., we know what hammers look like) and motorically (i.e., we know what it feels like to wield a hammer). Thus, embodied language theories postulate that upon encountering the word *hammer*, experiential traces stored in modality specific (e.g., visual and motor) brain areas are activated. Importantly, these activations are seen to contribute to the lexical-semantic meaning of the word form *hammer*. This view differs fundamentally from symbolic (disembodied) accounts, in which conceptual representations are symbolic and amodal, and lexical-semantic meaning independent of real-world experience (Kintsch, 2008).

Converging evidence for the idea that sensory-motor brain areas are involved in language comprehension comes from behavioral, electrophysiological, and neuroimaging studies. Glenberg and Kaschak (2002) present the action-sentence compatibility effect (ACE), in which reading a sentence that implies an action toward or away from the body (e.g., *open the drawer/close the drawer*) facilitates a congruent action (i.e., moving a hand toward or away from the body). The authors argue that responses are facilitated because comprehending language about action recruits the same neural resources as required for action execution; thus comprehending the sentence primes a congruent motor act. Zwaan and Taylor (2006) have shown that motor resonance is activated on-line during sentence comprehension. For example, while reading the sentence *The marathon runner opened the water bottle*, evidence for motor resonance is seen in conjunction with presentation of the verb *opened*. Furthermore motor resonance is elicited by words which unambiguously specify kinematic properties of actions, even if these words are not verbs (Taylor et al., 2008). For example, in the text *He looked at the pie and turned the oven dial. The baking time needed to be shorter/longer* evidence for motor resonance is seen in conjunction with the sentence final adjective (which specifies how the oven dial should be turned) rather than in conjunction with the verb *turned*.

In addition to words and sentences describing active events, isolated words denoting manipulable objects have also been shown to interact with motor behaviors and to activate neural motor areas. For example, Glover et al. (2004) show that words denoting objects that afford particular actions are sufficient to activate motor representations. Participants in their study were primed with the names of objects of different sizes (e.g., *apple*, *grape*) and required to reach out and grasp a wooden block. Interestingly, the maximum grip aperture during the reach (i.e., maximum distance between forefinger and thumb) was influenced by the size of the object denoted by the prime word, despite the fact that this information was irrelevant for executing the grasping task. Similarly, Rueschemeyer et al. (2010a) had participants perform a lexical decision task to words denoting objects typically brought toward or away from the body for functional use (e.g., *cup* or *key*, respectively). The authors found that responses to words were facilitated if the required response action was congruent to the

action typically performed on the word's referent (i.e., responses to *cup* were faster if a movement was made toward the body rather than away from the body). This indicates that very specific information about how an object is manipulated is activated during lexical retrieval.

Evidence in favor of embodied theories of language also comes from recent neuroimaging studies. Specifically, the comprehension of action verbs (Kemmerer et al., 2008), action sentences (Desai et al., 2010), and words denoting manipulable objects (Chao and Martin, 2000; Saccuman et al., 2006; Rueschemeyer et al., 2010b) all reliably activate the cerebral motor system. Furthermore, embodied lexical-semantic representations activate the neural motor system in a somatotopic manner (Hauk et al., 2004; Tettamanti et al., 2005; Aziz-Zadeh et al., 2006; but see also de Zubicaray et al., 2010). Findings from the studies reviewed above provide strong evidence for functional links between the neural motor system and lexical-semantic processing of words that entail a motor component.

It has been suggested that embodied lexical-semantic representations are fast, automatic and invariant. Pulvermueller et al. (2000) demonstrated that category-specific activation can be observed as early as about 200 ms after word onset (see also Hauk and Pulvermueller, 2004), and occurs irrespective of attention to action words (Shtyrov et al., 2004; Pulvermueller et al., 2005). Pulvermueller proposes that strong functional links between language and motor systems have developed as a consequence of the fact that actions and their referents often co-occur near-simultaneously. Specifically, the action and the word co-occur frequently, and thereby, neural populations recruited for processing a word form and those involved in processing the referent body movement frequently fire together and become strongly linked (Pulvermueller, 1999, 2001). Due to the strong within-assembly connections that link language and action representations, action word recognition will thereby automatically and invariably trigger activation in specific action-related networks.

However, in contrast to what would be expected if embodied representations are indeed automatic and invariant, several studies have failed to find motor-related activity for words with an action-semantic component (Rueschemeyer et al., 2007; Raposo et al., 2009). Raposo et al. (2009) showed in a recent fMRI study that action verbs in isolation (e.g., *kick*) or in literal sentences (e.g., *kick the ball*) elicit a response in motor/premotor cortices. Action verbs in an idiomatic context (e.g., *kick the bucket*), however, did not elicit such activations. These findings strongly challenge the automaticity of motor-related activity for action words and rather suggests that the activation of meaning attributes of words is a flexible and contextually dependent process (but see also Boulenger et al., 2009). In a similar vein, Rueschemeyer et al. (2007) showed that processing of morphologically complex verbs built on motor stems showed no differences in involvement of the motor system when compared with processing complex verbs with abstract stems. For example, the difference between the verb begreifen (*to comprehend*) vs. bedenken (*to think*). In the first case the morphologically complex verb is a prefixed form of the simple motor verb *greifen* (to grasp), whereas in the latter case it is a prefixed form of the abstract verb *denken* (to think). All these studies provide evidence that motor responses rather than being automatic and invariable, depend on the context (i.e., sentence context in the case of Raposo et al., 2009, and morphological context in the case of Rueschemeyer et al., 2007) in which action words are encountered. That is, a crucial factor for observing activity in motor and premotor regions during action word processing seems to be that the context emphasizes motor properties (suggesting that representations are flexible).

The failure of some studies to report motor activation for action-related words embedded in various contexts already speaks against a strictly automatic interpretation, and suggests that motor activations may be called on in a flexible manner during word processing. In the two studies cited above, motor information is not useful in comprehending the given language utterances (i.e., idiomatic phrases or morphologically complex words), and indeed the results indicate that in these cases the motor system is not reliably activated. However, even in cases in which motor information *is* helpful in processing semantic content, a certain degree of flexibility on the level of specific motor programs might be expected. For example, the motor programs associated with the word *cup* are quite different in the sentences *She filled her cup at the tap/She drank from the cup*. In the first case, the actor in the sentence brings the cup away from her body and toward the tap, in the second she brings the cup toward her mouth. Previous research has shown that words referring to manipulable objects activate motor areas (Chao and Martin, 2000; Saccuman et al., 2006; Rueschemeyer et al., 2010a), but since most objects can be used in multiple ways, does the motor contribution to lexical-semantic comprehension also vary depending on the language context in which a word is presented?

In the current study, we investigated precisely this issue: within a given modality (i.e., action) we investigated whether a word always activates a specific motor program, or whether the motor program activated depends on the context in which a word is presented. Previously we (Rueschemeyer et al., 2010a) showed that comprehension of words denoting manipulable objects (e.g., *cup*, *hammer*) are facilitated by the prior planning of an action congruent to the prototypical use of the object. In other words, participants were faster to respond to the word *cup* when they had planned an action toward their body (the most common use of the cup being a vessel to drink from) than when they had planned an action away from their body; the opposite pattern was true for words denoting objects typically brought away from the body (e.g., *hammer*). We thus show a strong congruency effect between the typical action associated with a word's referent and processing of the individual word. The target words in this previous study were presented in isolation, i.e., without any explicit language context. In the current study we investigated whether embedding the same words in a lexical context suggesting a non-prototypical (but not unfamiliar) use of the objects would influence the observed congruency effects. To this end object words were always preceded by a prime word providing a context that emphasized either the dominant action feature (*thirst* – ***cup***) or a non-dominant action feature (*sink* – ***cup***). As previous studies have shown that the retrieval of a particular conceptual feature depends on the context in which a word is encountered, we hypothesized that the effect of movement preparation on word processing would interact with the

semantic context in which a word is encountered. Specifically, in trials in which the semantic context emphasizes the dominant motor properties of a concept, we expect faster onset latencies to words denoting objects for which the functional use is congruent with the prepared movement in directionality. Furthermore, we expect that this facilitation effect is not present if the semantic context emphasizes motor properties of a concept that are not related to the object's functional use.

EXPERIMENT

METHOD

Participants

Thirty-five subjects participated; the average age was 21.3 years. All subjects were students at the Radboud University Nijmegen and participated either for money or course credit. No subject was aware of the purpose of the experiment.

Stimuli

Words were presented in white Arial fonts on a black background. The average word length was 10 letters. For a viewing distance of 100 cm, the stimuli subtended on average a visual angle of 0.57° × 2.86°. A total of 120 letter string stimuli were created for the experiment (stimuli with English translations can be seen in **Table 1**). Eighty were real Dutch words denoting familiar objects and comprised the critical experimental stimuli. The remaining 40 stimuli were Dutch pseudowords (i.e., phonotactically and orthographically legal letter strings with no meaning in Dutch). Target stimuli belonged to one of two experimental conditions: (1) words denoting objects for which the functional use is associated with a movement toward the body (e.g., *telephone*, *photo camera*), and (2) words denoting objects for which the functional use is associated with a movement away from the body (e.g., *hammer*, *pencil*). Target stimuli were presented in two contexts: (1) target words were preceded by a word that emphasized the action feature related to the functional use (e.g., *conversation-telephone*, *nail-hammer*), and (2) target words were preceded by a word that emphasized an action feature not related to the functional use (e.g., *adapter-telephone*, *tool belt-hammer*).

To test that stimuli were truly matched with regards to important psycholinguistic variables, a questionnaire was administered to 15 native Dutch speakers who did not participate in the behavioral experiment (see **Table 2** for results). In this questionnaire, participants were asked to rate words on a 7-point scale with respect to (1) the imageability of the noun (1 = very difficult to imagine the referent noun, 7 = very easy to imagine the referent noun), (2) whether the noun denoted an object that you typically bring toward or away from the body (−3 = toward the body, +3 = away from the body).

The results of the questionnaire showed that nouns were matched across the two conditions with respect to imageability (body: $M = 6.82$, SE = 0.024; world: $M = 6.76$, SE = 0.029), $t(1,78) = 1.46$, $p > 0.1$. For object nouns, participants agreed that body words referred to objects that you typically bring toward the body, world words referred to objects you typically bring away from the body (body: $M = -1.13$, SE = 0.049; world: $M = 2.33$, SE = 0.057), both means significantly differed from 0 as indicated by one-sample *t*-tests (all *p*-values <0.001). In order to obtain an objective measure of word frequency, we calculated the mean lemma frequency per million for each condition using the lexical database (Baayen et al., 1993). This gave a mean of 567 (SE = 128.3) for the body words and 487 (SE = 119.3) for the world words. An independent sample *t*-test indicated words were matched on frequency, $t(1,78) = 1.48$, $p > 0.1$. Additionally, independent sample *t*-tests indicated that nouns were matched with regard to length (body: $M = 6.8$, world: $M = 6.3$), $t(1,78) = 0.90$, $p > 0.2$.

Table 1 | *Dutch* words associated with a movement toward the body (body words) and a movement away from the body (world words).
English translations are printed in italics.

Body words		World words	
Haarband	*Hair ribbon/Hairband*	Zwabber	*Swab/Mob*
La	*Drawer*	Spaarpot	*Money box*
Sjaal	*Scarf*	Spade	*Spade/Shovel*
Microfoon	*Microphone*	Zaag	*Saw*
Loep	*Magnifying glass*	Vaas	*Vase*
Hoed	*Hat*	Naald	*Needle*
Nagelvijl	*Nail file*	Kaars	*Candle*
Pleister	*Band-Aid*	Plant	*Plant*
Fluit	*Flute*	Flesopener	*Bottle opener*
Bril	*(pair of) Spectacles*	Deegroller	*Rolling pin*
Wijnglas	*Wine glass*	Koekenpan	*Frying pin*
Mok	*Mug*	Voetbal	*Football*
Make-up	*Make up*	Theepot	*Teapot*
Zakdoek	*Handkerchief*	Speld	*Pin*
Lolly	*Lollipop*	Stempel	*Stamp*
Halssnoer	*Necklace*	Sleutel	*Key*
Helm	*Helmet*	Lamp	*Lamp*
Telefoon	*Telephone*	Schep	*Scoop/Shovel*
Shampoo	*Shampoo*	Knikker	*Marble*
Armband	*Bracelet*	Bijl	*Axe*
Tondeuse	*(pair of) Clippers*	Baksteen	*Brick*
Mondharmonica	*Mouth organ*	Fakkel	*Torch*
Want	*Toothbrush*	Bel	*Bell*
Tandenborstel	*Mitten*	Hamer	*Hammer*
Handdoek	*Towel*	Computer	*Computer*
Ring	*Ring*	Hengel	*Fishing rod*
Trompet	*Trumpet*	Pen	*Pen*
Schoen	*Shoe*	Boor	*Drill*
Lippenstift	*Lipstick*	Gloeilamp	*Lightbulb*
Lepel	*Spoon*	Trommel	*Drum*
Oorbel	*Earring*	Kapstok	*Coat rack*
Borstel	*Brush*	Verfpot	*Pot/Tin of paint*
Verrekijker	*Binoculars*	Vergiet	*Stainer/Colander*
Stropdas	*Tie/Necktie*	Karaf	*Decanter/Carafe*
Gordel	*Belt*	Garde	*Whisk*
Vork	*Fork*	Paraplu	*Umbrella*
Parfum	*Perfume*	Mes	*Knife*
Horloge	*Watch*	Potlood	*Pencil*
Jas	*Jacket*	Ventilator	*Fan*
Fototoestel	*Photo camera*	Dobbelsteen	*Dice*

Procedure and design

Participants were seated comfortably in front of a computer monitor, and responded by means of a key press (i.e., by pressing either a key that was located *nearer* or *further* from the body). To start a trial participants had to press the start button of a response device (located in the middle of the response device). Subsequently, they received a cue (i.e., the letters A or B) that signaled them to prepare a movement (either *toward* or *away* from the body) which they only executed if the second word was lexically valid. Participants were instructed to read both words carefully. The advent of the two words was signaled by a fixation cross (appearing centrally for 500 ms). The first word was presented for 1000 ms. The second word appeared 1000 ms after the offset of the first word, calling for a speeded response to the identity of the word (i.e., a response in the case that the second word was a real word in Dutch). The second word remained visible until participants responded, or for a maximum of 2000 ms (see **Figure 1** for an illustration of a Go and NoGo trial). That is, the first presented word was not clearly associated with a specific direction or action, the differences we report are relative differences between reaction times (RTs) acquired in response to the second presented word. The experiment comprised 160 critical trials composed by 20 replications of the factorial combination of two movement cues, two word types, and two contexts. That is, the direction of the participant's response and the motor program generally associated with the word's referent either corresponded or not (action congruent vs. action incongruent) and the word was either presented in a language context highlighting the functional use or a less typical use of the object (focus on dominant vs. non-dominant action feature). The order in which word pairs were presented was counterbalanced over participants. That is, half of the participants first saw a target word (e.g., *telephone*) preceded by a word that emphasized the dominant action feature (*conversation*) and then the same target word preceded by a word that emphasized a non-dominant action feature (*adapter*) and vice versa.

Data analysis

We measured the latencies to recognize that a presented Dutch word was lexically valid defined as the time difference between word onset and release of the start button. Average RTs can be seen in **Figure 2**. Additionally, we recorded movement times (MTs; i.e., the time from releasing the start button until depressing the target button). Trials with extreme RTs or MTs ($2 \times$ STD $\pm$ mean) were treated as outliers and excluded from further analysis. This led to an exclusion of 9.12% of the data. The significance criterion for all analyses was set to $\alpha = 0.05$.

RESULTS

The error rates in the lexical decision task were on average lower than 2% and therefore not further analyzed. RTs were averaged for each participant in each condition (see **Table 3** for means) and submitted to a two-way repeated measures analysis of variance (ANOVA) with factors Action Congruency (congruent vs. incongruent) and Context (focus on dominant vs. non-dominant action feature). The main effect for Action Congruency and Context did not reach significance, both *p*-values >0.2. Importantly, in line with our hypothesis, the ANOVA revealed a significant interaction $F(1,34) = 5.76$, $p < 0.05$, $\eta_p^2 = 0.145$, indicating that the RT differences between action congruent and action incongruent words were modulated by the different context conditions. To further explore this interaction, we calculated *post hoc* paired sample *t*-tests. When the context emphasized the dominant action feature, words congruent with the prepared action were processed faster than incongruent words, $t(34) = -2.21$, $p < 0.05$. However, if the context focused on non-dominated action features, the mean RTs to action congruent and action incongruent words did not differ statistically, $t(34) < 1$, and the RT pattern reversed descriptively.

Movement times were averaged for each participant in each condition and submitted to a two-way repeated measures ANOVA. The main effect for Action Congruency and Context did not reach significance, both *p*-values >0.05. Furthermore, the ANOVA did not reveal a significant interaction ($p > 0.5$).

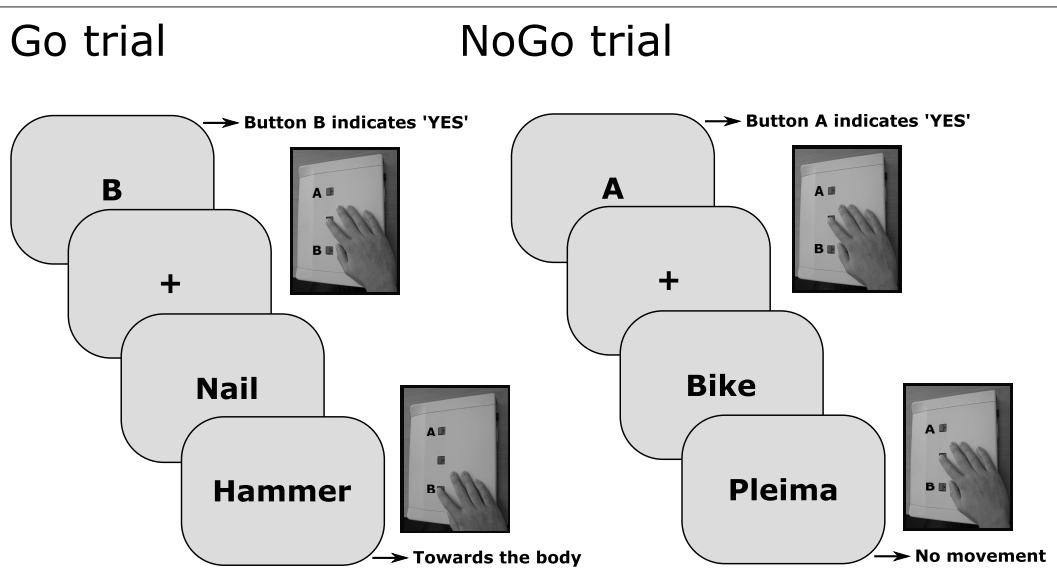


FIGURE 1 | Illustration of a Go and NoGo trial.

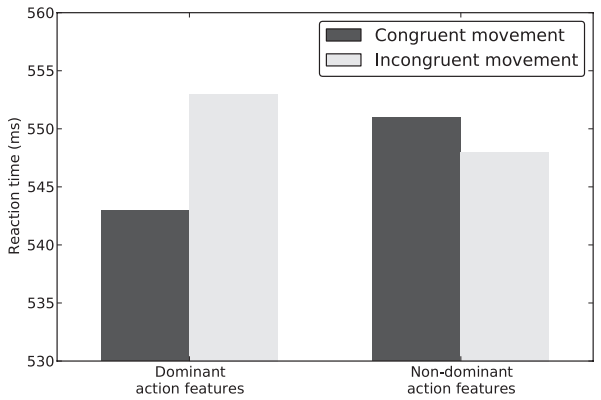


FIGURE 2 | Average reaction times (RTs) for words, as a function of the congruency between the cue and associated movement direction, and the contextual focus (focus on dominant action feature vs. non-dominant action feature).

Table 2 | Mean ratings of the pre-tests.

	Body words	World words
Length	6.8	6.3
Lemma frequency per million (CELEX)	567	487
Imageability	6.82	6.76
Action association	–1.33	2.33

Table 3 | Average performance rates (PR), reaction times (RTs), and movement times (MTs) with standard errors for congruent and incongruent trials in both the dominant focus and non-dominant focus condition.

	PR (SE)	RT (SE)	MT (SE)
DOMINANT FOCUS			
Congruent	98.4 (0.26)	543 (14.73)	276 (16.05)
Incongruent	98.3 (0.37)	553 (16.21)	272 (15.55)
NON-DOMINANT FOCUS			
Congruent	98.6 (0.35)	551 (15.54)	272 (15.51)
Incongruent	98.3 (0.37)	548 (15.84)	269 (15.69)

DISCUSSION

The present study clearly demonstrates contextual effects on embodied word processing, evidenced by a reliable Action Congruency × Context interaction effect. Specifically, a congruency effect (i.e., faster word recognition times) is observed for trials in which the direction of the participant's response and the motor program generally associated with the word's referent correspond, but only if the word is presented in a language context highlighting the functional use of the object (e.g., *thirst* – ***cup***). In a language context highlighting a less typical use of the object (e.g., *sink* – ***cup***), the congruency effect disappears. These findings suggest that activation of modality specific (in this case motor) information during word comprehension is a flexible and contextually dependent process, and not a process of an automatic and invariant simulation of a specific motor program. In other words, semantic features are flexibly recruited with respect to the type of properties emphasized by the semantic context (see also Hoenig et al., 2008; Masson et al., 2008; Raposo et al., 2009).

The congruency effect in the present study extends the results of our previous research on motor involvement in language processing (Rueschemeyer et al., 2010a) and demonstrates the important role of contextual information in the embodied processing of language. We show that preparing an action congruent to the typical, functional use of an object facilitates processing of the word denoting the object. We suggest that words presented in isolation rely on a default representation, which highlights the typical functional use of the object. In our current results we extend this finding to show (not surprisingly) that words presented in a context highlighting the denoted object's functional use are also facilitated by preparing a congruent action.

Interestingly, the action congruency did not affect the word processing time if the semantic context emphasized action features not belonging to the core of the concept. There are two alternative explanations for this finding: first, it is possible that motor semantic features are not activated at all in a non-dominant action context. Alternatively, it is possible that dominant conceptual features are co-activated along with non-dominant features. In other words, *cup* in the context of *sink* activates both motor programs associated with moving a cup toward the sink and motor programs associated with bringing the cup toward the mouth. This activation of two opposite motor programs might have resulted in a null effect for trials in which properties unrelated to the functional use of the object are emphasized. In this scenario, the priming effect of movement preparation on lexical access is canceled out, because motor codes underlying two opposing movements are activated simultaneously. This explanation is in line with the findings of Hoenig et al. (2008) in which they show that activation in regions coding non-dominant object features increased if semantic context encouraged participants to focus on the non-dominant feature. However, brain regions coding information about dominant features were co-activated, even when not directly probed by the task. The authors argue that dominant object features are co-activated with non-dominant object features through collateral support (see also Farah and McClelland, 1991).

The reliable congruency effect for trials in which there was a contextual focus on action properties related to the functional use of the object, bolsters the claim that in processing of a word with an action-semantic component we activate information stored in modality specific sensory-motor systems (Hoenig et al., 2008). The present study, however, cannot determine whether motor system involvement is a fundamental necessity or a consequence of word comprehension. One might argue that the reason for observing a congruency effect between prepared movement direction and associated movement direction is because the person voluntarily images the functional use of the referent object, after the meaning of the object word is already understood. The design of our current study prevents us from ruling out this possibility. For a better understanding of the exact nature of the interaction between context

and embodied word processing, we need to further investigate the temporal dynamics of the observed contextual effects on embodied word processing.

CONCLUSION

The current study demonstrates that the interaction between lexical-semantic processing and movement preparation processes was modulated by the context in which these words were encountered. Together the data suggest that context plays a fundamental role in sensory-motor activations during language processing. That is, the activation of specific motor properties in language comprehension is flexible and context-dependent.

ACKNOWLEDGMENTS

The authors thank P. de Water, and G. van Oijen for technical support. The study was supported by NWO-Veni grant nr (016-094-053) to Shirley-Ann Rueschemeyer and NWO-VICI grant nr (453–05–001) to Harold Bekkering.

REFERENCES

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823.

Baayen, R., Piepenbrock, R., and Van Rijn, H. (1993). *The CELEX Lexical Database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91.

Boulenger, V., Hauk, O., and Pulvermueller, F. (2009). Grasping ideas with your motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Chao, L. L., and Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12, 478–484.

Desai, R. H., Binder, J. R., Conant, L. L., and Seidenberg, M. S. (2010). Activation of sensory-motor areas in sentence comprehension. *Cereb. Cortex* 20, 468–478.

de Zubicaray, G., Postle, N., McMahon, K., Meredith, M., and Ashton, R. (2010). Mirror neurons, the representation of word meaning, and the foot of the third left frontal convolution. *Brain Lang.* 112, 77–84.

Farah, M. J., and McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *J. Exp. Psychol. Gen.* 120, 339–357.

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Glenberg, A. M., and Kaschak, M. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glover, S., Rosenbaum, D. A., Graham, J., and Dixon, P. (2004). Grasping the meaning of words. *Exp. Brain Res.* 154, 103–108.

Hauk, O., Johnsrude, I., and Pulvermueller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Hauk, O., and Pulvermueller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Hum. Brain Mapp.* 21, 191–201.

Hoenig, K., Sim, E.-J., Bochev, V., Herrnberger, B., and Kiefer, M. (2008). Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *J. Cogn. Neurosci.* 20, 1799–1814.

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., and Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI. *Brain Lang.* 107, 16–43.

Kintsch, W. (2008). "Symbol systems and perceptual representations," in *Symbols and Embodiment*, eds M. De Vega, A. Glenberg, and A. Graesser (Oxford: Oxford University Press), 145–164.

Masson, M., Bub, D., and Warren, C. (2008). Kicking calculators: contribution of embodied representations to sentence comprehension. *J. Mem. Lang.* 59, 256–265.

Pulvermueller, F. (1999). Words in the brain's language. *Behav. Brain Sci.* 22, 253–336.

Pulvermueller, F. (2001). Brain reflections of words and their meaning. *Trends Cogn. Sci.* 5, 517–524.

Pulvermueller, F., Härle, M., and Hummel, F. (2000). Neurophysiological distinction of verb categories. *Neuroreport* 11, 2789–2793.

Pulvermueller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892.

Raposo, A., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia* 47, 388–396.

Rueschemeyer, S.-A., Brass, M., and Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor stems. *J. Cogn. Neurosci.* 19, 855–865.

Rueschemeyer, S.-A., Pfeiffer, C., and Bekkering, H. (2010a). Body schematics: on the role of the body schema in embodied lexical-semantic representations. *Neuropsychologia* 48, 774–781.

Rueschemeyer, S.-A., van Rooij, D., Lindemann, O., Willems, R., and Bekkering, H. (2010b). The function of words: distinct neural correlates for words denoting differently manipulable objects. *J. Cogn. Neurosci.* 22, 1844–1851.

Saccuman, M. C., Cappa, S. F., Bates, E. A., Arevalo, A., Della Rosa, P., Danna, M., and Perani, D. (2006). The impact of semantic reference on word class: an fMRI study of action and object naming. *Neuroimage* 32, 1865–1878.

Shtyrov, Y., Hauk, O., and Pulvermueller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *Eur. J. Neurosci.* 19, 1083–1092.

Taylor, L. J., Lev-Ari, S., and Zwaan, R. A. (2008). Inferences about action engage action systems. *Brain Lang.* 107, 62–67.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–281.

Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *J. Exp. Psychol. Gen.* 135, 1–11.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 June 2010; paper pending published: 03 July 2010; accepted: 21 August 2010; published online: 04 October 2010.
Citation: van Dam WO, Rueschemeyer S-A, Lindemann O and Bekkering H (2010) Context effects in embodied lexical-semantic processing. Front. Psychology 1:150. doi: 10.3389/fpsyg.2010.00150
This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.
Copyright © 2010 van Dam, Rueschemeyer, Lindemann and Bekkering. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 27 January 2011
doi: 10.3389/fpsyg.2010.00243

Deriving motor primitives through action segmentation

Paul E. Hemeren* and Serge Thill

School of Humanities and Informatics, University of Skövde, Skövde, Sweden

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Lewis A. Wheaton, Georgia Tech, USA
Yann Coello, University of Lille Nord de France, France

****Correspondence:***
Paul E. Hemeren, School of Humanities and Informatics, University of Skövde, Box 408, 541 28 Skövde, Sweden.
e-mail: paul.hemeren@his.se

The purpose of the present experiment is to further understand the effect of levels of processing (top-down vs. bottom-up) on the perception of movement kinematics and primitives for grasping actions in order to gain insight into possible primitives used by the mirror system. In the present study, we investigated the potential of identifying such primitives using an action segmentation task. Specifically, we investigated whether or not segmentation was driven primarily by the kinematics of the action, as opposed to high-level top-down information about the action and the object used in the action. Participants in the experiment were shown 12 point-light movies of object-centered hand/arm actions that were either presented in their canonical orientation together with the object in question (top-down condition) or upside down (inverted) without information about the object (bottom-up condition). The results show that (1) despite impaired high-level action recognition for the inverted actions participants were able to reliably segment the actions according to lower-level kinematic variables, (2) segmentation behavior in both groups was significantly related to the kinematic variables of change in direction, velocity, and acceleration of the wrist (thumb and finger tips) for most of the included actions. This indicates that top-down activation of an action representation leads to similar segmentation behavior for hand/arm actions compared to bottom-up, or local, visual processing when performing a fairly unconstrained segmentation task. Motor primitives as parts of more complex actions may therefore be reliably derived through visual segmentation based on movement kinematics.

Keywords: action representation, event segmentation, motor primitives, mirror neurons, point-light displays, biological motion, action recognition, motor cognition

INTRODUCTION

Human actions have a spatiotemporal structure that can be accessed when we execute our own actions and when we view the actions of others. Understanding the structure of human actions allows us to identify and predict, or to see, the intentions of others (e.g., Blakemore and Decety, 2001; Saxe et al., 2004; Iacoboni et al., 2005). This ability to see a pattern of human motion as a coherent whole and not just as a complex pattern of movement of the arms and legs is referred to as epistemic visual perception (Jeannerod and Jacob, 2005). The meaning of actions is tied to the conceptual knowledge associated with a given pattern of bodily motion. Conceptual knowledge in turn includes knowledge about the goals of the movement as well as its sensory–motor patterns in motor execution and the visual recognition of actions. This linkage between perceptual and conceptual knowledge is apparent in tasks that demonstrate interference effects between perceptual and conceptual tasks (van Dantzig et al., 2008). The research presented here further addresses the relationship between perceptual and conceptual knowledge in the context of event segmentation (Zacks and Swallow, 2007).

ACTION REPRESENTATION, MOTOR PRIMITIVES, AND EVENT STRUCTURE

Motor execution and the visual recognition of actions appear to have a common neurological basis in mirror neurons in the primate brain, which become activated when an individual performs certain actions *and* when an individual observes the actions of another person performing the actions (e.g., di Pellegrino et al., 1992; Rizzolatti et al., 1996, 2006; Buccino et al., 2004). Mirror neurons have been described as a critical neurophysiological basis for internal models for action representation and for mediating the coupling between perception and action (Pozzo et al., 2006). These internal models have in other work been referred to as action prototypes (e.g., Pollick et al., 2001; Giese and Lappe, 2002; Pollick, 2004; Hemeren, 2008), motor schemata (Grafton et al., 1997), and motor prototypes (Borghi and Riggio, 2009). The ability to imagine, or simulate, movement relies on activation of cortical areas (e.g., primary motor cortex, the supplementary motor area, and the premotor cortex) that are involved in the execution of actions (Jeannerod, 1995; Michelon et al., 2006). One central idea regarding the structure and organization of the internal models concerns the existence of motor primitives that can be flexibly combined to create complex action sequences (e.g., Mussa-Ivaldi and Bizzi, 2000; Thoroughman and Shadmehr, 2000; Poggio and Bizzi, 2004; Chersi et al., 2006). The combination of motor primitives can also create hierarchical representations that allow generalization over specific situations (Poggio and Bizzi, 2004). Conversely, action hierarchies can also be understood on the basis of their component structure.

In order to understand our ability to represent actions, we need to understand what the motor primitives are and how they are combined. When dealing with motor execution, the basic question concerns the forces that are needed to produce the appropriate limb movement. Factors such as limb position, velocity, and acceleration determine the required forces (Mussa-Ivaldi and Bizzi, 2000; Poggio and Bizzi, 2004). In contrast, Chersi et al. (2006, 2010) referred to *reaching*, *grasping*, *transporting*, and *placing* as motor primitives in their proposed Chain Model for representing the separate

goal-based representations of (eating) and (placing). According to this model, pools of neurons in the mirror system, each encoding a motion primitive (e.g., *reaching for a peanut, grasping the peanut, bringing the peanut toward the mouth*) are linked together to form an overall action (e.g., *eating a peanut*). Regardless of whether motor primitives are described in terms of the kinematic variables in motor execution (limb position, velocity, etc.) or in terms of more higher-lever descriptions of motor execution (reaching for a peanut) further empirical work is needed to verify the relationship between primitives involved in motor execution and the visual recognition of actions.

For the visual recognition of actions, the visual system in most cases requires access to limb position, velocity, and acceleration. Given the tight coupling between perception and action, we would expect motor primitives for action execution and recognition to be very similar. A complementary method for determining motor primitives would be to engage participants in an action segmentation task and then assess the degree of agreement between the action kinematics and the segmentation behavior of the participants. Recent advances in understanding how humans represent events have been made by asking people to segment events (Newtson and Engquist, 1976; Newtson et al., 1977; Zacks, 2004; Zacks and Swallow, 2007; Zacks et al., 2009; Meyer et al., 2010). Zacks (2004) for instance presented participants with short animations of a circle and a square. They were then asked to segment the animations into both fine- and coarse-grained meaningful units. The results showed that movement features (e.g., position, velocity, acceleration, etc.) of the stimuli could reliably predict the event segmentation, especially for the fine-grained segmentation. In another set of experiments, Zacks et al. (2009) used more naturalistic action events. For these experiments, movies of a person folding laundry, building a house out of blocks, and assembling a video game system were segmented by participants. Movement variables associated with each event were recorded by a motion tracking system. Again, the participants were asked to segment the movies into fine-grained and coarse-grained units. The results confirmed previous findings showing that movement features were significantly correlated with segmentation behavior. For example, the speed and acceleration of body parts indicated breakpoints between action segments. This shows that the visual parsing of fairly complex events is tied to the kinematics of the stimuli.

Zacks et al. (2009) included an additional manipulation that investigated the influence of top-down conceptual knowledge vs. bottom-up driven processing on the segmentation behavior of the participants. This was done by converting the previous stimuli to only show the actor's head and hands, and their relation to one another. The idea here was that if participants have information about the meaning of the action, then they will be more inclined to segment the stimuli on the basis of that conceptual knowledge and less on the finer kinematic features, which should lead to a coarser segmentation. One group of participants (top-down) was told that they would see the same previously presented movies but this time as animations of the actor's head and hands. Another group was only told that they would see an animation of an actor performing a daily activity and that the motion of his hands and head are represented by objects used to record their movement. In this case, no contextual information about physical objects or surroundings was available to the participants. The results showed that this manipulation had no measurable effect on segmentation behavior. Regardless of previous knowledge, participants continued to base their segmentation on the perceptually salient movement features. When other sources of information were removed by just showing the simple animations, relations between movement and segmentation were strengthened. This result stands in contrast to results obtained by Castellini et al. (2007) for their model of motor execution for reaching and grasping. Prior knowledge of knowing the object (a can, a roll of duct tape, and a mug) involved in the action led to better performance for the model. Knowing the object involved in reaching and grasping allows for a more accurate hand shape in the early phase of the action.

The purpose of the present experiment is to further understand the effect of levels of processing (top-down vs. bottom-up) on the perception of movement kinematics and primitives for grasping actions in order to gain insight into possible primitives used by the mirror system as hypothesized for example by Chersi et al. (2006, 2010). Specifically, the present experiment investigates whether or not the definition of a motor primitive is affected by conceptual knowledge of the observed action. When activated knowledge about the object used in the action activates information about the possible goal of the action, this may lead to generalizations over different motor routes to the same goal. In this sense, action segmentation when guided by top-down processing may lead to fewer segments and perhaps greater agreement among people about where the relevant breakpoints between segments should be placed.

The experimental paradigm from previous studies on event segmentation (Zacks et al., 2009) will be used to investigate the role of top-down vs. bottom-up processing within the context of action segmentation for 12 grasping actions. If people are given prior information about the grasping actions being performed, this should be sufficient to activate conceptual knowledge by which to guide segmentation of the grasping actions. On the other hand, if people have no idea about the goals of the actions, then there may be a tendency to focus on more fine-grained segments of the actions because there is no information from which to abstract from small kinematic changes in the movements. Accordingly, there should be more segments and perhaps more variability about where the marks for the breakpoints should be placed for people who are only allowed to used bottom-up processing.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-four participants were recruited from the student population at the University of Skövde, Sweden. Participants were randomly assigned to one of two viewing conditions. For the picture condition (top-down processing condition; $n = 12$, six females, mean age = 28) participants were first given a picture of the object that was used in each action. In the inverted + no-picture condition (bottom-up processing condition; seven females, mean age = 23) each action was mirror inverted and no-pictures of the objects were shown to the participants. Participants were also paid approximately $6.50 for their participation. All participants were informed as to the nature of the experiment including possible risks and benefits. On the basis of this information, every participant

provided written consent to their participation in the experiment. This was in accordance with Swedish law and the World Medical Association Declaration of Helsinki.

STIMULI/MATERIALS

Twelve grasping actions using one arm/hand (**Table 1**) were recorded using a ShapeHandPlus™ motion capture device (**Figure 1**). The ShapeHand™ data glove was integrated with an arm tracking ShapeTape™. This device allowed the precise capture of finger, hand, and arm positions of the person performing the actions. It also allowed the recording of actions without the objects being visible. The primary purpose of the glove was to reliably capture the kinematics of fine-grained arm and hand movements, which was not possible with other motion capture equipment such as point-light motion capture systems. For the purposes of the present experiment, the 3D coordinates of 22 reference points on the limb (the tip of each finger and thumb as well as the joints of the hand and arm) were recorded.

All 12 actions were performed by the same right-handed person using the right arm/hand to perform each action. Each action started and ended at the same resting position with the arm at the side and was performed as naturally as possible with the actual relevant objects being used in each action. Natural arm and hand actions of different durations and complexity were selected in order to sample from a wide variety of such actions and to be able to text for the sensitivity of potential differences in the number of segments for different actions. All actions, however, included moving the arm toward an object, grasping the object, using the object, releasing the object, and bringing the hand/arm down to rest again. For example, we included cyclical (or iterative) actions such as sawing, writing, turning pages in a book, and non-cyclical actions such as opening a door, drinking from a mug, opening a can, and pouring from a bottle. We did not, however, create an *a priori* segmentation of these actions since a comparison between a predetermined number of segments and participant segmentation behavior was not the focus of the experiment. However, it was important to be able to demonstrate that the experimental method was sufficiently sensitive to detect differences in the number of segmentation marks between actions. Therefore, we included actions that seemingly had different numbers of segments. For example, *writing on a whiteboard* could be considered as consisting of the following segments: picking up a pen, taking the cap off, starting to write, underlining the text, putting the cap back on, and then putting the pen down on the table. *Opening a door* could be considered as consisting of: gripping the door handle, turning it, pushing it forward, and then releasing it. Whether or not participants would notice and mark the segments was one of the issues under investigation.

Table 1 | List of action sequences used in the experiment (Sequence duration rounded to the nearest second).

1. Cut with scissors	(22)	7. Drink from a mug	(16)
2. Lift a dumbbell	(15)	8. Open a can and take a drink	(19)
3. Open a door	(8)	9. Solve the tower of Hanoi	(21)
4. Pour from a bottle	(12)	10. Turn pages in a book	(20)
5. Saw wood	(15)	11. Unscrew a bottle cap	(14)
6. Spray from a spray bottle	(12)	12. Write on whiteboard	(22)



FIGURE 1 | ShapeHand™ motion capture system.

Each action was shown as a constellation of point-lights (white lights against a black background) corresponding to the reference points mentioned above. The motivation for using the point-light technique was to minimize the possible confounding of form information about the hand and arm with the kinematics of the actions. By using point-lights, participants should still be able to recognize the actions and yet not be influenced by the form information about the configuration of the hand and arm. This is certainly the case for full-body actions (e.g., Dittrich, 1993; Blake and Shiffrar, 2007) and previous results from Poinzner et al. (1981) show that point-lights attached to the fingers can reliably convey American Sign Language. There is thus reason to believe that observers should be able to recognize actions shown as point-light displays even for hand and arm actions. In addition, participants in the picture condition would be shown the objects involved in the different actions, which should facilitate recognition of the actions in a top-down manner.

Previous results using point-light displays have shown that when they are turned upside down (inverted), people have a greater difficulty recognizing the actions being performed (e.g., Dittrich, 1993; Pavlova and Sokolov, 2000; Shiffrar and Pinto, 2002). It appears that the global processing of inverted point-light displays is impaired (Hemeren, 2005), which makes action recognition difficult. Many people can see the movement of arms, legs, and hands (intact local motion processing) but have difficulty describing the movement at a higher global level of meaning (impaired global motion processing; e.g., Sumi, 1984). These results for point-light displays of biological motion are very similar to the visual processing limitations of viewing an inverted face (e.g., Carey and Diamond, 1994; Boutsen and Humphreys, 2003; Leder and Carbon, 2006). Therefore, in order to create a condition where top-down (conceptual) processing of kinematic information is severely impaired, we inverted the 12 point-light actions. It should be emphasized that despite inverting the actions the *same* kinematic information is available to the participants in both conditions. Inverting the actions does not change the kinematics present in the displays. Participants in this condition were not shown any pictures of the objects involved in the actions. All recorded action sequences were displayed in real time at 30 frames per second. Each action was also oriented to avoid as

much occlusion as possible. See **Figures 2 and 3** for an example of five frames from the *drink from a mug* sequence and the inverted version of the same action.

PROCEDURE

Participants were tested individually and were seated at a distance of about 60 cm to a laptop computer screen. Prior to segmenting the actions, participants were informed as to the nature of the experiment. They were told that they would see 12 brief arm-and-hand movements and that the movements would be presented as a constellation of moving point-lights. Each action was performed by one arm and hand. All participants were first instructed to simply watch an action all the way through before starting the segmentation procedure.

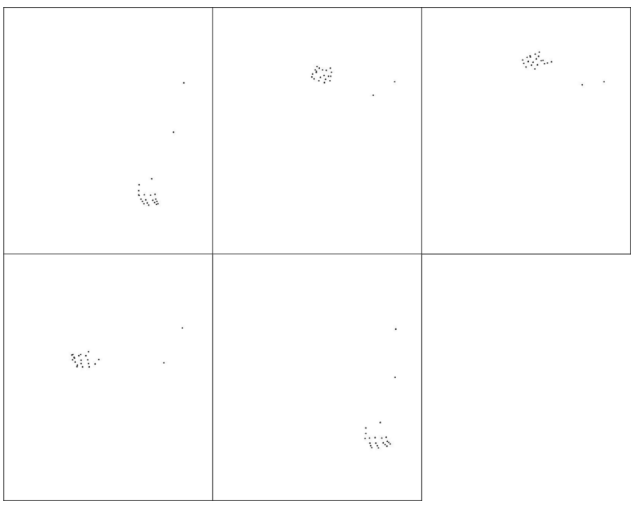


FIGURE 2 | Five frames (black-on-white for clarity) at 4 s intervals from the *drink from mug* sequence, right side up.

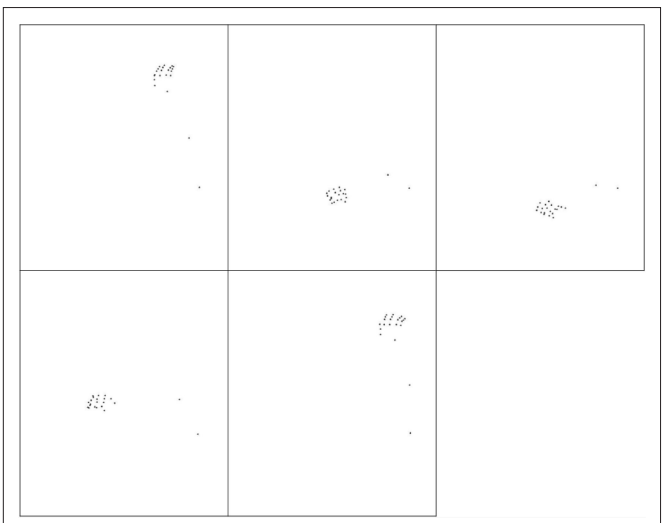


FIGURE 3 | Five frames at 4 s intervals from the *drink from mug* sequence, inverted.

Participants in the picture condition were first presented with a picture of the object that was involved in the action, and then they viewed the action all the way through once. After viewing the action, they were asked to describe what action the person was performing and then to begin segmenting the action. Participants in the no-picture + inverted condition viewed each action without previously viewing a picture of the object used in the action. After viewing an action all the way through, the participants in this condition were instructed to begin segmenting the action without having described the action the person was performing. Following the segmentation phase, participants were asked to describe the action they thought the person was performing.

For the segmentation phase, participants were told to mark breakpoints in the action sequences that constituted the transitions between different segments in the action sequences, if they thought there were any such segments. It was left up to the participants to determine whether or not the actions contained different segments and where to mark the possible breakpoints between action segments.

Participants were also instructed how to use the stimulus player (**Figure 4**) so that they could be fairly accurate at placing marks at the breakpoints between segments. In order to reduce potential variability in participant segmentation, the stimulus player included a pause function as well as rewind and fast-forward possibilities by using the yellow slider. In other words, participants were given the opportunity to manipulate the action in order to find the desired breakpoints between segments for each action. Once they found a breakpoint between two segments, they were instructed to mark this by pressing the space bar, in which case a red line would appear on a time line below the animated figure. They were also able to make self-imposed corrections if they determined that to be necessary.

The software displaying the actions recorded the number of marks made for each action and participant as well as which frames in the action sequence were marked. There were no set time limits. The experiment lasted approximately 45 min. Participants were

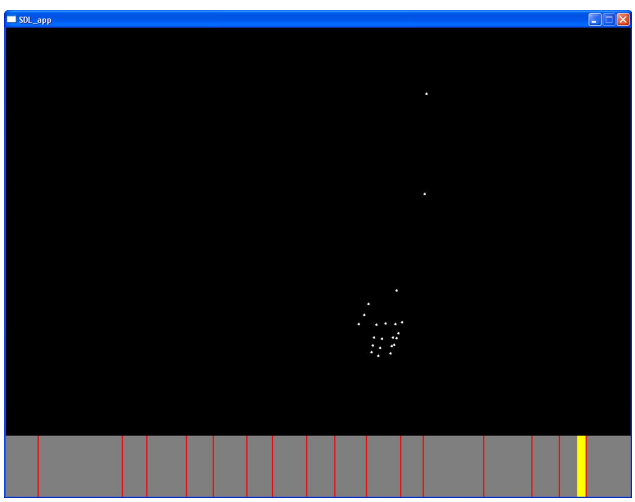


FIGURE 4 | Screen shot of an action (turning pages in a book) displayed in the stimulus player. The red lines are segmentation marks and the yellow bar is the slider.

given two practice actions to familiarize themselves with the stimuli and the tasks. The order of presentation for the 12 actions was randomly determined for each participant.

Almost all participants raised the question of how to define a segment. They were told that a segment could be viewed as a part of an action that can be used as a kind of building block to construct the whole action and that they should just mark the segments they judged to be necessary for that action. Participants had no trouble understanding the notion that actions can be segmented into action parts. The experimenter wrote down the verbal descriptions of the stimuli for later analysis. Firstly, all participant descriptions were judged for their correctness. A description was scored as correct if it included a correct identification of the action, not necessarily the object. For example, a description that included "pouring" for the *pour from bottle* action would be scored as correct. Participants' descriptions were generally clear about whether or not they recognized the actions.

In addition to participants' verbal descriptions and number of segmentation marks, the placements of the marks were analyzed in relation to the kinematic variables of velocity, acceleration, and change in direction. Velocity (measured in m/s) was obtained by approximating the first derivative of the position of recorded markers over time. It thus represents the mean tangential velocity. Acceleration (measured in m/s^2) was equivalently obtained by approximating the second derivative. The change in direction (measured in degrees/s) was computed as the angular difference between tangential vectors to the motion path at two consecutive frames in time (see **Figure 5**).

These variables were calculated for movements of the wrist as well as every finger tip during the execution of the action. To eliminate noise and minor sources of variability in the kinematic data, the obtained signals were smoothed using a standard low-pass filter (see **Figure 6**).

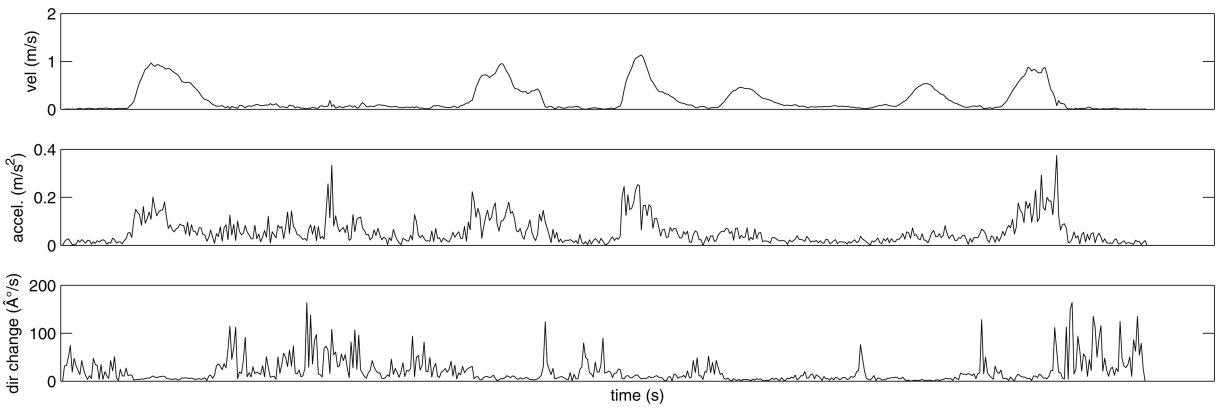


FIGURE 5 | Kinematic profile according to the time course for velocity (top), acceleration (middle), and changes in direction (bottom) for the wrist during the action *open can and drink*.

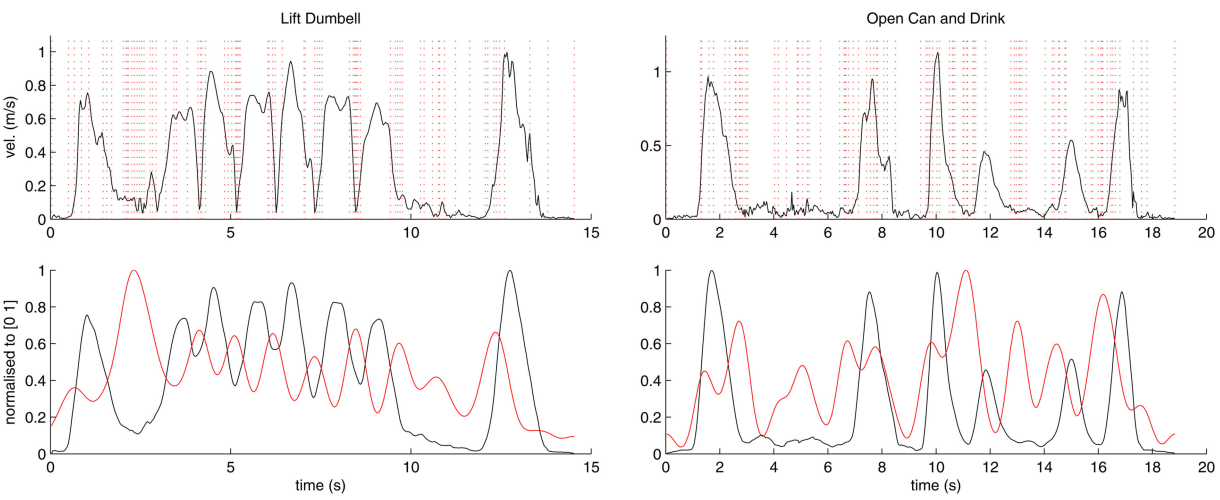


FIGURE 6 | Computation of the correlation coefficients between participant responses and kinematic data. The top row shows raw velocity data and marks placed by participants from group 1 (pictures) for two actions; *lift dumbbell* (left) and *open can and drink* (right). The bottom row shows the corresponding velocity after smoothing through a low-pass filter (black) as well as the density function computed from the marks placed by participants (red). Curves in the bottom row are normalized to the range of [0 1] for better visualization.

RESULTS

VERBAL DESCRIPTIONS

The results for the number of correct verbal descriptions as a function of picture condition and action are presented in **Table 2**. The results show clearly that participants who viewed that actions upright and together with a picture could identify the actions presented in the point-light displays. These participants could identify the immediate goal of the action. Here are some examples of the verbal responses:

- *The person is gripping the door handle and opening the door.*
- *The person is unscrewing the cap on the bottle.*
- *The person is pouring something from the bottle.*
- *A person opened a can and then lifted the can and took a drink.*

In contrast, the participants who saw the actions upside down and did not get to view the object used in the action were impaired at recognizing the goal, or higher-level purpose, of the actions. Here are some examples of the verbal responses for participants in that condition (with the displayed action in parentheses):

- *Someone sticking their thumb out and pointing in a certain direction. (opening a door)*
- *Moves the hand down, and the thumb switches place, the thumb and index finger, and the middle finger are touching each other. (unscrewing a bottle cap)*
- *Grabs something without using the index finger. (pouring from a bottle)*
- *First phase, the hand and fingertips do something, the second phase, grabs something and holds it close and then pushes it away. (open a can and drink)*

These descriptions are typical of the participant responses in this condition and show that participants are able to see and describe the motion of the fingers and arm but fail to interpret the motions according to any higher-level action description. It appears then that participants have visual access to the kinematics of the parts of the hand and arm.

NUMBER OF SEGMENTATION MARKS

The mean number of marks as a function of viewing condition and action are presented in **Table 3**. Regarding the differences between the number of marks for viewing condition (picture vs. no-picture-inverted), participants made slightly more marks in the picture condition than in the no-picture-inverted condition. A mixed ANOVA, 2 (viewing condition: picture vs. no-picture-inverted, between subjects) × 12 (action, within subjects) on the marking data, however, showed that the main effect of viewing condition was not significant, $F < 1$. The main effect of action, however, was significant, Greenhouse–Geisser adjusted $F(4.2, 92.32) = 6.51$, $p < 0.001$. This indicates that participants viewed different actions as consisting of different numbers of segments. The interaction was not significant, Greenhouse–Geisser adjusted $F(4.2, 92.32) = 1.53$, $p = 0.20$.

Regarding the main effect of the action variable, some actions received more segmentation marks than others. For example, *post hoc* comparisons showed that participants marked significantly more segments for *writing on the board* ($M = 7.7$) compared to

Table 2 | Number of correct verbal descriptions of the actions.

	Action					
Viewing condition	**Cut with scissors**	**Lift dumbbell**	**Open door**	**Pour from bottle**	**Saw wood**	**Spray with spray bottle**
Pictures	12	12	12	12	12	12
No-pictures inverted	4	1	1	2	2	1
	Drink from mug	**Open a can and drink**	**Move disks tower Hanoi**	**Turn pages in book**	**Unscrew bottle cap**	**Write on board**
Pictures	12	12	12	12	11	12
No-pictures inverted	1	1	1	0	1	3

Table 3 | Mean number of segmentation marks as a function of viewing condition and action (SDs in parentheses).

	Action					
Viewing condition	**Cut with scissors**	**Lift dumbbell**	**Open door**	**Pour from bottle**	**Saw wood**	**Spray with spray bottle**
Pictures	7.7 (5.6)	7.8 (3.7)	4.7 (2.1)	5.0 (2.0)	9.7 (6.9)	6.7 (3.0)
No-pictures inverted	6.5 (4.7)	6.5 (4.3)	4.3 (1.9)	5.5 (2.4)	6.0 (3.3)	6.5 (3.6)
	Drink from mug	**Open a can and drink**	**Move disks tower Hanoi**	**Turn pages in book**	**Unscrew bottle cap**	**Write on board**
Pictures	5.6 (1.8)	7.8 (2.7)	12.0 (7.9)	9.2 (6.0)	6.0 (3.8)	8.3 (5.3)
No-pictures inverted	5.3 (2.1)	6.9 (2.5)	7.6 (6.8)	7.8 (4.5)	5.8 (3.3)	7.0 (2.9)

opening a door ($M = 4.5$), $t(22) = 4.66$, $p = 0.008$. Other Bonferroni adjusted *post hoc* comparisons showed that *lifting a dumbbell* ($M = 7.1$), *opening a can and drinking* ($M = 7.4$), *move disks on the Tower of Hanoi* ($M = 9.8$), and *turning pages in a book* ($M = 8.5$) had significantly more segment marks than *opening a door*, ps < 0.05. *Opening a can and drinking* also had significantly more segmentation marks than *pouring from a bottle* ($M = 5.3$) and *drinking from a mug* ($M = 5.4$), $p < 0.05$. These differences will be further discussed in the Section "Discussion."

The large standard deviations for the segmentation marks for some of the conditions suggests that participants the number of segmentation marks varied quite a bit for some of the actions. This seems to be a result of a difference in marking behavior for the repetitive actions like *cutting with scissors*, *sawing wood*, *moving disks tower of Hanoi*, *turning pages in a book*, and *writing on a board*. This will also be discussed in the Section "Discussion."

CORRELATIONS BETWEEN KINEMATIC VARIABLES AND SEGMENTATION

For every action, a density function of the marks placed along the timeline by each group of participants was computed using a Gaussian kernel with a width of 0.3 s (see **Figure 6**). Linear (Pearson) correlation coefficients between the density function and each of the kinematic variables (velocity, acceleration, and change in direction) were then calculated (**Table 4**). Although kinematic variables were tracked for the wrist as well as for every finger, we did not in general find any noteworthy differences in the correlation coefficients between the different points on the hand and wrist and the density functions. For simplicity, we therefore limit the discussion in this paper to the behavior of the wrist kinematics. Varying the width of the kernel within reasonable limits (i.e., avoiding extreme cases in which the peaks of the density function are exceedingly narrow or in which the density function is so smeared out that details of the marking behavior are lost) did not generally affect whether or not correlations were significant.

The results in **Table 4** show that change in direction is significantly inversely correlated with 10 of the 12 actions in the picture condition and with 11 of the 12 actions in the no-picture-inverted condition, indicating that change in direction is associated with fewer segmentation marks. Segmentation marking is more positively associated with other kinematic variables. Velocity, for example, is significantly positively correlated with seven actions in the

Table 4 | Correlations (Pearson r) between mark density function and kinematic variables measured from the wrist: change in direction of movement, velocity, and acceleration.

			Viewing condition: with pictures			
			Action			
Wrist	**Cut with scissors**	**Lift dumbbell**	**Open door**	**Pour from bottle**	**Saw wood**	**Spray with spray bottle**
Change direction	**–0.18***	**–0.29***	**–0.34***	**–0.35***	**–0.48***	0.04
Velocity	**0.12***	0.03	0.02	–0.01	**0.43***	–0.02
Acceleration	**0.12***	**0.29***	**0.31***	0.11	**0.42***	0.09
	Drink from mug	**Open a can and drink**	**Move disks tower Hanoi**	**Turn pages in book**	**Unscrew bottle cap**	**Write on board**
Change direction	**–0.22***	**–0.28***	**–0.26***	**–0.12***	–0.02	**–0.45***
Velocity	**0.33***	**0.18***	0.01	**0.19***	**0.37***	**0.44***
Acceleration	**0.23***	**0.20***	0.06	**0.35***	**0.41***	**0.34***
			Viewing condition: no-picture-inverted			
			Action			
	Cut with scissors	**Lift dumbbell**	**Open door**	**Pour from bottle**	**Saw wood**	**Spray with spray bottle**
Change direction	**–0.49*!**	–0.11	**–0.33***	**–0.47***	**–0.52***	**–0.40*!**
Velocity	**0.46*!**	**–0.14***	**0.45*!**	**0.38*!**	**–0.13***	**0.51*!**
Acceleration	**0.32*!**	**0.20***	**0.75*!**	**–0.41*!**	**–0.40***	**0.28*!**
	Drink from mug	**Open a can and drink**	**Move disks tower Hanoi**	**Turn pages in book**	**Unscrew bottle cap**	**Write on board**
Change direction	**–0.22***	**–0.32***	**–0.36***	**–0.33*!**	**–0.15***	**–0.54***
Velocity	**0.53*!**	**0.36*!**	**0.48*!**	**0.46*!**	**0.54*!**	**0.51***
Acceleration	**0.50*!**	**0.29***	**0.44*!**	**0.40***	**0.53***	**0.30***

A (*) indicates a significant coefficient at the 0.01 level. Statistically significant coefficients are in bold-face type. A (!) indicates a significant difference between that coefficient and the coefficient in the corresponding condition for the viewing condition with pictures at the 0.01 level.

picture condition and with 10 actions in the no-picture-inverted condition, which indicates that segmentation marking increases with velocity. Acceleration is significantly positively correlated with nine actions in the picture condition and with 10 actions in the no-picture-inverted condition.

Higher values of velocity appear to signal the start of a segment whereas an episode of ongoing changes in direction tends to be associated with the carrying out of a part of an action. For example, for the action of *drinking from a mug*, there are changes in direction during the drinking phase which consists of tilting the mug and consequently thereby changing the direction of the wrist point. The act of drinking as such appears to be a whole segment and is not further segmented. An analysis of this inverse relationship between change in direction and velocity (**Table 5**) shows that it is significant for all actions. This effect indicates that it is not the occurrence of a change in direction for the wrist that is associated with action segmentation.

The size of many of the correlation coefficients for the no-picture-inverted group are slightly higher than the corresponding coefficients in the picture group, which indicates that the segmentation behavior of the participants in the no-picture-inverted group is more strongly related to the kinematics for those correlations.

Table 5 | Correlation coefficients (Pearson *r*) for the relationship between change in direction and velocity for the actions.

Action	*r*	Action	*r*
1. Cut with scissors	–0.23	7. Drink from a mug	–0.45
2. Lift a dumbbell	–0.51	8. Open a can and take a drink	–0.23
3. Open a door	–0.45	9. Solve the tower of Hanoi	–0.21
4. Pour from a bottle	–0.30	10. Turn pages in a book	–0.34
5. Saw wood	–0.50	11. Unscrew a bottle cap	–0.30
6. Spray from a spray bottle	–0.32	12. Write on whiteboard	–0.39

All coefficients are significant at the 0.001 level.

Comparisons (z-transformed) between the correlation coefficients for the two viewing conditions, actions, and kinematic variables showed that 18 of the 36 coefficients for the no-picture-inverted group were significantly larger (assuming the same direction of correlation) than the corresponding coefficients for the picture group. This finding will be further in the Section "Discussion."

AGREEMENT ACROSS CONDITIONS

Despite some differences between the picture and the no-picture-inverted conditions, participants in both groups appear to be marking similar segments. Correlation coefficients for the relationship between the density functions for the picture and no-picture-inverted groups show significant agreement (**Table 6**). The range between the highest (unscrew a bottle cap) and lowest (tower of Hanoi) correlation coefficient is quite large and suggests different levels of agreement for the different actions. Plots of the density functions for the segmentation marks for the two groups and for two actions (*drink from mug* and *tower of Hanoi*) are presented in **Figure 7**.

Table 6 | Correlation coefficients (Pearson *r*) for the relationship between the density functions for the picture and no-picture-inverted groups.

Action	*r* (*n*)	Action	*r* (*n*)
1. Cut with scissors	0.72 (650)	7. Drink from a mug	0.86 (477)
2. Lift a dumbbell	0.71 (435)	8. Open a can and take a drink	0.50 (564)
3. Open a door	0.60 (236)	9. Solve the tower of Hanoi	0.18 (634)
4. Pour from a bottle	0.60 (343)	10. Turn pages in a book	0.58 (584)
5. Saw wood	0.46 (457)	11. Unscrew a bottle cap	0.93 (404)
6. Spray from a spray bottle	0.66 (349)	12. Write on whiteboard	0.67 (670)

All coefficients are significant at the 0.001 level (n = number of observations of velocity and mark density function over the time course of the action, i.e., number of frames for each action).

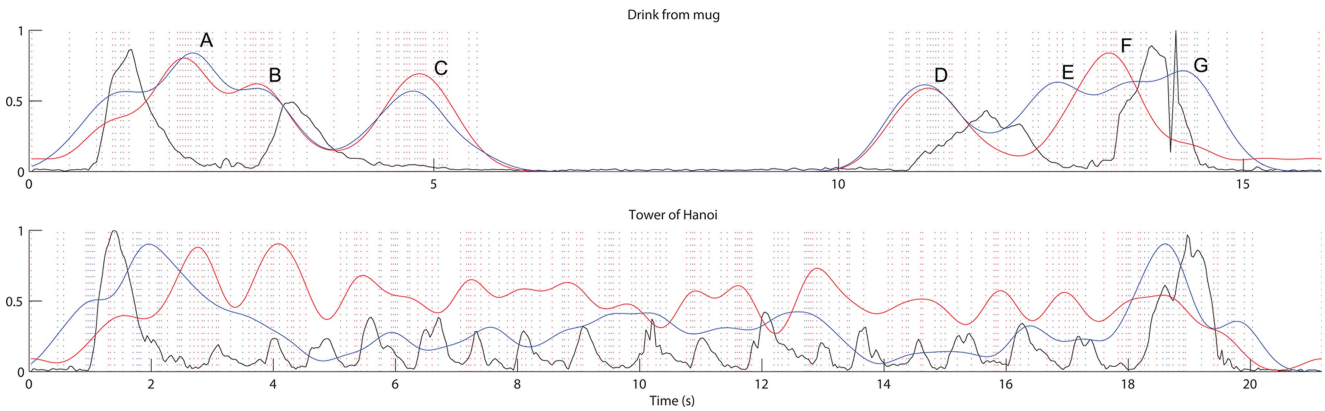


FIGURE 7 | Relationship between density functions for the segmentation marks for the picture and no-picture-inverted groups for two actions. Black = velocity. Red = picture group. Blue = no-picture group. Dotted vertical lines indicate marks placed by participants in the picture (red) and no-picture group (blue) respectively.

The top density function profile in **Figure 7** shows the considerable agreement between participants for the two different groups for the action *drinking from a mug*. For this action (top figure in **Figure 7**), participants marked gripping the mug (A), starting to lift the mug (B), starting to drink (C), stop drinking (D), putting the mug down, blue line (E), releasing the grip, red line (F), and setting the mug down again (G). In this case there was a small difference between the groups regarding the marking of release of the grip. The picture group appeared more inclined to mark that segment than the no-picture-inverted group.

For the bottom density function profile (tower of Hanoi), the main difference between the groups seems to concern whether or not to mark the recurring grasping–moving–releasing motions involved in the action. These differences will be further discussed in the Section "Discussion" below.

DISCUSSION AND CONCLUSION

In response to the central question of the experiment reported here, the results show that there is a significant relationship between segmentation behavior and velocity or acceleration and change of direction for most of the hand/arm actions. This result holds for the picture group and the no-picture-inverted group, which suggests that differences between top-down and bottom-up visual processing of the point-light hand/arm actions did not carry over to obvious differences in overall segmentation behavior. This conclusion is also supported by the significant correlations between the density profiles for the two different conditions (**Figure 7**), which indicate that participants in the two groups tend to place their segmentation marks in similar locations. The segmentation behavior of participants suggests that an increase in velocity (or acceleration for the picture group) is an important signal for denoting a segment boundary, or breakpoint.

There were also no significant differences between the number of segmentation marks for the picture and no-picture-inverted conditions (**Table 3**), which is consistent with the conclusion that the two groups of participants are similar in their segmentation behavior. Admittedly, drawing conclusions on the basis of a null-effect is somewhat problematic in that it can be difficult to determine whether the null-effect is the result of an insufficiently sensitive method or the result of there actually being no effect of the independent variable. With regard to the issue of an insufficiently sensitive method, there were significant differences between segmentation marks for the different actions, which suggests that the sensitivity of the method was sufficiently high to also detect potential differences between the different viewing conditions. There were, however, some rather high standard deviations for a few of the conditions. We purposely included a number of different actions in order to detect possible differences between different kinds of actions. This was done because there is no previous research that has investigated the visual segmentation of hand/arm actions. By including many different actions, the results could potentially have a greater external validity than if the segmentation stimuli were limited to fewer actions.

A similar line of reasoning can be applied to the freedom that was given to participants for the task of placing segmentation marks, which also likely contributed to the variability. Participants were given the possibility of moving the yellow slider (**Figure 4**) in order to find the frame they wanted to mark as a breakpoint between segments, and all participants used the slider to try and find the breakpoints between segments, but they realized it took far too much time to exactly specify the frame for each breakpoint for 12 actions. We did not require exact precision in the segment marking task because we wanted to avoid demand characteristics associated with too many constraints on the segmentation task. The smoothing function mentioned above was used to treat this variation and has been used by other researchers on similar data (Meyer et al., 2010), or alternatively, a binning technique has been used (e.g., Zacks, 2004).

VERBAL DESCRIPTIONS AND RECOGNITION

The results from the verbal descriptions show that the orientation manipulation for the point-light hand/arm actions successfully created top-down and bottom-up driven visual processing of the stimuli. This is consistent with the findings from previous studies of whole body point-light displays (e.g., Bertenthal and Pinto, 1994; Pavlova and Sokolov, 2000). Recognition of the actions was obviously impaired by inverting the action. Despite the severe impairment of not being able to recognize the actions, participants were still able to consistently use the kinematics to mark segments of the actions, which was indicated by the significant correlations with change in direction, velocity, and acceleration. Furthermore, the segmentation behavior of participants who viewed the inverted displays was reliably correlated with the segmentation behavior of the participants who viewed the actions in an upright orientation. This shows that both groups seem to base their segmentation of the actions on the more low-level kinematics rather than high-level knowledge associated with the conceptual understanding of the viewed actions. The participants in the no-picture-inverted condition seem to do this a bit more. Top-down influences do not seem to modulate segmentation behavior such that they lead to very different segmentation behavior. Our results are consistent with the results from Zacks et al. (2009) and Bidet-Ildei et al. (2006). In the results from Bidet-Ildei et al. (2006) participants were able to reliably discriminate between natural and unnatural arm movements of point-light displays of elliptic motion but were very poor at identifying the display as a specific arm movement of a human making an elliptic motion. Results from our experiment show that the inversion manipulation seems to work on even smaller grained actions (more local limb motions). There appears to be local motion processing of the limb parts, e.g., fingers and hand, and more global (holistic) processing seems to be impaired.

The fact that participants can see and describe the movements of body parts but fail to identify the higher-level semantic meaning of the actions is similar to association agnosia for objects where patients can see the parts of objects but fail to identify the object as such (Farah, 2004). There is no strictly visual deficit as such but rather an inability to recognize the object. When our participants view the inverted point-light actions, they are able to visually discern the relevant body parts and segment the actions on the basis of changes in the direction of movement and velocity. What seems to be missing is the epistemic visual perception (Jeannerod and Jacob, 2005).

POSSIBLE INFLUENCE OF TOP-DOWN PROCESSING

One difference mentioned above between the two viewing conditions is that some of the correlation coefficients were somewhat higher for the no-picture-inverted group. Although segmentation marking appears to be similar for the two conditions, participants in the no-picture-inverted condition tend to base their segmentation marks somewhat more on the kinematics than the participants in the picture group. One explanation for this behavior could be that participants with access to higher-level meaning of actions tend to rely less on the precise pattern of the kinematics during action segmentation. The results from the verbal descriptions seem to confirm this as participants with access to higher-level information tend to describe segmentation instances based on this [e.g., "reached out his hand, gripped the (door) handle and opened the door"] whereas the second group (no-picture-inverted) tended to refer more closely to actual movements (e.g., "grips something from underneath and turns it"). One such example of where top-down information appears to influence the segment marking can be seen in the case of *opening a door*. Toward the end of that action, participants who were shown a picture of a door handle clearly marked the releasing of the grip on the door handle, i.e., segmentation was about the grip. For the group that saw the action upside down and without a picture, marks were made in connection with the motion of the hand downward toward a resting position. The major difference here is that for the picture group, the segmentation was about the grip, and for the no-picture group the segmentation was about the movement of the arm/hand. This suggests that top-down knowledge seems to involve gripping as an important segment for opening a door, which is not similarly marked for the group that did not have access to any top-down knowledge about the action. Future experiments will have to look more closely at this potential difference.

Another possibility for a potential role of high-level knowledge concerns action hierarchies. Poggio and Bizzi (2004) discuss the combination of motor primitives into hierarchical representations that allow generalization over specific situations. If participants are given the task of segmenting inverted and upright actions on the basis of very coarse-grained patterns and not allowed to place as many breakpoints as they wish, we should see more pronounced differences between top-down and bottom-up processing.

We are thus not making the strong claim that there is no modulation of the segmentation of action sequences by top-down visual processing. It could be case that given other manipulations, epistemic visual perception may lead to different segmentation strategies. However, our results regarding the overall segmentation behavior do indicate that the different strategies may nonetheless converge onto similar marking behavior.

MOTOR PRIMITIVES AND THE MIRROR SYSTEM

As discussed in the introduction, it is still an open question regarding what motion primitives are encoded by the mirror system. The chain model (Chersi et al., 2006) is primarily based on neurophysiological data from Macaque monkeys (e.g., Fogassi et al., 2005) and therefore includes only very basic primitives that correspond to easy tasks that monkeys can carry out in laboratory settings. It is an open question how well, if at all, those findings can translate to a human mirror system and therefore, it is of interest to identify potential "human" motion primitives.

Since it is at present almost impossible to record from relevant human neurons at an adequate resolution to answer this question, we have chosen an action segmentation task instead. Critically, while the segmentation behavior of the no-picture-inverted (bottom-up) group in several instances correlated more closely with the kinematics than the picture group, our results indicate that the overall segmentation behavior nonetheless remains similar; it thus seems that primitives which are identifiable in a task such as the one in the present paper are defined primarily by the kinematics of the actions rather than additional contextual information, although the latter can clearly influence the identification of start and end points of primitives.

This is interesting in the light of the argument that mirror neurons encode not only a motion primitive but also the goal of the action an observed or executed primitive is part of, e.g., Fogassi et al. (2005), Umiltà et al. (2008). One could infer from such an entanglement of encodings that higher-level contextual information affects the definition of primitives. However, our results indicate that this is not the case.

It should also be noted here that kinematics are not necessarily equal to motor commands or muscle activations. Umiltà et al. (2008) for instance have shown that mirror neurons in monkeys do not encode the motor commands needed to execute an action but rather the behavior of the end-effector, and the role of kinematic variables in action segmentation studied here reflect that aspect.

CONCLUSION

When participants are given the task of segmenting hand/arm actions presented as point-light displays, segmentation is largely based on the kinematics, i.e., the velocity and acceleration of the wrist and hand extremities (finger tips), regardless of whether or not participants have access to higher-level information about the action. If access to high-level information about the identification of the action, e.g., drinking) is impaired by inverting the point-light displays, the kinematic information remains a salient source of information on which to base action segmentation. If participants have access to the high-level information, they still tend to rely on the kinematics of the hand/arm actions for determining where to place segmentation marks. This indicates that top-down activation of an action representation leads to similar segmentation behavior for hand/arm actions compared to bottom-up, or local, visual processing when performing a fairly unconstrained segmentation task. Future studies will need to address the issues of more precisely identifying motor primitives and determining their hierarchical organization in relation to high-level knowledge structures.

ACKNOWLEDGMENTS

This work has been supported by the FP7 project ROSSI, "Emergence of communication in RObots through Sensorimotor and Social Interaction," Grant agreement no. 216125, and by the University of Skövde. We thank the reviewers for constructive comments and suggestions and to Mikael Lebram for his valuable technical assistance.

SUPPLEMENTARY MATERIAL

The Movies 1 and 2 for this article can be found online at http://www.frontiersin.org/cognition/10.3389/fpsyg.2010.00243/abstract

REFERENCES

Bertenthal, B. I., and Pinto, J. (1994). Global processing of biological motions. *Psychol. Sci.* 5, 221–225.

Bidet-Ildei, C., Orliaguet, J.-P., Sokolov, A. N., and Pavlova, M. (2006). Perception of elliptic biological motion. *Perception* 35, 1137–1147.

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73.

Blakemore, S.-J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567.

Borghi, A. M., and Riggio, L. (2009). Sentence comprehension and simulation of object temporary, canonical and stable affordances. *Brain Res.* 1253, 117–128.

Boutsen, L., and Humphreys, G. W. (2003). The effect of inversion on the encoding of normal and "thacherized" faces. *Q. J. Exp. Psychol.* 56A, 955–975.

Buccino, G., Binkofski, F., and Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang.* 89, 370–376.

Carey, S., and Diamond, R. (1994). Are faces conceived as configurations more by adults than by children? *Vis. Cogn.* 1, 253–275.

Castellini, C., Orabona, F., Metta, G., and Sandini, G. (2007). Internal models of reaching and grasping. *Adv. Robot.* 21, 1545–1564.

Chersi, F., Mukovskiy, A., Fogassi, L., Ferrari, P. F., and Erlhagen, W. (2006). A model of intention understanding based on learned chains of motor acts in the parietal lobe. *Comput. Neurosci.* 69, 48.

Chersi, F., Thill, S., Ziemke, T., and Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Front. Neurorobotics* 4:4. doi: 10.3389/fnbot.2010.00004

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events. *Exp. Brain Res.* 91, 176–180.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception* 22, 15–22.

Farah, M. J. (2004). *Visual Agnosia*. Cambridge, MA: MIT Press.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667.

Giese, M. A., and Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Res.* 38, 1847–1858.

Grafton, S. T., Fadiga, L., Arbib, M. A., and Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage* 6, 231–236.

Hemeren, P. E. (2005). "Orientation specific effects of automatic access to categorical information in biological motion perception," in *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, eds B. G. Bara, L. W. Barsalou, and M. Bucciarelli (Hillsdale, NJ: Erlbaum), 935–940.

Hemeren, P. E. (2008). *Mind in Action: Action Representation and the Perception of Biological Motion*. Lund University Cognitive Studies 140. Sweden: Lund University.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3, e79. doi: 10.1371/journal.pbio.0030079

Jeannerod, M. (1995). Mental imagery in the motor cortex. *Neuropsychologia* 33, 1419–1432.

Jeannerod, M., and Jacob, P. (2005). Visual cognition: a new look at the two-visual systems model. *Neuropsychologia* 43, 301–312.

Leder, H., and Carbon, C.-C. (2006). Face-specific configural processing of relational information. *Br. J. Psychol.* 97, 19–29.

Meyer, M., DeCamp, P., Hard, B., Baldwin, D., and Roy, D. (2010). "Assessing behavioral and computational approaches to naturalistic action segmentation," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson and R. Catrambone (Austin, TX: Cognitive Science Society), 2710–2715.

Michelon, P., Vettel, J. M., and Zacks, J. M. (2006). Lateral somatotopic organization during imagined and prepared movements. *J. Neurophysiol.* 95, 811–822.

Mussa-Ivaldi, F. A., and Bizzi, E. (2000). Motor learning through the combination of primitives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 1755–1769.

Newtson, D., and Engquist, G. (1976). The perceptual organization of ongoing behavior. *J. Exp. Soc. Psychol.* 12, 436–450.

Newtson, D., Engquist, G., and Bois, J. (1977). The objective basis of behavior units. *J. Pers. Soc. Psychol.* 35, 847–862.

Pavlova, M., and Sokolov, A. (2000). Orientation specificity in biological motion perception. *Percept. Psychophys.* 62, 889–899.

Poggio, T., and Bizzi, E. (2004). Generalization in vision and motor control. *Nature* 431, 768–774.

Poinzner, H., Bellugi, U., and Lutes-Driscoll, V. (1981). Perception of American sign language in dynamic point-light displays. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 430–440.

Pollick, F. E. (2004). "The features people use to recognize human movement style," in *Gesture-Based Communication in Human–Computer Interaction*, Vol. 2915, eds A. Camurri and G. Volpe (Berlin: Springer-Verlag), 10–19.

Pollick, F. E., Fidopiastis, C., and Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception* 30, 323–338.

Pozzo, T., Papaxanthis, C., Petit, J. L., Schweighofer, N., and Stucchi, N. (2006). Kinematic features of movement tunes perception and action coupling. *Behav. Brain Res.* 169, 75–82.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2006). Mirrors in the mind. *Sci. Am.* 295, 54–61.

Saxe, R., Xiao, D.-K., Kovacs, G., Perrett, D. I., and Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446.

Shiffrar, M., and Pinto, J. (2002). "The visual analysis of bodily motion," in *Common Mechanisms in Perception and Action: Attention and Performance*, Vol. 19, eds W. Prinz and B. Hommel (Oxford: Oxford University Press), 381–399.

Sumi, S. (1984). Upside-down presentation of the Johansson moving light pattern. *Perception* 13, 283–286.

Thoroughman, K. A., and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature* 407, 742–747.

Umiltà, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V., and Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2209–2213.

van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cogn. Sci.* 28, 979–1008.

Zacks, J. M., Kumar, S., Abrams, R. A., and Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition* 112, 201–216.

Zacks, J. M., and Swallow, K. M. (2007). Event segmentation. *Curr. Dir. Psychol. Sci.* 16, 80–84.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 August 2010; accepted: 23 December 2010; published online: 27 January 2011.

Citation: Hemeren PE and Thill S (2011) Deriving motor primitives through action segmentation. Front. Psychology ***1****:243. doi: 10.3389/fpsyg.2010.00243*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Hemeren and Thill. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 16 November 2010
doi: 10.3389/fpsyg.2010.00184

The embodied statistician

Elizabeth R. Marsh[1]* and Arthur M. Glenberg[1,2]

[1] *Laboratory for Embodied Cognition, Department of Psychology, Arizona State University, Tempe, AZ, USA*
[2] *Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA*

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Sian Beilock, University of Chicago, USA
Michael Spivey, University of California, USA

****Correspondence:***
Elizabeth R. Marsh, Department of Psychology, Arizona State University, 950 S. McAllister Ave, Tempe, AZ 85287, USA.
e-mail: ermarsh@asu.edu

How do infants, children, and adults learn grammatical rules from the mere observation of grammatically structured sequences? We present an embodied hypothesis that (a) people covertly imitate stimuli; (b) imitation tunes the particular neuromuscular systems used in the imitation, facilitating transitions between the states corresponding to the successive grammatical stimuli; and (c) the discrimination between grammatical and ungrammatical stimuli is based on differential ease of imitation of the sequences. We report two experiments consistent with the embodied account of statistical learning. Experiment 1 demonstrates that sequences composed of stimuli imitated with different neuromuscular systems were more difficult to learn compared to sequences imitated within a single neuromuscular system. Experiment 2 provides further evidence by showing that selectively interfering with the tuned neuromuscular system while attempting to discriminate between grammatical and ungrammatical sequences disrupted performance only on sequences imitated by that particular neuromuscular system. Together these results are difficult for theories postulating that grammatical rule learning is based primarily on abstract statistics representing transition probabilities.

Keywords: statistical learning, implicit imitation, artificial grammar, embodied cognition, fluency

INTRODUCTION

We speak grammatical English without being able to say much about the rules of English grammar. How is that procedural knowledge learned, and where does it reside? One answer is given by the study of statistical learning, which is the ability to learn (often without intent) which sequences of stimuli are consistent with a set of rules. It has been observed in both infants and adults, in the visual, auditory, and tactile domains, and in a variety of stimulus displays from simple to complex to real-world scenes (Reber, 1967; Saffran et al., 1996, 1999; Fiser and Aslin, 2001, 2005; Saffran, 2001; Creel et al., 2004; Conway and Christiansen, 2005; Brady and Oliva, 2008). One explanation of the phenomenon is that it reflects a domain-general learning process: the brain is wired to pick up and represent statistics. Another (Conway and Christiansen, 2006) is that this learning is modality-constrained. We describe and test an embodied mechanism for statistical learning that is consistent with the notion of modality constraints but that also suggests why the phenomenon is robust enough to be classified as a general learning process.

Artificial grammar learning (AGL, e.g., Reber, 1967; Conway and Christiansen, 2005) is a paradigm case of statistical learning. Participants are exposed to sequences of stimuli such as tones, visual patterns, or tactile sensations. Attention to the stimuli is enforced by, for example, asking the participant to judge if two successive sequences are identical. After exposure to the sequences, participants are asked to discriminate between (a) sequences consistent with the rules used to generate the training sequences and (b) sequences generated by other rules. Although participants often profess no knowledge of the rules, they are successful in making the discrimination. How do they do this?

Our theory has three tenets. First, when people attend to stimuli, they concurrently imitate successive stimuli, often without awareness. Imitation of this sort is documented both behaviorally (e.g., Chartrand and Bargh, 1999) and using neurophysiological measures (e.g., Rizzolatti and Craighero, 2004). Wilson (2001) presents a review of the literature supporting the claim that people reliably imitate, and she suggests a wide variety of functions of imitation.

Second, imitation is a neuromuscular phenomenon, that is, it requires cortical mechanisms that control particular effectors (e.g., the speech articulators, the hands, etc.) to generate the motor commands that would produce an imitation of the stimulus, even if the actual production is inhibited (e.g., Grush, 2004; Iacoboni, 2008). As elements of the sequence are imitated, the particular neuromuscular system used to produce the imitations is forced to make transitions between neuromuscular states. In this manner, the neuromuscular system becomes tuned to the environment (the grammar) so that some sequences, the grammatical ones, can be imitated with alacrity. This tuning is consistent with the literature on use-induced plasticity (e.g., Classen et al., 1998).

Third, when participants are asked to determine if a sequence is grammatical, they base their judgments on the implicit fluency with which they can imitate the sequence. Because of previous tuning, grammatical sequences are imitated more fluently than non-grammatical sequences. This account of statistical learning is similar in kind to the account of the mere exposure effect provided by Topolinski and Strack (2009), as well as the differences observed in the recognition memory of skilled versus unskilled typists for fluent letter dyads (Yang et al., 2009).

We tested this embodied account of statistical learning in two experiments. In the first experiment, we show that people have little difficulty discriminating between grammatical and ungrammatical sequences when the successive stimuli in a sequence can be imitated within the same neuromuscular system. However, when the same grammatical rules are instantiated by stimuli imitated using different neuromuscular systems, performance drops to chance levels.

These results are consistent with the embodied account of statistical learning presented here, but they leave open the possibility of an alternative, attention-based account of domain-general statistical learning. In the second experiment, we directly test the importance of the domain-specific fluency gained from imitation of the training sequences by demonstrating that occupying neuromuscular system A disrupts the ability to discriminate between grammatical and ungrammatical sequences that are imitated using system A. Importantly, however, occupying system A has little effect on discrimination when the sequence is imitated using system B. The results suggest that the ability to access information in the form of domain-specific fluency is a large contributor to the subsequent recognition of grammatical sequences.

EXPERIMENT 1

We used a modified AGL paradigm similar to the one used by Conway and Christiansen (2005). During training, participants were exposed to pairs of grammatical sequences, and decided if the sequences within each pair were identical. During the test phase, participants were exposed to novel sequences. Half of the test sequences were generated by the grammar and half violated the grammar. Participants were asked to decide if each novel sequence was grammatical.

The key to the experiment is that participants received a single grammar expressed simultaneously in two modalities, auditory tones and visual boxes at different locations (**Figure 1**). During the test phase, however, they received only half of the information contained in the training sequences: auditory-only, visual-only, or alternating modalities. An alternating sequence begins with either a visual or auditory stimulus, and then it is followed by a stimulus in the other modality. This modality alternation continues until the end of the sequence.

Exposure to the dual modality sequences during training should engender imitation of the tones using the vocal folds (i.e., humming) and subsequent tuning of the laryngeal neuromuscular system. This tuning should allow easy discrimination between grammatical and ungrammatical single-modality auditory sequences at the test. Similarly, the dual modality sequences at training should engender tuning of the eye movement system to follow the sequence of spatial locations. This tuning should allow easy discrimination between grammatical and ungrammatical single-modality visual sequences at the test.

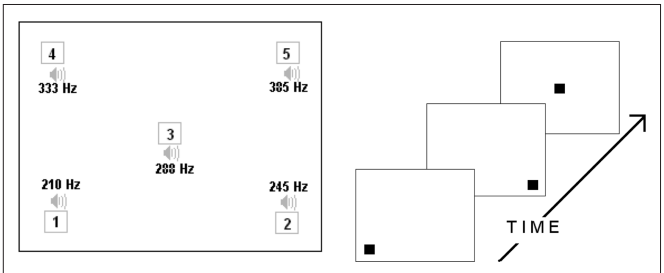


FIGURE 1 | Left: The large rectangle represents the layout of the computer screen for Experiments 1 and 2. The five numbered boxes show the five possible locations of the visual stimuli and beside them are the five possible tones (in Hz) associated with each location. Right: A graphical representation of the visual sequence "1 2 3" as it would have appeared.

Consider, however, the alternating sequences. The alternating grammatical sequences instantiate exactly the same grammar used during training and used during the tests in the single-modality conditions. Thus, the alternating statistical regularities are exactly the same as the single-modality statistical regularities. If those regularities alone are the primary basis of performance, then performance should be the same as in the single-modality case.

On the embodied account, however, the imitation of sequences during training does not produce cross-modality tuning as readily as single-modality tuning. This is because the eye movement system tunes itself to produce sequences of eye movements, and the laryngeal system tunes itself to produce sequences of hummed tones. However, the eye movement system does not tune itself to produce a hummed tone, and vice versa. This is not to say that cross-modal tuning is impossible given the embodied account of statistical learning. Cross-modal tuning likely occurs on a regular basis in natural settings (e.g., the visual stimulus "D-O-G" may tune the speech articulators to produce the sound "dog"); this type of cross-domain mapping may only occur after considerable repetition relative to the sort of single-system tuning described here, which should arise out of biomechanical necessity.

To better understand why cross-modal tuning should not occur here, consider the neuromuscular systems used in imitating an alternating sequence. Starting with a stimulus 1 at a particular visual location, the eye movement system is tuned to move to other locations (for stimulus 2) that form a grammatical sequence. However, stimulus two is a tone. On imitating the tone, the laryngeal system is tuned to produce other tones that conform to stimulus 3, however stimulus 3 is a visual stimulus. Note that we left the eye movement system ready to move to the location of stimulus 2, but instead it must direct the eyes to the location of stimulus 3. This is an unpracticed transition that should be produced with less fluency than a practiced transition. Thus, when imitating the alternating test sequences, grammatical and ungrammatical sequences should be imitated with equal dysfluency, and discrimination performance should drop dramatically.

MATERIALS AND METHODS

This study was approved by the Arizona State University (ASU) IRB and was conducted in accordance with the approved standards. All participants gave informed consent. The participants were 20 Arizona State University undergraduates recruited from introductory psychology classes who received course credit. Four participants were excluded due to computer errors.

Visual stimuli were presented on 15" LCD monitors, and auditory stimuli were presented using headphones. Participants responded using the keyboard.

A modified version of the Gomez and Gerken (1999) grammar was used to create 12 grammatical sequences, ranging from 3 to 6 stimuli in length and composed of the numbers 1–5. Each of the numbers referred to one of five locations on the computer screen (each corner and the middle) at which a small black box was presented. In addition, each number referred to one of five tone frequencies (210, 245, 288, 333, and 385 Hz). The training sequences were constructed so that the box locations and tones were completely redundant with one another. Items within a single

sequence were displayed for 250 ms with a 150 ms inter-item interval; a sequence was separated from its mate within the training pair by a 1000 ms.

Half of the 24 novel test sequences conformed to the training grammar and half did not. The ungrammatical sequences contained the same beginning and ending stimuli as the grammatical sequences and differed only in internal transitions. Eight test sequences were used for each of the auditory, visual, and alternating tests. We counterbalanced both the order of the test-sequence modalities and the assignment of eight sequences to the modalities.

RESULTS

Proportions correct on the auditory and visual test sequences were above chance (0.50); Auditory-only sequences: $M = 0.625$, SEM = 0.045, $t(19) = 2.77$, p-rep = 0.942; visual-only sequences ($M = 0.625$, SEM = 0.054), $t(19) = 2.33$, p-rep = 0.907. However, performance on the alternating sequences was at chance ($M = 0.481$, SEM = 0.032), $t(19) = -0.590$, p-rep = 0.456. Tests of within-subjects contrasts demonstrated a significant difference between the auditory-only and the alternating sequences, $F(1,19) = 6.24$, p-rep = 0.979, and a significant difference between the visual-only and the alternating sequences, $F(1,19) = 7.33$, p-rep = 0.940.

These results are in accord with the predictions of the embodied account of statistical learning. There remains, however, an alternative explanation that is in keeping with a domain-general mechanism. It is possible that the learning of alternating sequences is constrained by attention to one modality at a time. Though it has been demonstrated that people can learn two separate grammars at one time when the stimuli comprising each grammar are perceptually distinct (Conway and Christiansen, 2006), the cost of switching attention between perceptually distinct stimuli within a single sequence may account for the apparent dysfluency for alternating sequences. Thus, the attention-switching account would claim that it is not necessarily the motor-fluency induced by modality-specific imitations that contributes to learning, but that attention is modality-specific. The resulting knowledge may still be domain-free statistics that can be applied to any single-modality sequences at the time of test (as this should produce no added demand on attentional resources). In this case, statistics abstracted from the auditory sequences alone should provide adequate recognition of grammatical sequences instantiated in the visual modality, and vice versa. However, the alternating sequences (and consequent attention-switching costs) disrupt the application of this domain-free statistical knowledge. Consequently, we tested another prediction of the embodied account.

EXPERIMENT 2

We used a selective interference task to test between the embodied account and the attention-switching account. The learning phase of this experiment was identical to that of Experiment 1. In the test phase, participants judged the grammaticality of both auditory and visual sequences while performing one of three secondary tasks: humming two notes requiring alternating changes in the laryngeal system (making the high-low sound of a siren), mouth sounds requiring alternating changes in the speech articulators (saying "da-da"), and feet sounds (alternating stomps with the left and right feet). Sounds were demonstrated for the participants via the headphones to ensure that they understood the task. Furthermore, this demonstration ensured that participants produced the sounds with similar timing (i.e., one hummed note took approximately the same time to produce as one "da" and as one stomp).

The embodied stance predicts that humming should drastically interfere with imitation using the laryngeal system and thereby disrupt discrimination between grammatical and ungrammatical auditory sequences. The other two tasks should be far less disruptive for the auditory sequences (note that the da-da task requires transitions in the speech articulators but few transitions in the vocal folds). None of the tasks should interfere with discrimination of the visual sequences because they do not affect transitions of the eyes between successive stimulus locations.

The attention-switching hypothesis, however, suggests a number of possible outcomes. First, it might be the case that the presence of the secondary task will disrupt performance ubiquitously, since attention is being divided in all test conditions. Second, it might be the case that performance on visual sequences alone is affected, since attention is being divided between production of an auditory stimulus and perception of a visual sequence. Though it cannot be ruled out, this seems a slim possibility. More likely is the third possibility. That is, the production of auditory stimuli should interfere with the perception of auditory stimuli alone because attention is being divided within a single modality. Note, however, that this account provides no clear reason to assume any difference in the level of interference associated with the production of one sound versus another. The disruption of auditory sequences should be ubiquitous.

MATERIALS AND METHODS

This study was approved by the Arizona State University (ASU) IRB and was conducted in accordance with the approved standards. All participants provided informed consent. The participants were 69 ASU undergraduates and staff. Participants received $10 in exchange for 1 h of participation. During the hour, participants completed this and an unrelated study. Three participants were excluded for not following instructions.

All materials were identical to those used in Experiment 1, as was the training phase procedure. In the testing phase, participants were exposed to 16 sequences in each modality while engaged in one of the three secondary tasks (manipulated between participants). The order of the modalities was counterbalanced.

RESULTS

Means are presented in **Table 1**.

The groups did not differ in regard to performance during training (all ps > 0.25). The major analysis was based on two, single df contrasts to test predictions derived from the embodied theory. The first contrast compared the difference between the auditory and visual test sequences during the siren task to the same difference during the da-da and stomping tasks. As predicted, the difference was larger for the siren task than for the others, $t(66) = 1.97$, p-rep = 0.88. The second contrast compared the visual-auditory difference for the da-da task to the difference for the stomping task, and as predicted this effect was not significant, $t(66) < 1$.

Table 1 | For each of the three secondary tasks, number correct during training (standard errors in parentheses), and proportion correct during the test phase.

Group	Training	Auditory test	Visual test
Siren (n=22)	69.32 (0.59)	0.59 (0.03)	0.70 (0.04)
Da-da (n=25)	67.76 (0.81)	0.66 (0.03)	0.66 (0.04)
Stomp (n=22)	68.77 (0.67)	0.66 (0.04)	0.66 (0.04)

These data challenge the attention-switching alternative in a number of ways. First, there is no clear reason why any of the interfering tasks should disrupt performance based on these abstract symbols unless attention was being divided at the time of test. On this account, however, there is no reason why the siren task should disrupt performance with the auditory sequences more than performance with the visual sequences. If the disruption was based on a masking effect of the sounds (unlikely given that headphones were used to present the stimuli) or the subdivision of attention within the auditory modality, then the da-da task should also have disrupted performance, but it did not.

In contrast, the results are in accord with the embodied account. Discriminating between grammatical and ungrammatical auditory sequences requires using the larynx to imitate the auditory stimuli. The siren task occupies that neuromuscular system and thereby reduces performance. Because the da-da task requires changes in the lips and tongue more than the larynx, it does not disrupt performance. Discriminating between grammatical and ungrammatical visual sequences requires using the eyes, possibly along with neck musculature, to imitate (or follow) the locations of the visual stimuli. None of the secondary tasks disrupt eye movements, and hence there is little disruption in discrimination performance.

Another reason for discounting the attention-switching account is that we have been able to demonstrate selective interference in the visual modality in the absence of any need for attention-switching. Marsh and Glenberg (2010) manipulated the participants' head orientation relative to the computer screen from training to test. In this experiment, participants either faced the screen directly or at a 45° angle during training. During test, this orientation was either preserved or switched to the alternative orientation. Participants who maintained the same head orientation (thus using the same eye movements during training and testing) performed normally on the visual and auditory test sequences. Participants who switched head orientation between training and test (thus using different, un-tuned eye movements during the test), however, performed poorly on the visual test sequences while performing well above chance on the auditory sequences. This finding provides further support for the generality of neuromuscular tuning account. Note that changing the orientation of the head to the screen does not require any attention-switching, so an attention-switching account would have difficulty predicting these results.

Thus, in Experiment 2, selectively interfering with the laryngeal system reduces discrimination between grammatical and ungrammatical auditory stimuli, but not visual stimuli. And, as found in Marsh and Glenberg (2010), selectively interfering with the eye movement control system reduces discrimination between grammatical and ungrammatical visual stimuli, but not auditory stimuli. Although one might postulate special attentional demands within each of these systems (e.g., the siren task disrupts auditory attention but the da-da task does not), such an account would be highly unparsimonious to say the least.

DISCUSSION

Can our simple embodied account accommodate all of the facts of statistical learning? Probably not. For instance, the last parts of an auditory sequence are learned best, compared to the first parts of a tactile sequence (Conway and Christiansen, 2005). This is not predicted by our account; but then neither is it by the analytic process account.

Nonetheless, embodiment offers a compelling account of a good proportion of the literature as well as novel predictions. For example, it predicts the learning of non-adjacent dependencies (e.g., between the first and third stimulus), will be facilitated if the intervening stimulus is imitated with a different neuromuscular system. Suppose that the first stimulus is tonal, the second visual, and the third tonal. Imitation of first stimulus leaves the laryngeal system in a particular state that is not affected by imitation of the visual stimulus. Then, imitation of the third stimulus produces a transition from the state of the laryngeal system after humming the first stimulus to the state of the system used to hum the third. In fact, Gebhart et al. (2009) report that the learning of remote dependencies only occurs when intervening stimuli are dissimilar to their flankers.

The embodied account also explains how Conway and Christiansen (2006) participants could manage to learn two different grammars when sequences from the grammars were interleaved and presented in different modalities. In Conway and Christiansen, learning tuned two neuromuscular systems, and testing required tuned transitions within each of the systems. In, contrast, in Experiment 1, learning tuned two systems, but testing required un-tuned transitions across systems.

Finally, Saffran et al. (2008) demonstrated that cotton-top tamarin monkeys can learn simple grammars instantiated by five spoken syllables, but not a more complex grammar instantiated by eight syllables. Our account predicts that if the complex grammar were instantiated in a neuromuscular system the tamarins can use more successfully for imitation (e.g., a reaching system), then the tamarins would have greater success in learning the complex grammar.

In the first paragraph of the introduction, we noted that the embodied mechanism for statistical learning is consistent with the notion of modality constraints, but that it also suggests why the phenomenon is robust enough to be classified as a general learning process. We think that neuromuscular tuning is likely to be found in all neuromuscular systems. Thus, one should be able to find statistical learning in for example, eye movement control, speech articulation, humming, finger movements, and so on. Thus, the learning mechanism (neuromuscular tuning) is found across many domains, and at the same time it is highly embodied.

Few native speakers can explain the rules with which grammarians describe their behavior. The present study suggests that such rules are embodied in differentially conditioned neuromuscular networks; the body has rules that its mind does not, and need not, know.

ACKNOWLEDGMENTS

We thank Andre Valdez for discussions about testing the embodied hypothesis, and Jenny Saffran and Peter Killeen for their insightful comments. We also thank Mathew Dexheimer and Frederick Raehl for assistance in data collection. Work on this article was supported by the National Science Foundation under Grant No. BCS 0744105 to Arthur Glenberg, and an NSF IGERT award to Elizabeth Marsh. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Brady, T. F., and Oliva, A. (2008). Statistical learning using real world scenes: extracting categorical regularities without conscious intent. *Psychol. Sci.* 19, 678–684.

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910.

Classen, J., Liepert, J., Wise, S. P., Hallett, M., and Cohen, L. G. (1998). Rapid plasticity of human cortical movement representation induced by practice. *J. Neurophysiol.* 79, 1117–1123.

Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39.

Conway, C. M., and Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychol. Sci.* 17, 905–912.

Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1119–1130.

Fiser, J., and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504.

Fiser, J., and Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 4, 521–537.

Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–490.

Gomez, R. L., and Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70, 109–135.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442.

Iacoboni, M. (2008). The role of premotor cortex in speech perception: evidence from fMRI and rTMS. *J. Physiol. (Paris)* 102, 31–34.

Marsh, E. R., and Glenberg, A. M. (2010). "Head orientation affects visual statistical learning," in *Poster Presented at the Annual Meeting of the Psychonomics Society*, St. Louis, MO.

Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition* 81, 149–169.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, J. R., Hauser, M., Seibel, R., Kpfhamer, J., Tsao, F., and Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

Topolinski, S., and Strack, F. (2009). Motormouth: mere exposure effects depend on stimulus-specific motor-simulations. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 423–433.

Wilson, M. (2001). Perceiving imitatible stimuli: consequences of isomorphism between input and output. *Psychol. Bull.* 127, 543–553.

Yang, S., Gallo, D. A., and Beilock, S. L. (2009). Embodied memory judgments: a case of motor fluency. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1359–1365.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 July 2010; paper pending published: 20 August 2010; accepted: 10 October 2010; published online: 16 November 2010.

Citation: Marsh ER and Glenberg AM (2010) The embodied statistician. Front. Psychology ***1****:184. doi: 10.3389/fpsyg.2010.00184*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Marsh and Glenberg. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 25 November 2010
doi: 10.3389/fpsyg.2010.00212

Embodied conceptual combination

Dermot Lynott[1]* and Louise Connell[2]

[1] *Manchester Business School, University of Manchester, Manchester, UK*
[2] *School of Psychological Sciences, University of Manchester, Manchester, UK*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Michiel van Elk, Radboud University Nijmegen, Netherlands
Lara Jones, Wayne State University, USA

****Correspondence:***
Dermot Lynott, Decision and Cognitive Sciences Research Centre, Manchester Business School, University of Manchester, Manchester M15 6PJ, UK, e-mail: dermot.lynott@manchester.ac.uk;
Louise Connell, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, UK. e-mail: louise.connell@manchester.ac.uk

Conceptual combination research investigates the processes involved in creating new meaning from old referents. It is therefore essential that embodied theories of cognition are able to explain this constructive ability and predict the resultant behavior. However, by failing to take an embodied or grounded view of the conceptual system, existing theories of conceptual combination cannot account for the role of perceptual, motor, and affective information in conceptual combination. In the present paper, we propose the embodied conceptual combination (ECCo) model to address this oversight. In ECCo, conceptual combination is the result of the interaction of the linguistic and simulation systems, such that linguistic distributional information guides or facilitates the combination process, but the new concept is fundamentally a situated, simulated entity. So, for example, a *cactus beetle* is represented as a multimodal simulation that includes visual (e.g., the shiny appearance of a *beetle*) and haptic (e.g., the prickliness of the *cactus*) information, all situated in the broader location of a desert environment under a hot sun, and with (at least for some people) an element of creepy-crawly revulsion. The ECCo theory differentiates interpretations according to whether the constituent concepts are destructively, or non-destructively, combined in the situated simulation. We compare ECCo to other theories of conceptual combination, and discuss how it accounts for classic effects in the literature.

Keywords: embodied cognition, conceptual combination, linguistic distributional knowledge, situated simulation

INTRODUCTION

Cognition is inherently constructive. Our cognitive functioning is not confined to retrieving familiar ideas and concepts, but rather is predicated upon the ability to understand new things and represent new concepts. Conceptual combination research investigates the processes involved in creating and understanding new meanings from old referents. For example, how do people interpret novel combinations such as *cactus beetle*, *mouse potato*, or *fame advantage*? Such combinations are used liberally (conversations, newspaper headlines, signage, novels, etc.) and people generally have little difficulty in constructing plausible interpretations, even where the surrounding context may be quite limited or uninformative.

Of course, central to understanding how people process these combinations is an understanding of what constitutes the representations of these concepts. Of existing theories of conceptual combination, many take an explicitly a modal view of the conceptual system (Wisniewski, 1997; Estes and Glucksberg, 2000). That is, concepts are represented in some abstracted format (e.g., feature lists, propositional representations) that do not relate directly to the modality-specific experiential basis of how these concepts were acquired. Other theories are agnostic as to the nature of the underlying representation (Gagné and Shoben, 1997; Costello and Keane, 2000), often using propositions as a descriptive or computational shortcut, but without making strong representational commitments. We suggest that any theory that fails to take an embodied or grounded view of the conceptual system cannot account for the role of perceptual, motor, affective, introspective, and social information in conceptual combination and cognition more generally (Barsalou et al., 2008). The flip side of this argument is that current embodied theories of cognition must also draw on theories of conceptual combination in order to explain the constructive, generative, and creative capacities of human cognition. In this paper, we aim to address these oversights in proposing a theoretical model of conceptual combination, embodied conceptual combination (ECCo) which draws on recent empirical and theoretical work in areas of language processing, mental representation, and links between our perceptual and conceptual systems.

In ECCo, we outline an embodied conceptual combination system based on a representation of knowledge that incorporates linguistic distributional information and situated simulation. Linguistic information guides or facilitates the simulation process, but the new concept created during conceptual combination is fundamentally a situated, simulated entity. The paper is divided into four main sections. In the rest of the introduction, we outline the structure of the conceptual system and review evidence for the roles of the linguistic and simulation systems during conceptual processing. In Section "ECCo: Embodied Conceptual Combination", we outline the core tenets of the ECCo theory and explain how it accounts for classical conceptual combination effects. In Section "Comparison of ECCo with Previous Theories of Conceptual Combination", we illustrate how ECCo relates to existing theories of conceptual combination and highlight key differences. Lastly, in Section "Concluding Remarks", we summarize and briefly consider future directions for conceptual combination research.

THE CONCEPTUAL SYSTEM

Embodied theories of cognition hold that our perceptual, motor, and other experience plays a fundamental role in how we talk about, think about and interact with people, objects and the world around us. In essence, the same neural systems that are

responsible for representing information during perception, action, and introspection are also responsible for representing (or *simulating*) the same information during conceptual thought (e.g., Barsalou, 1999, 2008; Glenberg and Robertson, 2000; Wilson, 2002; Gallese and Lakoff, 2005; Gibbs, 2006). A concept is an aggregated memory of aspects of experience that have repeatedly received attention in the past, and incorporates perceptual, motor, affective, introspective, social, linguistic, and other information. For instance, a concept of *dog* could potentially include a host of perceptual-motor information, possibly including visual information of the color and shape of a dog, tactile information regarding the feel of a dog's coat, olfactory information of the smell of a dog, auditory information of a dog's bark, motor information about patting a dog, social information about the status of dogs in human households, along with positive or negative affective valence depending on one's experience with dogs in the past. Any time the word "dog" is encountered, a subset of these aspects will be retrieved to suit the task at hand. However, human language is full of statistical regularities. Words and phrases tend to occur repeatedly in similar contexts, just as their referents tend to occur repeatedly in similar situations, which allows people to build up substantial distributional knowledge of linguistic associations. In this way, lexical associates of "dog" are also activated (e.g., "bark," "pet," "cat," etc.), which might in themselves suffice for a response and which in turn can activate their own simulation information. Importantly, the concept retrieved is situated and context-specific, with linguistic and simulation content changing dynamically with our experiences, current goals, and available resources. One cannot, in effect, retrieve the same concept twice.

Thus, both linguistic and simulation systems are central to human conceptualization (Clark, 2006; Barsalou et al., 2008; Louwerse and Jeuniaux, 2010); language bootstraps simulations to facilitate more complex conceptual processing than would otherwise be possible. The Language and Situated Simulation theory (LASS: Barsalou et al., 2008; see also the Symbol Interdependency Hypothesis, Louwerse and Jeuniaux, 2008), for instance, describes a general framework where both linguistic and simulation systems are simultaneously activated on encountering a word, with the linguistic system reaching peak activation slightly sooner than the simulation system. While it is a statistical trend that shallow, linguistic distributional responses are faster than responses that rely on deeper, situated simulation, the relative importance of each type of system will change according to the current context or specific task demands. In short, the concept to which a word refers is ultimately grounded in the simulation system, but a word does not need to be fully grounded every time it is processed (Louwerse and Connell, in press). It is important to note that distributional information in the linguistic system arises not only from associations between lexical items (e.g., between "dog" and "cat"), but also from associations between their referents in past experience (e.g., encountering cats and dogs in household pet situations). This constant interactivity between the linguistic and simulation systems means that they are, to some extent, partial reflections of each other. However, the linguistic system offers a fuzzy approximation that can provide an adequate heuristic in certain tasks, whereas the simulation system provides representational precision for more complex conceptual processing.

Affordances are key to the simulation system, and refer to the ways in which a particular object enables interaction (or *meshing*) with other entities (Gibson, 1979; Glenberg, 1997). A sweater affords filling with leaves in a way that a chair does not, and a leaf-filled sweater affords use as a pillow in a way that a rock-filled sweater does not (Glenberg and Robertson, 2000). In this example, the affordances of leaves, sweaters, and pillows mesh successfully within the situation of a person improvising a pillow on a camping trip. When affordances mesh successfully, they form a coherent and stable simulation, which is what allows conceptual processing to be both productive and creative.

EVIDENCE FOR LINGUISTIC DISTRIBUTIONAL INFORMATION IN CONCEPTUAL PROCESSING

The linguistic system contains statistical distributional information that is powerful enough to support superficial strategies in a broad range of linguistic and conceptual tasks that might otherwise be assumed to require deeper processing (Glaser, 1992; Solomon and Barsalou, 2004). For example, Solomon and Barsalou (2004) showed that responses in property verification tasks, where participants judge whether a property is usually true of an object (e.g., *lemon-yellow*) are predominantly based on word associations (e.g., "lemon" and "yellow" are closely associated, therefore respond yes) rather than on conceptual access. Indeed, this shallow, linguistic shortcut is the norm unless special care is taken to include filler items that are closely associated but nonetheless false (e.g., *monkey-banana*), which forces people to simulate the entity in question in order to avoid associative errors. In terms of conceptual combination, knowledge of how words have previously combined affects how people understand and evaluate future word combinations. Lapata et al. (1999) showed that the co-occurrence frequencies of adjective-noun combinations (e.g., *strong tea* versus *powerful tea*) were highly correlated with human plausibility ratings of those combinations, while the frequency of the noun alone was not. Indeed, the influence of the linguistic system is not limited to language stimuli. When participants were presented with two images in vertical alignment (e.g., a lamp above a table) and asked to judge whether the items usually appeared in those relative positions in the real world, Louwerse and Jeuniaux (2010) found that word order was a significant predictor of response times. Even though lamps are usually found above tables (and seldom below), people's ability to perform this ostensibly visuospatial memory task was affected by the fact that "table … lamp" is a more common linguistic construction than "lamp … table".

Regarding the timecourse predictions of both LASS (Barsalou et al., 2008) and the SIH (Louwerse and Jeuniaux, 2008), recent evidence supports the notion of the linguistic system offering a fuzzy heuristic that operates faster than the more precise simulation system. Louwerse and Connell (in press) analyzed the corpus distributions of a large set of perceptual object properties and found that, while human ratings are distinct for five perceptual modalities (i.e., auditory, gustatory, haptic, olfactory, visual: Lynott and Connell, 2009), distributional statistics identified only three "linguistic modalities" (i.e., auditory, visuo-haptic, and olfacto-gustatory). Previous work had shown that that switching between perceptual modalities in consecutive trials incurs a processing cost (e.g., Pecher et al., 2003). In a modality switching paradigm that

asked people to verify modality-specific properties (e.g., haptic *marble can be cool*), Louwerse and Connell tested whether switching costs were better predicted by switches between three distributional linguistic modalities or five simulated perceptual modalities. Consistent with LASS and SIH predictions, fast responses showed an effect of linguistic switching, while slow responses showed an effect of simulated perceptual switching. In other words, not only do these findings demonstrate distinct roles for the linguistic and simulation systems, but also their relative impact in the timecourse of responses.

EVIDENCE FOR SITUATED SIMULATED INFORMATION IN CONCEPTUAL PROCESSING

However, linguistic distributional information has limits. Previous experience with language will not suffice when trying to judge whether a description of a novel situation is sensible. People's capacity to understand how a sweater filled with leaves can be used as a pillow is rooted in their ability to simulate the objects' affordances and mesh them into a coherent situation (Glenberg and Robertson, 2000; see also "Affordances and Meshing in the Simulation System"). Just as objects, people, ideas, and emotions are encountered as part of broader, situated experience, the representations that people create during conceptual processing are situated simulations. When reading about everyday objects, people simulate perceptual properties such as shape (Zwaan et al., 2002), color (Connell, 2007; Connell and Lynott, 2009), and spatial location (Estes et al., 2008). For example, Estes et al. showed that people were slower to respond to an X at the top of the screen after having read *cowboy hat* (as opposed to *cowboy boot*) because the simulation of a *cowboy hat* was occupying their attention in its typical, high location. In a property-listing task, Wu and Barsalou (2009) found that people listed visual features of novel adjective-noun combinations that were occluded for the canonical noun: for example, *roots* and *dirt* were rarely listed for *lawn*, but were frequently listed for *rolled-up lawn*. Crucially, Wu and Barsalou showed that the pattern of property listing was not due to shallow processing in the linguistic system, but came from the visual simulation of the conceptual combination. Neuroimaging studies have also demonstrated modality-specific perceptual simulation during conceptual processing (see Barsalou, 2008, for review). González et al. (2006), for instance, found that passively reading scent-related words increased activation in the piriform cortex, an area normally activated during olfaction. Furthermore, Goldberg et al. (2006) showed that, when people verified properties related to color, sound, touch, and taste, regions of the cortex normally associated with perceiving visual, auditory, haptic, and gustatory information were activated.

Indeed, the emergence of several perceptual phenomena during conceptual processing strongly suggests that the conceptual system has co-opted the perceptual system for the purposes of representation. One such phenomenon is the tactile disadvantage: in perception, people are generally slower to detect tactile stimuli (e.g., finger vibration) than visual (e.g., light flash) or auditory (e.g., noise burst) stimuli, even when they are told which modality to expect (Spence et al., 2001; Turatto et al., 2004). Connell and Lynott (2010) replicated this effect in conceptual processing by using a modality detection task, where participants were asked to judge whether a particular word corresponded to a particular target modality. They showed that people were slower and less accurate in responding to touch-related words (e.g., *warm*, *itchy*) than words related to vision, sound, taste, or smell. In both perceptual and conceptual processing, the tactile disadvantage reflects people's difficulty in focusing attention on the tactile modality.

Another key phenomenon is the modality switching effect. In perception, switching costs arise when attention must be reallocated from one modality-specific neural system to another in successive trials: processing an auditory beep following a visual light flash following incurs a cost compared to when the trials are in the same modality (Spence et al., 2001; Turatto et al., 2004). Similar modality switching costs emerge when verifying properties in successive conceptual trials (e.g., auditory *leaves can be rustling* following visual *apple can be shiny:* Pecher et al., 2003; but see also Louwerse and Connell, in press), or when verifying a property following a perceptual stimulus (e.g., auditory *leaves can be rustling* following a visual light flash*:* van Dantzig et al., 2008). Moreover, modality switching costs are not just restricted to the retrieval of familiar conceptual information, but also emerge during conceptual combination when a new conceptual entity is created. Connell and Lynott (in press) asked participants to interpret adjective–noun combinations that had been normed to produce interpretations that related strongly to one perceptual modality (e.g., interpretations for *jingling onion* were predominantly auditory). They found that people were slower to interpret novel compounds when they followed familiar compounds in a different modality (e.g., auditory *jingling onion* following visual *shiny penny*). Importantly, switching costs in this interpretation paradigm were not subject to the linguistic shortcut that Louwerse and Connell (in press) observed in a property verification paradigm. Rather, situating the simulation of a novel conceptual combination in one perceptual modality incurs a switching cost if attention has already been grabbed by another modality.

Situated simulations are not just perceptual, but also extend to motor, affective, and other representations. Glenberg and Kaschak (2002) found that, when people read a sentence such as "You handed Courtney the notebook," they were faster to make a movement *away* from their bodies (compared to *towards* their bodies), which is consistent with having simulated the situationally appropriate movement. Even the current physical situation of the body can influence conceptual processing. For example, participants who made their responses into a microphone were faster to process phrases about mouth-related actions (e.g., "to suck the sweet") than hand-related actions (e.g., "to unwrap the sweet"), even though there was no response time difference for participants who responded with a foot pedal (Scorolli and Borghi, 2007). Similar effects extend into the simulation of affective situations. People's speed in understanding sentences that describe sad or unpleasant situations (e.g., "You hold back your tears as you enter the funeral home") is facilitated when the mouth forms a pouting expression by holding a pen between the lips (Havas et al., 2007), but inhibited when botulinum toxin has been used to immobilize the frown muscles (Havas et al., 2010). Similarly, adopting a congruent bodily posture facilitated people's recall of various social situations such as opening a door for a visitor or applauding at a concert (Dijkstra

et al., 2007). Taken collectively, the above evidence indicates that conceptual processing routinely requires perceptual, motor, affective, and social situated simulations.

ECCo: EMBODIED CONCEPTUAL COMBINATION

Both linguistic and simulation information are central to conceptual representations, and are therefore also central to the processes of conceptual combination. When we refer to conceptual combination, we mean creating or understanding a new concept by actively combining two already-known concepts (e.g., *mushroom chair* as a chair shaped like a mushroom). Our goal in presenting an embodied theory of conceptual combination is to put forward a single framework that can accommodate all the above evidence regarding the roles of the linguistic and simulation systems alongside the plethora of findings that have accumulated over the years in the conceptual combination literature. ECCo is thus the first theory of conceptual combination to do so.

The mechanisms described in this paper encompass the processing of both lexicalized and novel compounds, since the end representation in both cases is still a situated simulation. However, actively constructing a meaning for a novel combination should be distinguished from simply retrieving already-known concepts labeled with a lexicalized phrase. For example, processing the compound *office chair* also relies on linguistic and simulation information, but it does not require an active conceptual combination process in order to be successfully simulated and understood because it has a strong, frequency-reinforced link between the phrasal unit and its simulation (i.e., it is easily retrievable). A novel compound, conversely, is missing this link and therefore requires other means to arrive at a simulation and interpretation (i.e., a combination process).

COMBINATION PROCESSES

All conceptual combinations are situated, meaning that they involve representing a broader setting as part of the simulation. When a compound is presented, both linguistic and simulation systems are rapidly engaged; activation begins to spread out from the words to other linguistic tokens, the neural mechanisms involved in direct experience begin to simulate perceptual, motor, affective, and other situated information, and the two systems continually feed into one another (i.e., words help to activate simulations, and simulations help to activate words). In ECCo, as in other embodied theories of language comprehension (Barsalou et al., 2008), peak activation of the linguistic system is usually reached before peak activation of the simulation system. It is important to note that this is a statistical trend only – fast, shallow responses tend to rely more on linguistic information, and slow, deep responses tend to rely more on simulation information (e.g., Louwerse and Connell, in press) – but this trend can still influence the conceptual combination process, depending on the task at hand.

Differential task demands

If a participant is simply asked whether a noun–noun compound is sensible (i.e., whether or not it makes sense: Gagné and Shoben, 1997; Estes, 2003a), then this is a relatively shallow judgment for which the linguistic system offers a quick and dirty shortcut. If a compound consists of two words that have no shared statistical, distributional history, then the linguistic system will offer an heuristic for rejecting the compound as non-sensical without any attempt at conceptual combination actually taking place. On the other hand, if a compound consists of two words that are very frequently juxtaposed, then the linguistic heuristic will lead to its acceptance as sensible. Of course, participants do not have to rely solely on this linguistic shortcut just because it exists – they may use the simulation system as a double-check on any compounds that seem linguistically sensible, or some individuals may even base every decision on whether the concepts can combine into a coherent simulation – but an easy shortcut is hard to refuse. Because the linguistic heuristic is faster and computationally cheaper than basing a judgment on the simulation system, and because there are no penalties within the sensibility judgment paradigm to prevent its use (e.g., Solomon and Barsalou, 2004), participants can safely exploit it.

Interpretation tasks contrast with sensibility judgments in requiring deeper processing in the simulation system (Tagalakis and Keane, 2006; Lynott and Connell, 2010). If a participant is asked to give an interpretation for a noun–noun compound, then there must be an attempt to actually combine the concepts before making a response. If the concept affordances cannot be meshed in a situated simulation, then the compound will not be interpretable, but a successful situated simulation can be described in words as an interpretation for the compound. However, it is still sometimes possible for a noun–noun compound to be given a definitional interpretation predominantly on the basis of shallow, linguistic information. For example, a typological definition of *cactus beetle* as "a type of beetle" does not necessarily require any deep processing. Or if someone is told that a *sun holiday* is a holiday in the sun, she should be able to define a *snow holiday* as a holiday in the snow, or a *desert holiday* as a holiday in the desert, without necessarily requiring the simulation system. Because the linguistic and simulation systems operate in overlapping waves, with only a statistical tendency for the linguistic system to be faster, such rapid definitions do not mean that the simulation system is not engaged at all. Rather, the definition can be triggered and the participant can respond just from linguistic information, but, even while speaking or pressing the response button, the situated simulation is still taking shape.

Affordances and meshing in the simulation system

Each concept in the compound has a myriad of potential affordances based on past experience, and many more can be created on the fly if the situation requires (Glenberg and Robertson, 2000). We use the term *affordances* in a broader sense than just the perceptual-motor properties proposed by Gibson (1979) and the action-enabling view proposed by Glenberg (1997; Glenberg and Robertson, 2000). Similar to Estes et al. (in press), we view affordances as embodying much of what is often described as relational information, by referring to the ways in which a concept offers opportunities for meshing with other concepts. When the head and modifier concepts[1] are paired in a compound, they mutually constrain the number and type of affordances that can be meshed (see Maguire et al., 2010).

[1]The constituent concepts in a combination (e.g., *cactus beetle*) have traditionally been referred to as the modifier concept (e.g., *cactus*) and the head concept (e.g., *beetle*); in English, the compound word order means that the modifier is assumed to come first and the head second. While we feel that this terminology is misleading – the "modifier" concept does not necessarily modify the head, and the "head" concept is not necessarily the primary focus of the combination – we have retained these terms for the sake of consistency with prior research.

Meshing describes the process of integrating the complementary and potentially interactive aspects of two or more concepts, and "underlies our ability to understand novel combinations" (Glenberg, 1997, p. 6). Both concrete and abstract concepts can mesh affordances. Because relatively abstract concepts are heavily reliant on simulating perceptual, social, introspective (Barsalou and Wiemer-Hastings, 2005; Wiemer-Hastings and Xu, 2005), and affective information (Kousta et al., 2010), they still afford meshing with objects, agents, and other entities that can cause changes in mental states. An *elephant complaint* (see "Choice of Process") allows the concepts to mesh in a situation where either the *complaint* itself is large and important, or where the *elephant* is the originator or subject of the complaint. Indeed, the affordances of relatively abstract concepts can be meshed in a variety of situations, including perceptual (e.g., *value sandwich* as a sandwich that is cheap or good value for money), social (e.g., *fame advantage* as the favorable position conferred by being well-known), affective (e.g., *stress season* as a particular time of year when people are extra-stressed), and so on. A successful mesh will result in a coherent and stable simulation, which is the goal of the conceptual combination process.

There are no hard and fast rules regarding whether a particular concept is suitable for meshing in a particular situation – it entirely depends on the other concept used. For example *tree snake* is unlikely to be a snake that eats trees, and so one could argue that, in fact, *tree* can never be meshed with an eating situation because trees just aren't eaten. However, this assumption would be inaccurate. A *tree termite* could easily be a termite that eats trees; suddenly, trees afford eating. In short, it is not the case that one can independently slot the head and modifier concepts into a particular role or frame. The affordances of the head or modifier concept are affected by the other concept in the combination.

Affordances can be meshed in one of two ways. Sometimes, the head and modifier concepts are meshed directly with each other even if this involves substantial destruction of one of the concepts. For example, *cactus* and *beetle* can destructively combine into a spiky beetle because *cactus* is reduced to its spikiness, and *beetle* affords having a variety of exoskeleton shapes for defense or camouflage (thus giving rise to a situated simulation where the beetle wards off predators with its sharp spikes, or uses its green and spiky casing to hide on the surface of a cactus, etc.). However, sometimes this meshing is non-destructive as it incorporates the head and modifier concepts in a situation that requires little adaptation. For example, *cactus* and *beetle* mesh easily with an eating situation because beetles must eat something and cacti are a plausible food for beetles (thus giving rise to a situated simulation where the beetle is sitting on a cactus and eating away, or is munching a piece of cactus flesh as pet food, etc.).

Destructive and non-destructive processing

All else being equal, it is quicker to leave two concepts intact than to engage in situationally appropriate destruction. However, it's rare that all else is equal between two possible interpretations, which means that non-destructive processing is often, but not necessarily, faster than destructive processing. The length of interpretation time depends on the associations from the linguistic representation (e.g., does this compound resemble any lexicalized compounds? what are the close associate words?), the interaction between the linguistic and simulation systems (e.g., has a similar simulation been created before for these kinds of concepts?), and the ease of creating the situated simulation itself (e.g., can the mutually constrained affordances mesh in a plausible situation?). Sometimes interpretation is easy, whether destructive or non-destructive, and sometimes effortful.

The main difference between destructive and non-destructive processes lies in how the affordances are constrained and meshed. The destructive process seeks to mesh the head and modifier concepts together even if it means substantially reducing one of them, while the non-destructive process seeks to mesh the head and modifier affordances in a situation that allows both concepts to remain relatively intact. Note that in both destructive and non-destructive processes, meshing the concept affordances is not solely the province of the simulation system. The linguistic system also helps to cue and create these affordances, and thus helps to determine which process is followed. Take the compound *whale seal*: immediately on encountering the word "whale," closely associated linguistic tokens will be activated, including "fish," "big," "ocean," and so on. Such tokens will, in their turn, begin to activate their relevant representations in the simulation system, such as "big" rapidly and automatically drawing attention to the visual and haptic modalities (Lynott and Connell, 2009; Connell and Lynott, 2010). Furthermore, *whale seal* is analogous to the (for some people) lexicalized phrasal token *elephant seal*, and so the simulation of this existing species of large seal will also begin to be formed. This rolling wave of linguistic and simulation activations will help to reduce *whale* to its bigness and to cue the affordance of *seals* coming in a variety of sizes, and so lead to the common interpretation of *whale seal* as a type of large seal. There are, of course, many other interpretations possible for *whale seal* (e.g., a seal that hunts whales, a seal with black-and-white orca-like markings), but they will all follow a similar course of affordance cuing and meshing.

Choice of process

Critically, destructive and non-destructive interpretations do not compete in parallel within the mind of any one individual. It is cognitively wasteful to pursue destructive and non-destructive processes concurrently, and, while possible, it is not the norm. Rather, even though both processes may be open to pursuit at the start of interpretation, one process is preferentially enabled by a number of interactive factors (depending on, e.g., frequency of encountering similar compounds, previous attentional focus on aspects of the concept, experience with a plausible non-destructive situation, available perceptual resources for representing the destructive form of a concept). An individual thus commits quickly to either a destructive or non-destructive interpretation, and attempts to create a coherent simulation using that process.

Take the concept *elephant*: usually, when people refer to elephants they mean the holistic animal. Additionally, people have plenty of experience of the word being used to refer to something large and ungainly, both in isolation (e.g., "he's such an elephant", meaning he is of large build and/or is clumsy in his movements) and in lexicalized compounds (e.g., *elephant seal* and *elephant*

garlic both emphasize larger than normal size). Thus, experience has built up a link between the linguistic system's "elephant" token and a simulation of largeness. Furthermore, because elephants tend to be larger than any surrounding creatures in most situations in which they are encountered, people have plenty of experience of their attention being drawn to the elephant's large size, meaning that the simulation system is also likely to emphasize largeness in the visual and haptic components of the *elephant* simulation. Therefore, when one encounters the compound *elephant complaint*, one can either commit to keeping the elephant in its intact form or to using a reduced version.

For some people, the holistic form of *elephant* is highlighted (due to recency and priming effects as well as cumulative experience) and so they will attempt a non-destructive combination. Here, because *elephant* and *complaint* can mesh together in a situation where the elephant constitutes the reason for the complaint, the non-destructive interpretation of *elephant complaint* could be a complaint that people make about the behavior of an elephant at a zoo. For others, a reduced form of *elephant* is highlighted by its presentation in the compound and so they will attempt a destructive interpretation. Here, the elephant's largeness can mesh with *complaint* in a situation where size is equated to seriousness, giving the destructive interpretation of *elephant complaint* as a large and important complaint. Of course, other destructive interpretations are possible if some other reduced form of *elephant* is highlighted for a particular person (e.g., a long-living complaint that is never resolved), because considerable individual differences exist in linguistic and simulation experience.

DESTRUCTIVE INTERPRETATIONS

A destructive interpretation occurs when one, or both, constituent concepts are reduced during the interpretation process from their intact holistic forms to some situationally appropriate aspect(s) of the concepts. Sometimes one concept is reduced to a particularly salient or diagnostic aspect (e.g., the slowness of a *snail*, the black-and-white stripes of a *zebra*, the coldness of *icicle*), but, since both concepts mutually constrain each others' affordances, what appears salient or diagnostic for a concept in isolation may not apply to a concept in combination. For example, *icicle fingers* may be interpreted as freezing cold fingers: here, *icicle* is reduced to its coldness because it can mesh with the affordance of *fingers* to have a variety of temperatures. However, *icicle fingers* can also be interpreted as cold and stiff fingers, even though stiffness is not usually a salient or diagnostic aspect of *icicle* in isolation, because fingers also afford variations in flexibility according to temperature. We have experience of fingers being difficult to bend when they are particularly cold, and so reducing *icicle* to its coldness and stiffness allows the two concepts to mesh together. Indeed, such complementary affordances are part of situating the combination in our wider experience of cold and physical sensation.

Reversals

With destructive interpretations, the head concept usually remains intact while the modifier is destructively reduced, but this is only a general pattern rather than a golden rule. Nothing precludes the modifier staying intact while the head is reduced (or, indeed, both concepts being reduced: see *llama camel* in "Types of Interpretation"). For example, take *butter police* as referring to the dietary advisors who replace pats of butter in university canteens with low-saturate butter substitutes; here, *police* is reduced (to its function of enforcing regulations), while *butter* remains intact (because it is the thing being regulated). A *stone duck* (i.e., a statue), *toy duck* (i.e., a child's plaything), or *cloud duck* (i.e., a distinctively shaped cloud) all reduce *duck* to its general shape, and in the toy's case, maybe also its color: there is no longer an actual animal present in the simulation. Sometimes, reducing the head and leaving the modifier intact means that the focus of the interpretation is actually on the modifier concept. In the earlier example of *icicle fingers*, most of the interpretations kept the focus on the head (i.e., cold fingers are still fingers). However, other interpretations of *icicle fingers* could focus on *icicle*, such as "finger-shaped icicles forming outdoors in the cold." Such cases can be described as reversals, because the same interpretation could be produced from switching the order of the head and modifier (e.g., *finger icicles*). Similarly, *cloud duck* and *duck cloud* are both interpretable as a duck-shaped cloud. In this way, ECCo does not distinguish in principle between non-reversal and reversal destructive interpretations: both result from reducing one concept to certain situationally relevant aspects and meshing with the other concept. Whether an individual chooses to reduce the head or modifier will depend on their past experience with similar words, concepts, combinations, and situations.

Figurative combinations

Some compounds that could be described as having figurative interpretations actually use existing meanings of polysemous words. For example, *tiger executive* could refer to an executive who is fierce or ruthless in business dealings, but this interpretation makes use of the fact that the word "tiger" already has a standard figurative meaning of fierce or ruthless (e.g., "used to refer to someone fierce, determined or ambitious": New Oxford American Dictionary, 2009). In this case, the destructive interpretation process is assisted by the rapid retrieval and simulation of fierceness from the *tiger* modifier, which can mesh with the head concept as a trait of the executive in question. Similarly, *taste explosion* exploits the standardized use of *explosion* to refer to suddenness in sensory experience (e.g., "a sudden outburst of something such as noise, light, or violent emotion": New Oxford American Dictionary, 2009), which can easily mesh with *taste* in the simulation of a sudden burst of flavor in the mouth. In other words, such combinations are destructive interpretations that are greatly assisted by previous experience of a concept's usage in a reduced form.

However, many other combinations that could be described as figurative are more novel in their juxtapositions. Such combinations tend to be destructive, with one or both concepts being reduced in the situated simulation to an adapted form of some situationally appropriate aspect. For example, the compound *dragon soup* can be interpreted as a hot and spicy soup with chili, which involves reducing *dragon* to its hot, fire-breathing aspect. The synesthetic conversion of *hot* from the sense of high temperature to that of chili spiciness is facilitated by the fact that the word "hot" is polysemous, with a conventional meaning that refers to the gustatory heat of

chilies (see also Lynott and Connell, 2009). Hot, spicy chili is thus not a directly reduced aspect of *dragon*, but is rather an aspect that has adapted from tactile heat to gustatory heat, assisted by the linguistic system that allows the simulation of spicy taste from the word "hot," which then affords meshing with *soup*. In other words, many ostensibly figurative interpretations are destructive interpretations that are greatly assisted by previous experience of how a concept's associates may be simulated in more than one form.

NON-DESTRUCTIVE INTERPRETATIONS

A non-destructive interpretation occurs where the constituent concepts remain relatively intact in a shared situation. Both concepts, in their holistic forms, have mutually constrained affordances that mesh together in a situated simulation. An *octopus apartment*, for example, could be an apartment where an octopus lives: *octopus* affords having a place to live, and *apartment* affords providing a home, and so the two concepts mesh in a living arrangements situation. As with destructive interpretations, participants frequently specify details of how they have situated their simulation when they give interpretations. Many of our participants (Lynott and Connell, 2010) situated their simulations in ways that explain why an octopus might be living in an apartment, such as "an apartment for an octopus in an octopus sanctuary," "an apartment that has a pet octopus in it," or "an underwater apartment block for an octopus, like in Spongebob" (see "The Importance of Experience"). Although each of these participants situated their simulations slightly differently, they all succeeded in combining the concepts non-destructively as variants of the "place where an octopus lives" interpretation. Indeed, a conceptual combination can often be non-destructively interpreted in very different ways because the concepts have meshed in different situations. For example, a *kidnapper killer* could either be someone who kills kidnappers (because a *killer* must have a victim, and a *kidnapper* affords being the target of a killer for a variety of revenge or vigilante reasons), or a kidnapper who kills his or her victims (because both *kidnapper* and *killer* have victims, and afford merging the two crimes into the actions of one individual).

Reversals

In many non-destructive combinations, the order of the head and modifier concepts is not particularly important to the way in which the affordances mesh. The only difference is attentional focus. For example, an *octopus apartment* (as an apartment where an octopus lives) and an *apartment octopus* (as an octopus who lives in an apartment) essentially describe the same situation, with attention focused on different elements according to which concept is in the head position (although attentional focus can also be influenced by contextual and prosodic effects: e.g., Fernald and Mazzie, 1991). The level of detail is likely to differ according to attentional focus, so that a simulation of a *murder town* may situate extra details on the safety and desirability of the *town* itself, while a *town murder* may situate extra details on the nature or victim of the *murder*. ECCo therefore takes the same position with non-destructive interpretations as it does with destructive interpretations: it does not distinguish in principle between non-reversal and reversal interpretations because, in many cases, the simulation is essentially the same and only differs in attentional focus.

Representational potential

Sometimes, a non-destructive combination does not need to have both head and modifier concepts fully present in the simulation, but instead the situation allows one concept to exist *in potentia*. For example, *coffee cup* is a lexicalized compound, but it can be represented in more than one way: as a cup that contains coffee, or as a cup that can *potentially* contain coffee as its usual purpose. In the latter case, there is technically no coffee present. However, this absence does not mean that the interpretation is destructive. Rather, because the situated simulation involves a representational placeholder for an intact concept of coffee, it is a non-destructive interpretation in which the modifier concept exists in potential form. Similarly, *cactus beetle* (meaning a beetle who eats cacti) does not necessarily require a cactus to be present in the simulation, but the representational potential for an intact cactus concept means that the interpretation is non-destructive.

In terms of simulation content, the potential *coffee* or *cactus* in such interpretations is similar to the representation of a negated object. Kaup et al. (2006) found that, after reading a sentence such as "there was no eagle in the sky," people were faster to respond to a picture of a eagle with outstretched wings than if they had just read "there was no eagle in the nest." In other words, even though there was no eagle present in the described situation, the potential shape of the eagle was nevertheless simulated. Likewise, even though there may be no cactus present in the above interpretation of *cactus beetle*, the potential existence of cactus (as a source of food for the beetle) is still simulated.

THE IMPORTANCE OF EXPERIENCE

Any two randomly paired nouns can be processed as a conceptual combination, but past experience will make some compounds more likely than others to produce a plausible interpretation in most people. An interpretation will be plausible if it fits with prior experience and knowledge (Costello and Keane, 2000; Connell and Keane, 2004, 2006). It is important to note that plausibility does not depend solely on experience of the real, mundane world, but that experience of fictional words also counts. The consequence is that interpretations do not have to adhere to the conventional laws of gravity (e.g., *elephant bubble* as an elephant floating around in a bubble), genetic combination (e.g., *canary pear* as a cross between a canary and a pear), actuality (*dragon soup* as a soup made with dragon meat), or animacy (*chair complaint* as a chair complaining about being sat upon) in order to be plausible. In other words, prior experience counts whether vicarious and fictional or direct and physical. For example, more than one of our participants (Lynott and Connell, 2010) interpreted *octopus apartment* as an underwater apartment where an octopus lives, like in Spongebob. While there is no octopus character in the cartoon Spongebob Squarepants, there are, nonetheless, a number of other sea creatures, such as crab, squid, and starfish, who live in underwater houses and apartments. Thus, the cartoon world of Spongebob Squarepants provides a useful set of situational affordances into which *octopus* and *apartment* can plausibly fit.

If the simulation can mesh the head and modifier into a familiar situation (e.g., *horse house* as a stable), then this interpretation will be readily accepted. Even if the head or modifier concepts do not fit the situation exactly, past experience may still provide a useful

basis for interpretation because compounds are often interpreted by analogy with a more familiar compound (Lynott et al., 2004; Tagalakis and Keane, 2006). A *bullet car*, for instance, is similar to the lexicalized compound *bullet train*, and people interpret it similarly (i.e., as a fast car) rather than some other plausible interpretation (i.e., a car for transporting bullets). Such use of past experience with related combinations is fundamental to ECCo's account of the combination process. The fact that "car" and "train" are closely related linguistic tokens, and the fact that a simulation of *car* can be situationally adapted in many of the same ways as a simulation of *train*, means that familiarity with a *bullet train* makes a *bullet car* easier to interpret.

CONCEPTUAL COMBINATION IN DEVELOPMENT

Children are capable of both destructive and non-destructive conceptual combination from quite early stages of linguistic and conceptual development. By the age of three, most children can process a variety of non-destructive conceptual combinations. For example, Clark et al. (1985) found that, when asked to point to the *mouse hat*, most 3-year olds could reliably point to the relevant picture (i.e., a mouse wearing a hat) as opposed to distractor pictures of a mouse, a hat, or a fish wearing a hat. Performance for these non-destructive combinations was at ceiling by the age of four. Nonetheless, children of this age group are also capable of destructive conceptual combination. When asked to point to the picture of a *rabbit car*, 3-year olds preferred to point to a destructive interpretation (i.e., a car with rabbit ears and a fluffy tail) than a non-destructive alternative (i.e., a car beside a rabbit) or pictures of either object alone (Nicoladis, 2003). Four-year olds showed the same pattern, but were even more likely to choose the destructive interpretation. While it could be argued that the available non-destructive interpretations were in some way inferior or unlikely (e.g., a *sun bag* as a bag beside a multi-rayed cartoon sun), they nonetheless represented a valid means of distinguishing the compound subcategory from the head category (e.g., the bag beside the sun as opposed to the bag by itself); a pragmatic reason for conceptual combination (Downing, 1977; Clark et al., 1985; Wisniewski, 1997). Indeed, when children were given only two options to chose from – the destructive and non-destructive interpretations – they showed equal preference for both pictures (Gottfried, 1997). In other words, while children were capable of both destructive and non-destructive conceptual combination, they were not always sure which was the "correct" strategy for interpreting the compound.

Many of the difficulties experienced by children in understanding novel compounds are consistent with a preference to simulate two intact concepts, with younger children in particular having problems with combinations that require extensive concept destruction or representation of a concept *in potentia*. Regarding destructive conceptual combination, Gottfried (1997) found that certain types of destruction were harder than others for children to process. Compounds like *fish plate* or *butterfly mask*, where the modifier concept has been reduced to multiple visual features such as shape, texture, and color, posed few problems for children. In a picture-pointing task, 3-year olds could successfully identify the destructive interpretation (e.g., a mask decorated to look like a butterfly) as opposed to distractors (e.g., a butterfly, a mask, or a mailbox), and 5-year olds' performance was at adult level. In contrast, performance of both age groups was much worse for compounds that reduced the modifier concept to just a single visual feature, such as basic shape (e.g., *mitten leaf* as a leaf shaped like a mitten), or color/pattern (e.g., *zebra shells* as shells patterned with black-and-white stripes). Indeed, for these items, the 3-year-old children were close to chance in choosing the correct picture. These findings suggest that, by the age of three, children are willing and able to destructively interpret noun–noun compounds, but find it easier to do so when more of the modifier concept is left intact in the simulation.

Furthermore, there is some evidence that children aged three and younger may have difficulty with non-destructive interpretations where one concept exists *in potentia* (e.g., a *baby bottle* does not require that a baby is actually present in the simulation, but instead has the representational potential for the baby's existence). Krott et al. (2010) used novel words and objects in an attempt to control the amount of information children had about the concepts being combined. Children were introduced to a pair of novel objects in two different configurations, either showing possession/attachment (e.g., two objects that have been glued together, described as "a donka that has a kig") or function (e.g., one object is actively placed inside the other, described as "a donka that holds a kig"). When asked to point to the *kig donka*, children aged four and over behaved like adults in choosing each type of combination in approximately equal numbers. However, very young children (2- and 3-year olds) tended to prefer the simple combination where both objects were permanently joined to one another. In other words, the youngest children found it difficult to represent a *kig donka* as an object whose function is to hold a kig, because such a combination would require that the kig be represented *in potentia* (i.e., it is still a *kig donka* whether or not a kig is present). Thus, between the ages of two and four, children begin to lose their preference for representing two intact concepts and become capable of simulating the potential existence of one concept in a combination.

ACCOUNTING FOR CLASSICAL EFFECTS IN CONCEPTUAL COMBINATION

As well as offering a theoretical model that is based on the importance of grounded simulations in conceptual representation, the ECCo theory is also consistent with empirical findings from decades of classical conceptual combination research.

Property specificity

People represent the same color term differently depending on the object it describes. For example, Halff et al. (1976) found that people represented the color red differently when paired with *hair*, *wine*, *flag*, *brick*, and *blood*. Using a similarity-rating paradigm, they found that people rated the similarity of *red flag* and *red light* to be greater than the similarity between *red light* and *red wine*. Similarly, Medin and Shoben (1988) found that, when asked to compare the color gray with both black and white, people considered gray to be more similar to white in the context of *hair*, but more similar to black in the context of *clouds*. In terms of perceptual simulation, these results are unsurprising and largely inevitable. Since all of these words and combinations refer to known or lexicalized concepts no actual combination process is required to understand

them. Rather, a simple retrieval process will suffice. Because the redness of wine and the redness of a light are initially perceived as being different hues, they will be perceptually simulated as different hues. The same could be said of other object properties such as size: a *tall ladder* and a *tall man* are perceived differently (i.e., a tall ladder would be considerable taller than a tall man) and so their perceptual simulation will reflect these differences.

Typicality gradients

For any given category it is possible to list members of that category in descending order of their typicality. So, people judge members of the category "spoon" to be typically *small* and *metal.* However, people also readily agree that *large wooden spoons* are typical members of the "spoon" category, equal to *small metal spoons* (Medin and Shoben, 1988). As with the property specificity, this effect is not surprising when such retrieval is based on situated simulations of prior experience rather than the rearrangement or modification of correlated size and substance attributes within an amodal SPOON concept. From an embodied perspective, small metal spoons and large wooden spoons are used in very different situations with different accompanying objects (e.g., adding sugar to a teacup versus stirring ingredients in a mixing bowl), different grips (precision versus power), different motor actions (finger and wrist movement versus full arm and shoulder movement), and even different bodily postures (often sitting versus standing). Thus, in ECCo, because people simulate situationally appropriate information when they retrieve concepts, we should not be surprised that people are happy to accept both *large wooden spoons* and *small metal spoons* as representing typical *spoon* experiences.

Emergent properties

When people are asked to list features or properties of a combined concept they often list features that are not mentioned for the concepts in isolation. For example, *pet birds* are described as living in cages and able to talk, even though these features are not listed for *pet* or *bird* in general: such features have been described as emergent properties (Hampton, 1987). Since *pet bird* is a lexicalized compound, people will be able to form a situated simulation by retrieval (rather than by active conceptual combination), which is likely to contain situational information as to where the bird lives (in a cage or aviary) and what sounds it makes (learned words and phrases as well as squawks). Thus, these so-called emergent features do not materialize from the ether, but rather come from the situated nature of the simulation, based on each individual participant's own experience of pet birds (see also Barsalou, 1999).

To take a more novel example, a *helicopter blanket* is often said to be waterproof, even though neither helicopters nor blankets are generally described as such (Wilkenfeld and Ward, 2001). Here, although the combination *helicopter blanket* may not be directly retrievable due to its novelty, the same story applies. Because people create a situated simulation for any combination they fully interpret, its situation will often contain information that may not necessarily be present in more usual experiences of a concept (e.g., a *blanket* on a bed). By situating a *helicopter blanket* outdoors as part of the process of meshing the concepts into a type of cover for a helicopter, many of Wilkenfeld and Ward's participants included situationally-appropriate details to their simulation that suggest why a helicopter might need to be covered, such as the cover being camouflaged, waterproof, fireproof, or durable. In ECCo, such emergent features are an inevitable consequence of a situated simulation.

Relation frequency

Interpretations that use a high-frequency relation of the modifier (*mountain lake, mountain cabin, mountain stream* are all located in mountains) are understood more quickly than ones that use a low-frequency relation (*mountain magazine* is about mountains; Gagné and Shoben, 1997). In ECCo, strong links between the modifier word in the linguistic system and particular situations in the simulation system (e.g., between "mountain" and a situation where entities or events have a mountain location) would lead to that situation being a strong candidate for meshing head and modifier affordances. However, recent evidence suggests that relation frequency applies to classes of groups of concepts rather than individual lexical items (e.g., the category *geographical location* rather than the specific modifier *mountain*), and that combination times are influenced by relational frequencies of the head interacting with those of the modifier (Maguire et al., 2010). In ECCo, such a finding is consistent with the idea that not only do people situate the entire combination during interpretation, but also that the affordances of the constituents of the combination constrain the possible interactions between the concepts. So, for example, a *mountain rat* is quickly and easily interpreted as a rat who lives in the mountains because *mountains* afford providing habitats for animals (similar to *mountain goat*, *mountain dog*), and rats afford having habitats in a variety of geographic locations (similar to *desert rat*, *river rat*), and so the affordances mesh in a habitat situation. In contrast, *mountain carpet* will take longer to interpret as a carpet with a pattern of mountains because there are fewer similar compounds to help mutually constrain the affordances: people have little experience of meshing *mountains* with fabric designs, or meshing *carpet*s with geographical locations.

Context

A key issue in conceptual combination research has been whether the processes involved in interpreting novel combinations in and out of context are the same. Gerrig and Bortfeld (1999) showed that, out of context, the combination *doll smile* is more quickly interpreted than the combination *baseball smile*, but they are understood equally quickly in a supportive context. In ECCo, whether a combination is encountered in or out of context, it must be appropriately situated to be understood. The only difference a surrounding discourse context makes is to allow some or most of this situation to be already in place when a person encounters the novel combination. Obviously, prespecifying a complete situation can tightly constrain the possible interpretations that are situationally appropriate, and even suggest affordances for combinations that would ordinarily be difficult to situate. In Gerrig and Bortfeld's work, the discourse contexts clearly established a meaning for the novel combinations in advance. Therefore, when people encountered the novel combination, much of the hard work of simulating a situation is already done, thereby minimizing the differences between the compounds that had appeared out-of-context.

Similarity

A compound is more likely to be interpreted in a destructive manner when its constituent concepts are similar (Wisniewski, 1997; Wilkenfeld and Ward, 2001). In such cases, it is more difficult to find a situation in which reasonably similar head and modifier concepts can mesh complementary affordances. For example, take the compound *zebra clam,* which combines two animals: it is difficult to generate a plausible situation that would allow both *zebra* and *clam* to be kept relatively intact because their affordances do not lend themselves to mesh in a single situation. Instead, it is easier to allow *zebra* to be destructively reduced to its color and pattern in a situation where *clam* remains intact, because *clam* affords having a variety of markings and textures on its shell. It is likely that such statistical regularities (i.e., that concepts from the same broad class, such as "living things," tend toward destructive combination) are reflected in the linguistic system. Thus, in ECCo, encountering two similar tokens in a compound will lead to preferential activation of potential situations that involve destruction. Such situations may involve meshing one concept with the other on the basis of visual markings (e.g., *zebra clam* as a striped clam), size (*elephant carrot* as a huge carrot), thickness and texture (*coat shirt* as a thick, heavy shirt), motor function (*hammer shoe* as a shoe used to hammer in a nail), and many more.

In addition, the destructive combination process decreases perceived similarity between the constituent concepts. Estes (2003b) found that people believed concepts such as *zebra* and *clam* to be moderately similar when simply asked for their rating, but less similar when they had first interpreted *zebra clam* to be a striped clam. This finding is consistent with ECCo's account of destructive interpretations; because there was relatively little of the original *zebra* concept left in the simulation, participants judged it to be quite dissimilar to *clam*. Furthermore, non-destructive interpretations show the opposite pattern by increasing perceived similarity between constituent concepts. Estes also found that people tended to judge concepts such as *mountain* and *snake* as more similar if they first interpreted the compound (i.e., as a snake that lives in mountainous areas). Since similarity between concepts is enhanced when they are incorporated in the same scenario (Wisniewski and Bassok, 1999), Estes's participants rated the concepts as more similar because their simulation of the non-destructive interpretation left the constituent concepts intact.

Compositionality

Evidence is mixed regarding whether emergent properties of a compound (e.g., green for *unripe banana*) are represented faster (Springer and Murphy, 1992) or slower (Swinney et al., 2007) than properties that are true of the head concept but not the compound (e.g., yellow for *unripe banana*). However, these experiments predominantly used lexicalized compounds like *boiled celery* or *peeled apple*, which constitute concept retrieval rather than true combination. We know from other work in sentence processing that, when context implies an atypical representation of a concept (e.g., an unripe banana as opposed to a typically ripe one), both typical and atypical versions of the concept are simulated in parallel (Connell and Lynott, 2009). A similar mechanism could operate in the processing of lexicalized compounds, where *unripe banana* leads people to rapidly simulate a typical yellow banana alongside the specified green banana. Thus, pending new evidence of compositionality in the processing of novel compounds, ECCo remains equivocal on whether parallel simulations occur when two concepts are being meshed for the first time.

COMPARISON OF ECCo WITH PREVIOUS THEORIES OF CONCEPTUAL COMBINATION

ECCo dispenses with many of the assumptions and dichotomies that are traditional in much conceptual combination research. In this section, we will concentrate on five current accounts of conceptual combination (see **Table 1**). Competition among relations in nominals (CARIN: Gagné and Shoben, 1997) posits that people interpret novel combinations using a set of thematic relations (e.g., made of, located, used by), where processing time depends on how often a particular relation has been previously used with the modifier concept. Dual process theory (Wisniewski, 1997) holds that two different processes compete in parallel to generate different types of interpretation for a compound: property-based interpretations are constructed by applying a property of the modifier to the head concept, while relation-based interpretations are constructed by binding the concepts to thematic roles in an augmented schema. The interactive property attribution model (IPA: Estes and Glucksberg, 2000) allows for both property and relational interpretations, but specifies that both types arise from the interaction of candidate modifier features and relevant head dimensions. Constraint theory (Costello and Keane, 2000) asserts that people use three pragmatic constraints – diagnosticity, informativeness, and plausibility – in order to narrow down the wide range of possible interpretations to an optimal few. Lastly, the retrieval–composition-analysis model (RCA: Prinz, 2002) argues for three stages of combination: attempt to retrieve lexicalized meaning, compositional integration of concepts (with two parallel processes for property and relational interpretations, as in dual process theory), and analysis using background knowledge. While there are many potential issues for discussion, the rest of this section will focus on the key areas in which ECCo diverges from previous theories.

THEORETICAL DIFFERENCES

ECCo differs from previous theories in many major ways. **Table 1** summarizes the principal positions, with the most fundamental issues discussed below.

Nature of conceptual representation

ECCo describes how both the linguistic and simulation systems are central to conceptual representations. Of the existing theories of conceptual combination, some are agnostic as to the nature of the underlying representation (CARIN, constraint theory), while others take an explicitly amodal view of the conceptual system (dual process and IPA theories). Such views of the conceptual system lie in contrast to the embodied perspective that views conceptual representations as situated simulations. Both ECCo and the RCA model commit to the perceptual and motor basis of much of the conceptual system, although the RCA model describes concepts as frames or schemata that contain feature slots with particular values. In contrast, ECCo highlights the importance

Table 1 | Comparison of ECCo with existing theories of conceptual combination.

Theoretical position	CARIN (Gagné and Shoben, 1997; Gagné, 2000; Gagné and Spalding, 2004)	Dual process theory (Wisniewski, 1997; Wisniewski and Love, 1998; Wisniewski and Murphy, 2005)	Constraint theory (Costello and Keane, 2000)	IPA (Estes and Glucksberg, 2000)	RCA (Prinz, 2002)	ECCo
Nature of representations in conceptual combination	Not specified, although includes distributional knowledge of relation frequency	Amodal schemata with slots and fillers (but see Storms and Wisniewski, 2005)	Not specified, but modeled as amodal schemata with slots and fillers	Amodal schemata with slots and fillers	Sensorimotor-based schemata with slots and fillers	Linguistic distributional information and situated simulation of meshed affordances
Distinct types of interpretation	Relational	Relational, property, hybrid	Relational, property, hybrid	Relational, property	Relational, property, hybrid	Destructive, non-destructive
Modifier-head reversals normal	No	No	Yes	No	No	Yes
Perceptual information affects combination process	No	No	No	No	Yes (by assumption, given nature of representations)	Yes
Different interpretation types arise from	Same process	Distinct, parallel processes	Same process	Same process	Distinct, parallel processes	Early commitment to one of two related processes
Role of conceptual knowledge (apart from constituent concepts) in the combination process	Limited to final elaboration stage	Central to scenario construction for relational interpretations	Central to applying plausibility constraint	Not addressed	Central to final analysis stage	Central to creating simulation
Role of surrounding context in the combination process	Can increase relation availability	Can indicate relevant modifier feature for property interpretations	Not addressed	Not addressed	Can affect integration of dimensions in composition stage	Central to creating simulation
Emergent properties arise from	Not addressed	Elaboration with background knowledge	Additions from instances or abstract domains	Not addressed	Instance retrieval or elaboration with background knowledge	Situated nature of simulation
Consistent with developmental trajectory	No	No	No	No	No	Yes

of affective and social information as well as sensorimotor (particularly for more abstract concepts), and describes conceptual structure in terms of affordances that are meshed when situating a simulation (see "A" ffordances and Meshing in the Simulation System"). Furthermore, no theory but ECCo underscores the importance of distributional linguistic information in conceptual combination[2], and how it can predominate depending on task demands (see "Differential Task Demands").

[2]CARIN does incorporate a type of distributional information in the form of relation frequencies. However, the scope of this information is much narrower in CARIN than in ECCo as it does not consider other types of statistical linguistic information nor its interaction with the simulation system (see also "Relation Frequency").

Types of interpretation

Embodied conceptual combination describes interpretations as destructive or non-destructive, depending on whether the constituent concepts are reduced or left intact when their affordances are meshed. In contrast, existing theories of conceptual combination tend to categorize interpretations as relation-based (e.g., *cactus beetle* as a beetle that eats cacti) and property-based (e.g., *cactus beetle* as a spiky beetle), although CARIN disagrees that property-based interpretations constitute a distinct type. Dual process, constraint and RCA theories also include hybrids (e.g., *llama camel* as a cross between a llama and a camel, or a creature that is half-llama and half-camel) and/or conjunctives (e.g., *pet rhino* is both a pet and a rhino). However, the need for this fragmentation of interpretation

types is questionable, with many of these categories serving only descriptive roles as a legacy of previous research (e.g., Downing, 1977). ECCo's destructive and non-destructive interpretations subsume these categories, although their overlap is not isomorphic: while property-based interpretations are principally destructive, relational and hybrid interpretations conflate destructive and non-destructive combinations.

Most, if not all, property-based interpretations are destructive combinations in ECCo. The IPA model tends to focus on the transfer of a single property, although dual process, RCA and constraint theories are clear that multiple properties may be transferred. In ECCo, a destructive interpretation involves the reduction of one or both concepts to situationally appropriate aspects that can be meshed with the other concept's affordances, which means that there is no default number of "properties" that may comprise a concept's reduced form. So, for example, a *zebra clam* may indeed reduce *zebra* to a visual black-and-white striped pattern, but *icicle fingers* reduces *icicle* to a haptic, motor and proprioceptive representation of coldness and stiffness (see "Destructive Interpretations").

Many relation-based interpretations qualify as non-destructive interpretations in ECCo. For example, CARIN specifies head-causes-modifier (e.g., *flu virus*), modifier-causes-head (e.g., *mall headache*), head-uses-modifier (e.g., *gas antiques*) and so on. However, one relation in CARIN's taxonomy is always destructive (e.g., head-resembles-modifier: *zebra clam*). Furthermore, the same relation can vary in whether the actual interpretation is destructive or non-destructive. For example, the head-has-modifier relation is destructive in *song book* (described as a book that "has" songs) because *song* has been reduced to a purely visual representation (i.e., the *song* in *song book* does not contain any auditory component, which is generally a core aspect of a song). In contrast, *picture book* (a book that "has" pictures) is non-destructive because the *picture*s in question are still intact entities in the pages of the *book*. Other inconsistent relations include head-made-of-modifier (e.g., destructive *stone lion* versus non-destructive *stone wall*) and head-is-modifier (e.g., destructive *horse toy* versus non-destructive *servant girl*). Because ECCo does not rely on a set of abstracted relations, focusing rather on situated simulations to derive meanings, interpretations can be more specific than is possible within a finite relational taxonomy.

Hybrid interpretations are also split between destructive and non-destructive interpretations in ECCo. For example, a *llama camel* may be destructively interpreted as a cross between a llama and a camel: here, the resulting creature is part-llama and part-camel, but neither *llama* nor *camel* is simulated in holistic form because their meshing involves representing an offspring that retains some aspects of both. On the other hand, *singer songwriter* and *pet fish* both have non-destructive interpretations because there is still an intact *singer* and *songwriter* (or *pet* and *fish*) in the simulation even though the concepts have been meshed into a single individual (see "Non-destructive Interpretations").

EVIDENTIAL DIFFERENCES

Embodied conceptual combination explicitly addresses many empirical phenomena in the conceptual combination literature (see "Accounting for Classical Effects in Conceptual Combination") that other theories have failed to address (see **Table 1**). For example, the IPA model does not specify any role for wider conceptual knowledge in the combination process, contrary to the other theories (which specify a limited role at some point during processing) and to ECCo (which regards wider conceptual information as an inevitable and important resource in situating the simulation). Similarly, neither the IPA model nor constraint theory have addressed how context affects conceptual combination, while other theories allow it to influence the availability of relations or properties, and ECCo regards wider context as playing a central role in how the simulation is situated (see "Context"). Emergent properties are not currently explained by either CARIN or the IPA model, while other theories account for their appearance via elaboration of the combined concept, and ECCo holds that they arise naturally from the situationally-appropriate details in the simulation (see "Emergent Properties"). Indeed, ECCo is the only theory that is consistent with children's developmental trajectory in first preferring to simulate two intact concepts to later becoming capable of simulating potential and much-reduced concepts in combination (see "Conceptual Combination in Development").

ECCo and the RCA model are the only accounts of conceptual combination that can accommodate the role of perceptual information in the combination processes. For example, Connell and Lynott (in press) showed that people are slower to simulate a novel conceptual combination (e.g., visual *shimmering tuna*) if their attention has already been engaged by a different perceptual modality in a previous trial (e.g., auditory *loud motorcycle*), and that this modality switching cost is not due to linguistic associations between words. Similarly, effects of visual occlusion (Wu and Barsalou, 2009) and the orienting of spatial attention (Estes et al., 2008) are only compatible with the ECCo and RCA frameworks.

However, only ECCo is compatible with evidence that different types of interpretation emerge from early commitment to a particular process. The RCA model adopts dual process theory's assumption that relation- and property-based processes compete in parallel in the mind of each individual, with the first process to be completed providing the interpretation. However, there is no positive evidence for this assumption, as much of the evidence cited in favor of parallel processes is consistent with ECCo's early commitment account. For example, relation-based interpretations are usually, but not necessarily, faster than property-based interpretations (Gagné, 2000; Estes, 2003a; Tagalakis and Keane, 2006); however this does not mean that relation-based processing tends to "win" a parallel race, but simply indicates that one process, from start to finish, is generally faster than the other (see "Destructive and Non-destructive Processing"). Also, the finding that property-based targets are slowed down by relational primes just as much as relation-based targets are slowed down by property primes (Estes, 2003a) does not mean that the processes compete with each other, but merely shows that the processes do not operate sequentially with property-based processing as a last resort.

Critical evidence for the early commitment account and against the parallel assumption comes from Lynott and Connell (2010), who showed that prosody affects the speed of property-based interpretations, but not relation-based interpretations. Dual emphasis (i.e., equal prosodic stress on both nouns in the compound)

led to faster response times for property-based interpretations (e.g., *octopus apartment* as "an apartment with eight rooms") than relation-based interpretations (e.g., *octopus apartment* as "an apartment where an octopus lives"). Crucially, the relative proportions of different interpretation types were unaffected by prosody, with relation-based interpretations remaining more frequent than property-based interpretations even under dual emphasis. Since the fastest interpretation type was not the most frequent, as would be expected in a parallel race between processes where the first-completed interpretation "wins," this finding is not consistent with the RCA assumption of parallel competition. Rather, it is consistent with the ECCo account that any one individual rapidly commits to either a destructive or non-destructive interpretation: those who committed to destructive processing were facilitated by dual emphasis, whereas those who committed to non-destructive processing were not.

CONCLUDING REMARKS

In ECCo, we have outlined a conceptual combination system based on the idea that people's conceptual representations are built through complex interactions between linguistic, perceptual, motor, affective, introspective, and social experience. Both linguistic and simulation systems are critical to the combination process, but the new concept is fundamentally a situated, simulated entity. So, for example, a *cactus beetle* is represented as a multimodal simulation that includes visual (e.g., the shiny appearance of a beetle) and haptic (e.g., the prickliness of the cactus) information, all situated in the broader location of a desert environment under a hot sun, and with (at least for some people) an element of creepy-crawly revulsion. While ECCo builds on the contributions of existing theoretical and empirical work in conceptual combination, as well as drawing on the wider literature on language processing and embodied cognition, it marks a clear departure from previous work on conceptual combination in terms of representation and processing. Importantly, from these proposals, specific predictions can be derived to test the claims of the theory in observable behavior. For example, the depth-of-processing differences that arise according to task demands, and the specified early commitment to destructive or non-destructive interpretation, provide clear avenues for further investigation.

It is not possible in the initial presentation of a theory to address every issue, make every possible comparison, or describe every piece of supporting evidence, but we aim to provide a framework that will enhance our understanding of conceptual combination. Future work in this area will endeavor to explore some issues in greater depth, such as the mechanisms by which concepts mutually constrain each other's affordances, the factors that enable children to develop their destructive combination skills, and the potential differences in brain localization between destructive and non-destructive interpretation processes.

ACKNOWLEDGMENTS

The order of authorship is arbitrary. This work was supported by the UK Economic and Social Research Council (Grant number RES-000-22-3248).

REFERENCES

Barsalou, L. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford: Oxford University Press), 245–283.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 129–163.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends Cogn. Sci.* 10, 370–374.

Clark, E. V., Gelman, S. A., and Lane, N. M. (1985). Compound nouns and category structure in young children. *Child Dev.* 56, 84–94.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition* 102, 476–485.

Connell, L., and Keane, M. T. (2004). What plausibly affects plausibility? Concept–coherence and distributional word–coherence as factors influencing plausibility judgements. *Mem. Cognit.* 32, 185–197.

Connell, L., and Keane, M. T. (2006). A model of plausibility. *Cogn. Sci.* 30, 95–120.

Connell, L., and Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychon. Bull. Rev.* 16, 573–577.

Connell, L., and Lynott, D. (2010). Look but don't touch: tactile disadvantage in processing modality-specific words. *Cognition* 115, 1–9.

Connell, L., and Lynott, D. (in press). Modality switching costs emerge in concept creation as well as retrieval. *Cogn. Sci.*

Costello, F. J., and Keane, M. T. (2000). Efficient creativity: constraint-guided conceptual combination. *Cogn. Sci.* 24, 299–349.

Dijkstra, K., Kaschak, M. P., and Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition* 102, 139–149.

Downing, P. (1977). On the creation and use of English compound nouns. *Language* 53, 810–842.

Estes, Z. (2003a). Attributive and relational processes in nominal combination. *J. Mem. Lang.* 48, 304–319.

Estes, Z. (2003b). A tale of two similarities: comparison and integration in conceptual combination. *Cogn. Sci.* 27, 911–921.

Estes, Z., and Glucksberg, S. (2000). Interactive property attribution in concept combination. *Mem. Cognit.* 28, 28–34.

Estes, Z., Golonka, S., and Jones, L. L. (in press). Thematic thinking: the apprehension and consequences of thematic relations. *Psychol. Learn. Motiv.*

Estes, Z., Verges, M., and Barsalou, L. W. (2008). Head up, foot down: object words orient attention to the object's typical location. *Psychol. Sci.* 19, 93–97.

Fernald, A., and Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Dev. Psychol.* 27, 209–221.

Gagné, C. L. (2000). Relation-based combinations versus property-based combinations: a test of the CARIN theory and dual-process theory of conceptual combination. *J. Mem. Lang.* 42, 365–389.

Gagné, C. L., and Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 71–87.

Gagné, C. L., and Spalding, T. L. (2004). Effect of discourse context and modifier relation frequency on conceptual combination. *J. Mem. Lang.* 50, 444–455.

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in reason and language. *Cogn. Neuropsychol.* 22, 455–479.

Gerrig, R., and Bortfeld, H. (1999). Sense creation in and out of discourse contexts. *J. Mem. Lang.* 41, 457–468.

Gibbs, R. W. (2006). *Embodiment and Cognitive Science*. New York: Cambridge University Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. New York: Houghton Mifflin.

Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105.

Glenberg, A. M. (1997). What is memory for? *Behav. Brain Sci.* 20, 1–55.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glenberg, A. M., and Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *J. Mem. Lang.* 43, 379–401.

Goldberg, R. F., Perfetti, C. A., and Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* 26, 4917–4921.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., and Ávila, C. (2006). Reading cinnamon activates

olfactory brain regions. *Neuroimage* 32, 906–912.
Gottfried, G. (1997). Using metaphors as modifiers: children's production of metaphoric compounds. *J. Child Lang.* 24, 567–601.
Halff, H. M., Ortony, A., and Anderson, R. C. (1976). A context-sensitive representation of word meanings. *Mem. Cognit.* 4, 378–383.
Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Mem. Cognit.* 15, 55–71.
Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., and Davidson, R. J. (2010). Cosmetic use of Botulinum Toxin-A affects processing of emotional language. *Psychol. Sci.* 21, 895–900.
Havas, D. A., Glenberg, A. M., and Rinck, M. (2007). Emotion simulation during language comprehension. *Psychon. Bull. Rev.* 14, 436–441.
Kaup, B., Lüdtke, J., and Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: is a door that is not open mentally closed? *J. Pragmat.* 38, 1033–1050.
Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (in press). The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.*
Krott, A., Gagné, C. L., and Nicoladis, E. (2010). Children's preference for HAS and LOCATED relations: a word learning bias for noun–noun compounds. *J. Child Lang.* 37, 373–394.
Lapata, M., McDonald, S., and Keller, F. (1999). "Determinants of adjective – noun plausibility," in *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, 30–36.
Louwerse, M. M., and Connell, L. (in press). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cogn. Sci.*
Louwerse, M. M., and Jeuniaux, P. (2008). "Language comprehension is both embodied and symbolic," in *Symbols, Embodiment, and Meaning*, eds M. de Vega, A. Glenberg, and A. C. Graesser (Oxford: Oxford University Press), 309–326.
Louwerse, M. M., and Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition* 114, 96–104.
Lynott, D., and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* 41, 558–564.
Lynott, D., and Connell, L. (2010). The effect of prosody on conceptual combination. *Cogn. Sci.* 34, 1107–1123.
Lynott, D., Tagalakis, G., and Keane, M. T. (2004). Conceptual combination with PUNC. *Artif. Intell. Rev.* 22, 247–267.
Maguire, P., Maguire., R., and Cater, A. (2010). The influence of interactional semantic patterns on the interpretation of noun–noun compounds. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 288–297.
Medin, D., and Shoben, E. (1988). Context and structure in conceptual combinations. *Cogn. Psychol.* 20, 158–190.
New Oxford American Dictionary. (2009). *New Oxford American Dictionary Application* (Version 2.1.2) [computer software]. Cupertino, CA: Apple Inc.
Nicoladis, E. (2003). What compound nouns mean to preschool children? *Brain Lang.* 84, 38–49.
Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychol. Sci.* 14, 119–124.
Prinz, J. J. (2002). *Furnishing The Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.
Scorolli, C., and Borghi, A. M. (2007). Sentence comprehension and action: effector specific modulation of the motor system. *Brain Res.* 1130, 119–124.
Solomon, K. O., and Barsalou, L. W. (2004). Perceptual simulation in property verification. *Mem. Cognit.* 32, 244–259.
Spence, C., Nicholls, M. E. R., and Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Percept. Psychophys.* 63, 330–336.
Springer, K., and Murphy, G. L. (1992). Feature availability in conceptual combination. *Psychol. Sci.* 3, 111–117.
Storms, G., and Wisniewski, E. J. (2005). Does the order of head noun and modifier explain response times in conceptual combination? *Mem. Cognit.* 33, 852–861.
Swinney, D., Love, T., Walenski, M., and Smith, E. E. (2007). Conceptual combination during sentence comprehension: evidence for compositional processes. *Psychol. Sci.* 18, 397–400.
Tagalakis, G., and Keane, M. T. (2006). Familiarity and relational preference in the understanding of noun–noun compounds. *Mem. Cognit.* 34, 1285–1297.
Turatto, M., Galfano, G., Bridgeman, B., and Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Exp. Brain Res.* 155, 301–310.
Van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.
Wiemer-Hastings, K., and Xu, X. (2005). Content differences for abstract and concrete concepts. *Cogn. Sci.* 29, 719–736.
Wilkenfeld, M. J., and Ward, T. B. (2001). Similarity and emergence in conceptual combination. *J. Mem. Lang.* 45, 21–38.
Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.
Wisniewski, E. J. (1997). When concepts combine. *Psychon. Bull. Rev.* 4, 167–183.
Wisniewski, E. J., and Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cogn. Psychol.* 39, 208–238.
Wisniewski, E. J., and Love, B. C. (1998). Relations versus properties in conceptual combination. *J. Mem. Lang.* 38, 177–202.
Wisniewski, E. J., and Murphy, G. L. (2005). Frequency of relation type as a determinant of conceptual combination: a reanalysis. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 169–174.
Wu, L., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol.*132, 173–189.
Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language comprehenders routinely represent the shapes of objects? *Psychol. Sci.* 13, 168–171.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2010; accepted: 09 November 2010; published online: 25 November 2010.
Citation: Lynott D and Connell L (2010) Embodied conceptual combination. Front. Psychology ***1****:212. doi: 10.3389/fpsyg.2010.00212*
This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.
Copyright © 2010 Lynott and Connell. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 06 May 2011
doi: 10.3389/fpsyg.2011.00084

How "social" is the social Simon effect?

Thomas Dolk[1]*, Bernhard Hommel[2], Lorenza S. Colzato[2], Simone Schütz-Bosbach[3], Wolfgang Prinz[1] and Roman Liepelt[1,4]

[1] *Department of Psychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2] *Cognitive Psychology Unit, Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands*
[3] *Independent Research Group "Body and Self," Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[4] *Department of Psychology, Junior Group "Neurocognition of Joint Action", Westfälische Wilhelms-University, Münster, Germany*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Klaus Kessler, University of Glasgow, UK
Andrew D. Wilson, University of Leeds, UK

***Correspondence:**
Thomas Dolk, Department of Psychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1A, 04103 Leipzig, Germany.
e-mail: dolk@cbs.mpg.de

In the standard Simon task, participants carry out spatially defined responses to non-spatial stimulus attributes. Responses are typically faster when stimulus location and response location correspond. This effect disappears when a participant responds to only one of the two stimuli and reappears when another person carries out the other response. This social Simon effect (SSE) has been considered as providing an index for action co-representation. Here, we investigated whether joint-action effects in a social Simon task involve mechanisms of action co-representation, as measured by the amount of incorporation of another person's action. We combined an auditory social Simon task with a manipulation of the sense of ownership of another person's hand (rubber hand illusion). If the SSE is established by action co-representation, then the incorporation of the other person's hand into one's own body representation should increase the SSE (synchronous > asynchronous stroking). However, we found the SSE to be smaller in the synchronous as compared to the asynchronous stroking condition (Experiment 1), suggesting that the SSE reflects the separation of spatial action events rather than the integration of the other person's action. This effect is independent of the active involvement (Experiment 2) and the presence of another person (Experiment 3). These findings suggest that the "social" Simon effect is not really social in nature but is established when an interaction partner produces events that serve as a spatial reference for one's own actions.

Keywords: joint action, social Simon, social cognition, rubber hand illusion

INTRODUCTION

Many activities we perform in daily life are carried out together with other people. But how do we mentally represent other people's actions and how does this affect our own behavior?

Recent research suggests that joint action can lead to the representation of one's own and other's actions. This "action co-representation" is thought to facilitate action prediction and coordination of one's own actions with those of others (Sebanz et al., 2006). Evidence for this view stems from the "social Simon task" developed by Sebanz et al. (2003). In the standard Simon task (Simon and Rudell, 1967; Simon, 1990), participants typically carry out spatially defined responses (e.g., left and right key presses) to non-spatial stimulus attributes (e.g., auditory pitch or visual color) that randomly appear on the left or right. For example, participants are required to press a right key whenever they perceive a high-pitched tone and a left key in response to a low-pitched tone. Although stimulus location is completely irrelevant in this task, responses are typically faster when they spatially correspond to the stimulus signaling them. That is, spatial stimulus–response compatibility facilitates task performance, a phenomenon that has come to be known as the Simon effect. Commonly, this effect disappears when a participant responds to only one of the two stimuli, rendering the task a "go–nogo task" (Hommel, 1996). However, if the same go–nogo task is shared between two participants so that each of them operates one of the two responses, a Simon effect is observed (Sebanz et al., 2003) – the "social Simon effect" (SSE).

According to the dimensional overlap model (Kornblum et al., 1990) the standard Simon effect can be explained by a match between the spatially irrelevant dimension of the stimulus and the relevant response dimension (Hommel et al., 2001). Accordingly, responses are assumed to be automatically activated if the stimulus spatially corresponds to the correct response and thus facilitate task performance, whereas a lack of correspondence between stimulus–response pairs leads to response competition.

It is fair to say that the mechanisms underlying the SSE are poorly understood. Some authors have claimed that, due to the fundamentally social nature of perception and action, people automatically co-represent other people's actions (Knoblich and Sebanz, 2006). However, a finding that does not seem to be completely in line with the idea of action co-representation being social, automatic, and mandatory is that the SSE is fully present in autistic participants (Sebanz et al., 2005), who can be assumed to have difficulties processing social information. According to Guagnanoa et al. (2010), the major role of the co-actor in the social Simon task might be to provide a spatial reference frame that allows coding of one's own action as left or right relative to the other person – just as one's own action alternatives provide a reference frame for relative response coding (Hommel, 1996). Guagnanoa et al. (2010) further claimed that this reference frame can only be used if the other person is located within a participant's peripersonal space. In line with a spatial reference explanation for the SSE, the authors were able to show that the SSE breaks down if the two co-actors

are seated outside of arm's reach. However, this approach does not easily explain why an individual's bad mood (Kuhbandner et al., 2010) or negative relationship with the co-actor (Hommel et al., 2009) eliminates the effect.

In the present study, we make a further attempt to clarify what the notion of action co-representation might mean, what it refers to, and in which sense it might account for the SSE. In essence, it may be possible to distinguish between three concepts of action co-representation, ranging from strong to weak. According to the first, *strong* concept, the SSE is assumed to be functionally similar to the effect obtained when one person is taking care of both responses (Sebanz et al., 2003). Following this line of reasoning, the SSE is due to the cognitive integration of the co-actor and his/her actions into the actor's body scheme. The second, *intermediate* concept, assumes that actors represent information about their co-actor and his/her actions without integrating it with representations of their own body and actions. This co-representation of the self and other provides a reference frame for the (e.g., spatial) coding of an individual's own actions relative to the other person and his/her actions (Guagnanoa et al., 2010). Thus, rather than incorporating the other person into the actor's body schema, the co-actor is represented as a social agent responsible for the alternative action separately from one's own body and action. According to the third, *weak* concept, the co-actor does not function as a social being but mainly by virtue of producing particular events (actions with perceivable effects), which serve as reference for coding one's own action.

Our experiments proceeded from testing the strongest to the weakest concept. In Experiment 1, we tested whether the SSE is affected by the perceived ownership of another person's hand as suggested by a strong conceptualization of action co-representation (Sebanz et al., 2003; Knoblich and Sebanz, 2006). A reliable paradigm to experimentally manipulate the sense of ownership of another person's hand is the rubber hand illusion (RHI; Botvinick and Cohen, 1998). Here, a rubber hand (or another person's hand) is stroked either synchronously or asynchronously. During synchronous stroking, the subject commonly feels the illusion that the seen rubber (or foreign) hand becomes a part of his/her own body.

We experimentally combined the RHI with an auditory social Simon task. In Experiments 2 and 3, we gradually de-socialized the task situation. In Experiment 2, we tested if we could find evidence of a SSE without the active involvement of the co-actor. In Experiment 3, we excluded the co-actor from the task setting altogether to test the weak concept of action co-representation.

EXPERIMENT 1

The aim of Experiment 1 was to investigate whether the SSE relies on or varies as a function of action co-representation induced by the RHI. Participants performed an auditory social Simon task while the perceived ownership of another person's hand (i.e., synchronous vs. asynchronous stroking) was manipulated.

The RHI is assumed to arise from a multimodal conflict between vision, touch, and proprioception (Ehrsson et al., 2004; Tsakiris and Haggard, 2005; Kammers et al., 2009). As vision usually dominates touch and proprioception (Constantini and Haggard, 2007), the RHI emerges as a consequence of synchronous but not asynchronous stroking. When stroking is synchronous, the sense of ownership is strong. As a result, the activity of the other hand should be more strongly attributed to one's own body and thus induce an integration of another person into one's own action representation. Conversely, in the asynchronous stroking condition, the other hand is more likely to be attributed to a different actor (Botvinick and Cohen, 1998) and thus, clearly separated from one's own action. This condition was hypothesized to work against the strong concept of action co-representation.

If the SSE relies on the cognitive integration of the co-actor and his/her actions into the actor's body schema (strong concept), synchronous stroking should create a more pronounced SSE compared to asynchronous stroking. However, if an actor tends to represent the co-actor as separate from him/herself and not integrate the other's actions into their own body schema (intermediate concept), synchronous stroking might actually lead to a smaller, rather than a larger SSE than asynchronous stroking does. This is because the asynchronous stroking might increase the saliency of the other person's hand and its actions, and thereby provide a stronger spatial reference for coding the actor's own action.

METHODS

Participants

Forty healthy undergraduate students (20 female; 20–25 years of age, mean age = 23.8) with no history of neurological or hearing problems participated in Experiment 1. Twenty served as actual participants (henceforth called actors) and 20 as co-actors (see **Figure 1**). The participants were all right-handed as assessed by the Edinburgh Inventory (Olfield, 1971), had normal or corrected-to-normal vision, were naive with regard to the hypothesis of the experiment and were paid €14 for participating.

Apparatus and stimuli

An auditory Simon task (go–nogo task) was used. In each trial, one of two sounds designed by van Steenbergen (2007) and chosen as go (sound A) and nogo (sound B) was presented via two loudspeakers separated by a distance of 1 m at approximately 60 dB to either the left or right side of both participants.

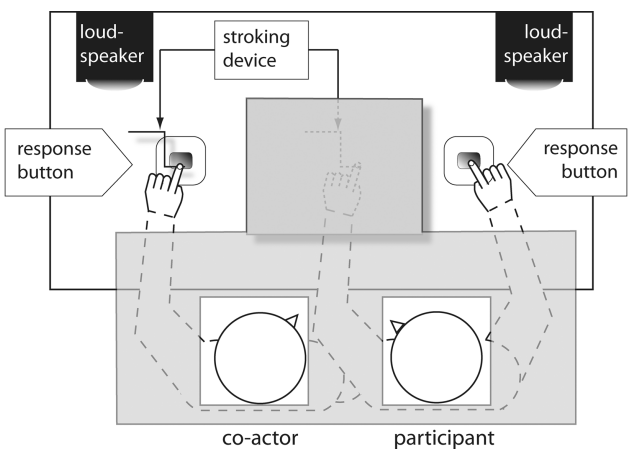


FIGURE 1 | Experimental setting in Experiment 1. Gray shaded areas indicate areas obscured from view.

To experimentally induce a sense of ownership of the other person's hand, we made use of the RHI. This involved stimulating the actor's and the co-actor's hand mechanically by means of two computer-controlled stepper motors, each with two identical paintbrushes attached, allowing the precise control of onset, direction, speed, and duration of both steppers independently. Following Lloyd (2007), the distance between both stroking devices was about 22.5 cm.

Subjective measures

Participants rated the perceived strength of the RHI by working through nine statements directly after each induction and experimental phase. The statements were translated from the original RHI Questionnaire (Botvinick and Cohen, 1998) and participants were to agree or disagree on a visual analog scale from left (0 = "completely disagree") to right (10 = "completely agree"). The first three statements are suggested to capture the core of the illusion (Botvinick and Cohen, 1998; Schütz-Bosbach et al., 2008; Kammers et al., 2009): (1) "It seemed as if I were feeling the touch of the paintbrush in the location where I saw the rubber hand touched"; (2) "It seemed as though the touch I felt was caused by the paintbrush touching the rubber hand"; (3) "I felt as if the rubber hand were my hand." A successful RHI induction would be indicated by higher ratings after synchronous than asynchronous visual–tactile stimulation.

Task and procedure

The experiment consisted of two consecutive sessions, each including an induction and an experimental phase. To avoid carryover effects, both sessions were separated by a 5-min mandatory break. Prior to the induction phase, participants were seated next to each other. The actual participant (see **Figure 1**) was always seated on the right and was asked to place his/her left index finger under the stroking device, so that the paintbrush could stimulate the occluded index finger from the knuckle to the fingertip or vice versa. His/her right index finger rested on the right response button. Randomly chosen co-actors, whose performance was not analyzed, were always seated on the left. They rested their left index finger on the left response button (80 cm between the two response buttons) directly under the left stroking device and their right hand on their lap under the table. After participant and co-actor were seated and had placed their hands in the correct positions, a white towel was placed over their shoulders and arms to obscure everything on the table except the co-actor's left and the participant's right hand (see **Figure 1**).

The experiment started with the induction phase. The stimulation was delivered mechanically by two stepper motors to which paintbrushes were attached. The amount of stimulation (onset, direction, speed, and duration) was precisely matched across conditions. To avoid habituation effects, the speed and direction of the paintbrushes were unpredictable and changed randomly every 5 s. In the synchronous condition, the participant's and the co-actor's left index fingers were stroked in synchrony, with identical location, timing, and trajectory parameters. In the asynchronous condition, the parameters differed between the two stroked fingers, while the total amount of stimulation for both index fingers was the same as in the synchronous condition. Thus, the synchronous and asynchronous stroking conditions differed only in the phase of the temporal structures of visual and tactile stimulation. The stroking procedure in each induction phase lasted for about 3-min. After the stimulation, both participant and co-actor were asked to fill out the RHI Questionnaire.

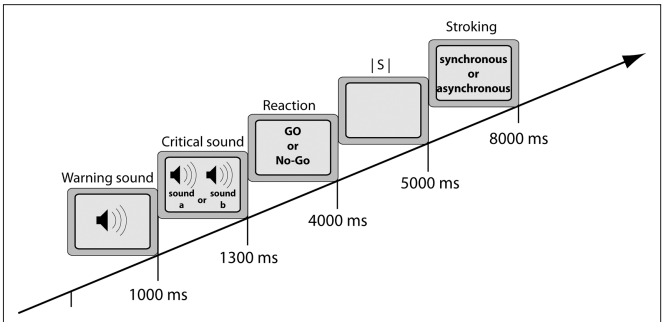


FIGURE 2 | Stimulus sequence in each trial. Trials started with the presentation of the warning sound. After 1000 ms, the critical sound (either A or B) appeared on the left or right of both participants. Participants had to respond within 3000 ms. The reaction was followed by an inter-stimulus-interval (ISI) of 1000 ms and 3000 ms stroking.

After completing the questionnaires, the experimental phase started. There were four blocks of 64 trials for each participant and co-actor (32 with spatially compatible stimulus–response relationships and 32 with spatially incompatible relationships). Each trial began with the presentation of the warning sound. After 1000 ms, the critical sound – either sound A or B – was presented to the right or the left side of both the participant and co-actor, requiring a response as quickly and as accurately as possible. Participant and co-actor were instructed to fixate on the other's hand and to respond exclusively to the sound assigned to them, irrespective of its location. Each response was followed by a 1000 ms inter-stimulus-interval (ISI) and 3000 ms stroking, which was always congruent to the stroking type of the corresponding induction period (either synchronous or asynchronous) to refresh the RHI (see **Figure 2**).

Feedback [mean reaction time (RT) and percentage correct] as well as a 2-min break were provided at the end of each block. After completing the first four blocks, participants were asked to fill in the RHI Questionnaire again, which was followed by a 5-min break to avoid carryover effects to the second session of the experiment. After the break, the second session started. The procedure was the same as in the first session except for the type of stimulation, which was always different from that in the first session. The order of stimulation type (synchronous followed by asynchronous stroking or vice versa) was counterbalanced across participants.

RESULTS

In the following, only data from the actual participants (actors) were analyzed.

Rubber hand questionnaire

Participants experienced the co-actor's hand as their own hand as a consequence of synchronous but not asynchronous stroking during both the induction and experimental phase: The RHI was significantly stronger after synchronous than after asynchronous

stroking (RHI-related questions 1–2 after the induction and 1–3 after the experimental phase; two-way paired-sample *t*-tests; all *p*s < 0.05).

Simon task

Reaction times. Responses were coded as compatible (stimulus ipsilateral to the correct response side) and incompatible (stimulus contralateral to the correct response side). Mean RTs on the auditory social go–nogo Simon task for the 20 actual participants were submitted to a 2 (Compatibility: compatible, incompatible) × 2 (Stroking: synchronous, asynchronous) within-subjects repeated measures analysis of variance (ANOVA). The analysis showed a significant main effect of Compatibility [$F(1,19) = 25.46$, $p < 0.001$, $\eta^2 = 0.57$] indicating that responses were faster with spatially compatible (mean RT = 291 ms) than with incompatible stimulus–response relationships (mean RT = 313 ms). More importantly, the compatibility effect varied with stroking, as indicated by a significant interaction of Compatibility × Stroking [$F(1,19) = 5.88$, $p < 0.05$, $\eta^2 = 0.24$; see **Figure 3**]. The 29 ms compatibility effect observed in the asynchronous stroking condition was significantly larger [$F(1,19) = 25.17$, $p < 0.001$, $\eta^2 = 0.57$] than the 15 ms compatibility effect in the synchronous stroking condition [$F(1,19) = 10.82$, $p < 0.01$, $\eta^2 = 0.36$; see **Figure 3**]. The main effect of Stroking was not significant [$F(1,19) < 1$, $\eta^2 = 0.01$]. To check for possible task order effects, we performed an additional ANOVA with Order as a between-subjects factor – but the three-way interaction was not significant [$F(1,18) = 1.40$, $p > 0.05$, $\eta^2 = 0.07$].

Error rates. We observed a significant main effect of Compatibility [$F(1,19) = 12.67$, $p < 0.01$, $\eta^2 = 0.40$], indicating higher error rates for incompatible (1.0%) than for compatible trials (0.3%). The interaction of Compatibility × Stroking was far from significance [$F(1,19) < 1$, $\eta^2 = 0.01$], which rules out a speed–accuracy trade-off.

DISCUSSION

The aim of this experiment was to test predictions of a strong concept of action co-representation accounting for the SSE. In particular, we investigated whether the SSE is mediated by the degree to which the active hand of a co-actor is perceived to be a part of the actor's own body.

First of all, we were able to replicate the findings of Sebanz et al. (2003), confirming that our particular setup was sufficiently sensitive to elicit the SSE in both the synchronous and the asynchronous stroking conditions. Second, we found the effect to be smaller, rather than larger, with synchronous than with asynchronous stroking. Thus, the incorporation of another person's hand into one's own body schema through the RHI (induced by synchronous stroking) reduces the SSE as compared to a condition where the co-actor is represented as a separate actor (induced by asynchronous stroking). This interpretation is supported by the subjective rating of the sense of ownership of the co-actor's hand in the synchronous stroking condition, which indicates that the experimental RHI manipulation was successful across all phases of the experiment.

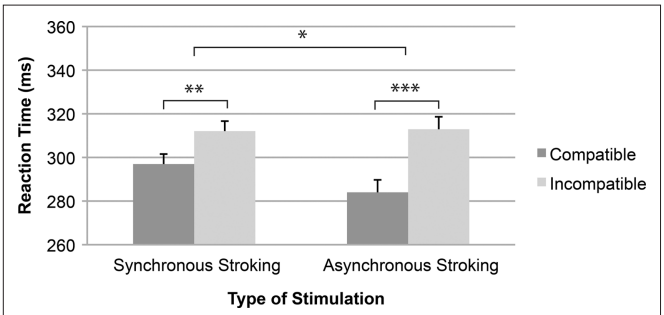


FIGURE 3 | Mean reaction time as a function of the type of stimulation and spatial stimulus–response compatibility. Error bars represent standard errors of the mean differences. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

These results provide considerable evidence against a strong concept of action co-representation as a mechanism underlying the SSE. That is, the SSE seems to occur even though actors represent their own action and the action of their co-actor separately. Emphasizing the difference between the two actions – or the related effectors – leads to a more pronounced SSE. This increase of the SSE in the asynchronous stroking condition is in line with the assumption that the SSE is established by the coding of one's action in reference to other actions (intermediate concept) or salient events (weak concept). Referential coding is known to be a basic principle operating in the Simon task (Hommel, 1993). Stimuli have been shown to be spatially coded relative to other stimuli that are either voluntarily attended to (Nicoletti and Umiltà, 1989) or that are salient enough to attract attention involuntarily (Treccani et al., 2006). With respect to action, response location has been shown to be coded in reference to other possible or recent responses (Hommel, 1996), in particular on spatial dimensions that help to discriminate between response alternatives (Ansorge and Wühr, 2004).

Given that most authors agree that the Simon effect is due to some sort of match or mismatch between spatial stimulus and response codes (Kornblum et al., 1990; Prinz, 1990; Hommel et al., 2001), the effect can only occur if stimulus location and response location are coded on the same dimension – as left and right in our case. In a standard Simon task, where the same participant performs both responses, this is very likely to happen, as the left–right dimension is particularly salient and provides the best discriminability between the two responses. In the social Simon task, however, participants operate only one response, so there is no actual need for spatial coding. Yet, if a co-actor (or perhaps another event) is sufficiently salient, people may nevertheless tend to code their response in reference to the spatial location of the other person or event (cf. Guagnanoa et al., 2010).

According to this reasoning, the social aspect of the joint-action situation created by the social Simon task may be just one of perhaps many factors that attract attention to other events and thereby induce the referential coding of one's own action, thus creating or enhancing the SSE. One implication of this possibility is that the active involvement of the co-actor in the present task might not necessarily induce referential response coding and elicit the SSE. To test this possibility, we performed a second experiment that included a now inactive but still salient "co-actor."

EXPERIMENT 2

The aim of Experiment 2 was to test whether the SSE can also be obtained with an inactive co-actor (to whom we will nevertheless keep referring to as "co-actor" for the sake of convenience). To do

so, we replicated Experiment 1 but now de-socialized the task to some degree: The co-actor no longer responded but sat passively next to the actual participant. If the co-actor provides a spatial reference frame for the coding of one's own action as left or right relative to the other person, one should expect a SSE even with an inactive co-actor. By contrast, however, if the active participation of the co-actor as a responding agent is crucial for the SSE to emerge as the original approach of Guagnanoa et al. (2010) suggests, the Simon effect should disappear.

METHODS

Participants

Twenty new healthy undergraduate students (10 female; 21–30 years of age, mean age = 24.8) with no history of neurological or hearing problems participated in Experiment 2. They fulfilled the same criteria and were treated in the same way as the participants in Experiment 1.

Apparatus, stimuli, task, and procedure

These were the same as in Experiment 1, with the following exceptions. The co-actor, who was the same for all participants, sat alongside the actual participant, and was no longer actively involved in the task. The left response button and the stroking device were visible on the co-actor's left side (see **Figure 4**). The co-actor was instructed to watch the behavior of the participant.

RESULTS

Reaction times

The 2 (Compatibility: compatible, incompatible) × 2 (Stroking: synchronous, asynchronous) within-subjects repeated measures ANOVA revealed a significant main effect for Compatibility [$F(1,19) = 14.05$, $p < 0.005$, $\eta^2 = 0.43$], showing that responses were faster with stimulus–response compatibility (mean RT = 335 ms) than with stimulus–response incompatibility (mean RT = 347 ms). The effects of Stroking [$F(1,19) < 1$, $\eta^2 = 0.03$] and the Compatibility × Stroking interaction [$F(1,19) < 1$, $\eta^2 = 0.01$; see **Figure 5**] were not significant. An additional ANOVA with Order as a between-subjects factor revealed no significant interaction [$F(1,18) = 1.66$, $p > 0.05$, $\eta^2 = 0.08$].

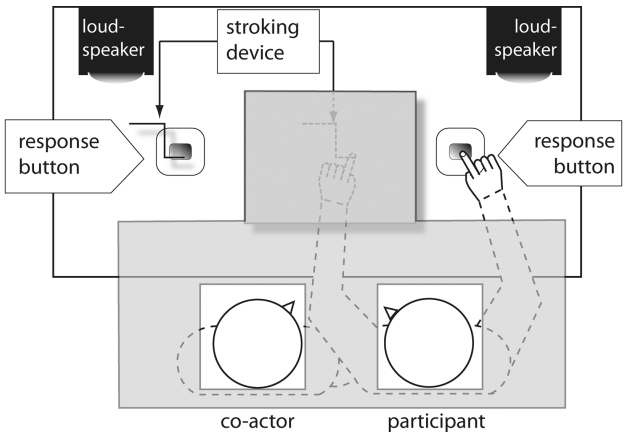


FIGURE 4 | Experimental setting in Experiment 2. Gray shaded areas indicate areas obscured from view.

Error rates

Neither the effects of Compatibility [$F(1,19) = 2.49$, $p > 0.05$, $\eta^2 = 0.12$] and Stroking [$F(1,19) < 1$, $\eta^2 = 0.01$], nor the interaction of Compatibility × Stroking were significant [$F(1,19) = 1.15$, $p > 0.05$, $\eta^2 = 0.06$].

DISCUSSION

Experiment 2 was designed to investigate whether the SSE can be obtained independent of the active involvement of another person. We found a significant overall Simon effect, which did not vary with the type of stroking. Extending previous findings (Guagnanoa et al., 2010), the present results suggest that the SSE can be established irrespective of what the other person is doing and whether this person is actively involved in the same or any other task.

This provides evidence against the interpretation of the SSE as a genuine joint-action effect (Sebanz et al., 2006) or as evidence for shared task representations (Knoblich and Sebanz, 2006; Ruys and Aarts, 2010). It also challenges the claim that it is "the presence of an active confederate" that provides the crucial reference for coding one's own action in space (Guagnanoa et al., 2010). This is not to say that the activity of the co-actor in Experiment 1 played no role at all. For one, the size of the SSE under asynchronous stroking was significantly smaller in Experiment 2 (13 ms) than it was in Experiment 1 [29 ms; $F(1,38) = 5.48$, $p < 0.05$, $\eta^2 = 0.13$] and, for another, the type of stroking affected the size of the SSE in Experiment 1 but not in Experiment 2. This suggests that the representation of one's own response is equipped with a spatial feature (right) if a second person is actively engaged in the task (Experiment 1) or if a non-social event, such as the movement of the stroking device, is present (Experiment 2). Thus, once another action alternative is sufficiently salient (e.g., the movement of the stroking device in the other response dimension), one's own action seems to be coded with reference to it, thus favoring the weak concept of action co-representation.

However, the finding of a Simon effect in Experiment 2 might alternatively be explained by the assumption that the mere presence of another person provided a reference for the coding of the alternative action event to one's own action (intermediate concept). In order to test this alternative explanation, we performed a third experiment.

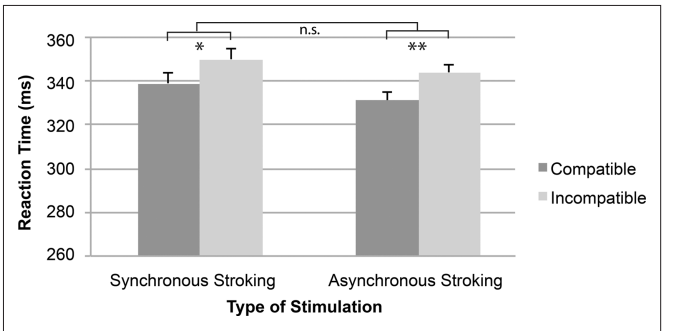


FIGURE 5 | Mean reaction time as a function of the type of stimulation and spatial stimulus–response compatibility. Error bars represent standard errors of the mean differences. $*p < 0.05$, $**p < 0.01$, n.s., not significant.

EXPERIMENT 3

The aim of Experiment 3 was to investigate the role of the presence of another person in the SSE. According to the weak concept of action co-representation, the co-actor does not function as a social actor but mainly as a source that produces particular events, which then serve as reference for coding one's own action. Guagnanoa et al. (2010) suggests that spatial response coding would be prompted by the presence of another active person in peripersonal space, irrespective of whether this person is working on the same task or not. Even though the present Experiment 2 suggests that this other person can just as well be inactive, the presence of a person may still be relevant. However, another possibility is that, even though the presence of another person is a particularly salient event, any salient event – social as well as non-social – could propagate referential coding (weak concept). With respect to our experimental setup, the mere manipulation of stroking might be sufficient to establish a salient non-social event that induces referential coding.

To test the latter hypothesis, we repeated Experiment 2 by further de-socializing the task: We no longer included another person. In one condition, the experimental procedure was the same as in Experiment 2, including the stroking device and the stroking manipulation (the "Device present" condition). This manipulation was expected to establish a salient event on the left side that could serve as a landmark for the participant to code his or her action as "right." If so, a SSE would be expected. In the other condition (the "Device absent" condition, see **Figure 6**), the stroking device on the participant's left was no longer present. However, the device above the participant's occluded left hand was still there, but there was no stroking manipulation any more, thus providing no landmark for referential coding.

METHODS

Participants

Twenty new healthy undergraduate students (11 female; 20–30 years of age, mean age = 24.9) with no history of neurological or hearing problems participated in Experiment 3. They fulfilled the same criteria and were treated in the same as the participants in Experiments 1 and 2.

Apparatus, stimuli, task, and procedure

These were the same as in the previous experiments, with the following exceptions. Participants carried out the task alone, in the absence of any other person, either with the moving stroking device ("Device present") or without the stroking device ("Device absent") operating at the alternative response side (see **Figure 6**).

The two conditions "Device present" and "Device absent" were manipulated within participants. In the "Device present" condition, the movements of the device stroking the participant's left hand and the visible device on the left side of the table were asynchronous. During the "Device absent" condition, both the response button and the stroking device on the participant's left were removed. However, to keep all other factors consistent with the previous experiments, the device above the participant's occluded left hand was still present, but no longer stimulated the hand (see **Figure 6**). The participants either started with the "Device present" or the "Device absent" condition; the order was counterbalanced across participants. Participants were seated on the right chair throughout the whole experiment and had to respond with their right index finger, pressing the right button. They were instructed to respond only to the tone assigned to them irrespective of the location and had to fixate on either the stroking device ("Device present") or a similar point on the empty table ("Device absent"; see **Figure 6**).

RESULTS

Reaction times

A 2 (Compatibility: compatible, incompatible) × 2 (Device: device present, device absent) within-subjects repeated measures ANOVA revealed a significant interaction of Compatibility × Device [$F(1,19) = 4.54$, $p < 0.05$, $\eta^2 = 0.19$; see **Figure 7**]. Follow-up analyses confirmed that the compatibility effect observed in the "Device present" condition (9 ms) was significant [$F(1,19) = 5.53$, $p < 0.05$, $\eta^2 = 0.23$], whereas the reversed compatibility effect in the "Device absent" condition (–7 ms, Compatible: mean RT = 314 ms; Incompatible: mean RT = 307 ms) was not [$F(1,19) < 1$, $\eta^2 = 0.05$; see **Figure 7**]. Participants responded faster with stimulus–response compatibility (mean RT = 304 ms) than with stimulus–response incompatibility (mean RT = 313 ms) in the "Device present" condition, whereas the RTs were slower for compatible

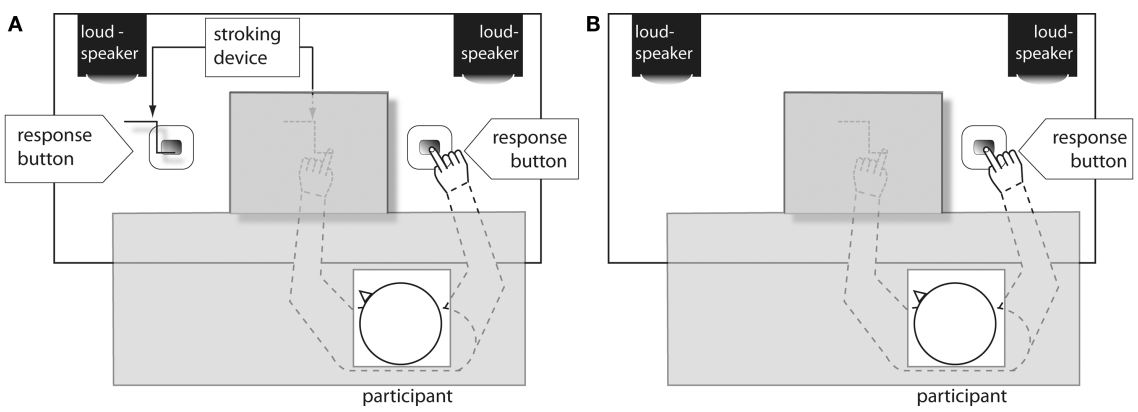


FIGURE 6 | Experimental setting in Experiment 3, in the "Device present" (A) and the "Device absent" condition (B). Gray shaded areas indicate areas obscured from view.

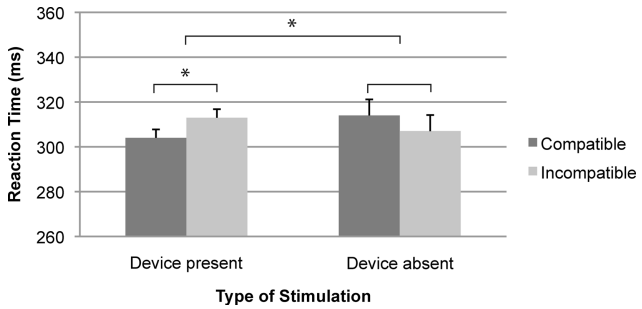


FIGURE 7 | Mean reaction time as a function of the type of stimulation and spatial stimulus–response compatibility. Error bars represent standard errors of the mean differences. *$p < 0.05$.

trials (mean RT = 314 ms) compared to incompatible trials (mean RT = 307 ms) in the "Device absent" condition. The effects of Compatibility and Device were not significant (all Fs < 1). An additional ANOVA with Order as a between-subjects factor revealed no interaction between Compatibility × Device × Order [$F(1,18) < 1$, $\eta^2 = 0.02$].

Error rates

Neither the effects of Compatibility [$F(1,19) = 1.72$, $p > 0.05$, $\eta^2 = 0.08$] and Device [$F(1,19) = 2.26$, $p > 0.05$, $\eta^2 = 0.11$], nor the interaction of Compatibility × Device were significant [$F(1,19) < 1$, $\eta^2 = 0.01$].

DISCUSSION

To disentangle whether it was the other person or the movement of the stroking device that acted as a reference frame for referential response coding, we repeated Experiment 2 without another person present. We found a significant Simon effect in the "Device present" condition but not in the "Device absent" condition, which suggests that the stroking device and/or its activities provided a reference for the spatial coding of the participant's own actions.

Given the rather small effects we obtained, these findings need to be handled with caution. Nevertheless, the interaction was clearly reliable and the sign of the SSE reversed in the "Device absent" condition. Moreover, the relatively modest size of the SSE in Experiment 3 fits well with the observations from other studies (e.g., Guagnanoa et al., 2010, Experiment 1 = 7 ms, Experiment 2b = 5 ms; Liepelt et al., 2011 = 9 ms; Sebanz et al., 2003 = 8 ms), suggesting that the present results are within a comparable range. Accordingly, we take these findings as evidence that the SSE can be obtained under entirely non-social circumstances.

To test whether the presence of the other person added to the effect in Experiment 2, we combined the data from Experiments 2 and 3 and performed an ANOVA with Compatibility (compatible, incompatible) as a within-subjects factor and Experiment (Experiment 2 – asynchronous stroking condition, Experiment 3 – "Device present" condition) as a between-subjects factor. This analysis revealed no significant interaction, suggesting that another person is neither necessary nor particularly relevant to provide a spatial reference for coding one's own action. Even a non-social salient event seems to have the potential to affect the way people represent their action and, thereby, to produce a (no longer so social) Simon effect.

GENERAL DISCUSSION

The present study was designed to clarify what the notion of action co-representation might be taken to mean and whether it is necessary for an account of the SSE. By experimentally inducing the RHI in an auditory social Simon task (in Experiment 1), we investigated whether the SSE is mediated by the degree to which actors incorporate other people's actions (or acting effectors) into their own body schema. Although we were able to replicate previous findings of Sebanz et al. (2003) and reproduced reliable RHIs in the actors, RHI-induced co-representation of a co-actor reduced, rather than increased, the SSE. This pattern of results provides clear evidence against a strong concept of action co-representation. Rather, asynchronous stroking apparently emphasized the existence of the co-actor, which was likely to provide a landmark for the spatial coding of the actor's response (Guagnanoa et al., 2010). This reasoning might be explained by the intermediate concept of action co-representation, which allows the separate representation of actor and co-actor (Hommel, 1993, 1996; Liepelt et al., 2011).

This interpretation received support from Experiment 2, where the other person was no longer actively involved in the task. In Experiment 3, however, the salient landmark was no longer a social event. In both cases, a reliable Simon effect was obtained even though – at least in Experiment 3 – the effect was clearly not a "social" effect anymore. In other words, the "SSE" can be induced through social as well as through non-social events, indicating that even an intermediate concept of action co-representation is not necessary to account for the SSE.

This assumption appears to be reasonable when considering the task requirements in the single and social Simon go–nogo task. In both cases, the spatial coding of responses is unnecessary and, as indicated by the absence of substantial Simon effects under some conditions (e.g., Hommel, 1996; Sebanz et al., 2003), apparently not obligatory either. This suggests that people are able to abstract from most aspects of their current environment and focus on what is currently relevant such as the action they are to carry out. However, irregular events are known to attract attention in a bottom-up fashion and it makes sense to assume that this is particularly true for events that are social in nature (Philipp and Prinz, 2010; Ruys and Aarts, 2010). This attraction of attention seems to be sufficient to induce the tendency or perhaps even the need to code one's action spatially in reference to this attention-attracting event.

In summary, testing predictions from three concepts of action co-representation revealed that neither the integration of another person's action into an individual's body representation (strong concept) nor the separate cognitive representation of one's own and the other person's action (intermediate concept) appear to be necessary for the SSE to occur. As even non-social events are sufficient to reliably influence an individual's own task performance (see Experiment 3), it seems to be the presence or expectation of salient events as such that underlies the SSE (Vlainic et al., 2010). Moreover, the present results suggests, that even the modality of these events does not matter much: Both proprioceptive events (resulting from the stroking of the participant's occluded hand) and visual events (resulting from the motion of the stroking device in the alternative response dimension) were functional in providing a reference frame for action coding. Identifying the factors that determine the relative saliency and the relative contributions of visual, proprioceptive, and other information to the SSE represents an important goal for future research.

The general observation that a salient event induces referential action coding is in line with the assumptions of the theory of event coding (Hommel et al., 2001; Hommel, 2004, 2009). The presence of another person sitting next to the actor performing his/her part of the task represents a salient event that provides an alternative for the actor's own response and thus calls for perceptual discrimination between the two (e.g., in terms of "left" vs. "right" and "me" vs. "not me"; Hommel et al., 2009; Liepelt et al., 2011). Once a response is spatially coded, it provides a reference that is compatible or incompatible with respect to the assigned stimulus (Guagnanoa et al., 2010). Establishing or using a task-relevant feature dimension may create feature overlap (Kornblum et al., 1990; Lam and Chua, 2009) and feature-correspondence effects, which are the basis of the Simon effect and other stimulus–response compatibility effects.

Our considerations should not be taken to mean that task sharing – in the sense of considering aspects of someone else's task in one's own cognitive task representation – is not social at all, nor does it imply that social aspects play no role at all in the SSE. As already mentioned, social factors like mood (Kuhbandner et al., 2010) or interpersonal relationship (Hommel et al., 2009) have been found to affect the size and presence of the effect. Nevertheless, these effects might well be mediated by the saliency they lend to the alternative action and/or actor, and the impact of this saliency on referential response coding. Further support for the claim that the understanding of social events is grounded in visual perception is given by the extensive literature on biological motion (Cross et al., 2009; Liepelt and Brass, 2010; for a review see Blake and Shiffrar, 2007) and emotions (Pollick et al., 2001; Heberlein et al., 2004). Accordingly, the SSE is another excellent example of the understanding of social events throughout salient signals we are sensitive to. Future research should take this into account to improve our understanding of social cognition.

Taken together, the present study suggests that the social Simon may be socially induced but is not really social in nature. Rather than requiring or necessarily reflecting the co-representation of the other person's action into an individual's own body and/or task representation, the effect seems to result from salient social or non-social actions or events that induce the coding of an individual's own action as left or right – a necessary condition for the Simon effect to emerge.

ACKNOWLEDGMENTS

We thank Henrik Grunert for technical support, Rosie Wallis for editing the article from a native speaker's perspective, Patricia Grocke for being the confederate in the second experiment and Anna Schmidt for her assistance with the data collection for the third experiment. The authors wish to thank Emily Cross and the reviewers for their helpful comments on a previous draft of this article.

REFERENCES

Ansorge, U., and Wühr, P. (2004). A response-discrimination account of the Simon effect. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 365–377.

Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73.

Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391, 756.

Constantini, M., and Haggard, P. (2007). The rubber hand illusion: sensitivity and reference frame for body ownership. *Conscious. Cogn.* 16, 229–240.

Cross, E. S., Hamilton, A. F. D. C., Kraemer, D. J. M, Kelley, W. M., and Grafton, S. T. (2009). Dissociable substrates for body motion and physical experience in the human action observation network. *Eur. J. Neurosci.* 30, 1383–1392.

Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). That's my hand. Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305, 875–877.

Guagnanoa, D., Rusconia, E., and Umiltà, C. A. (2010). Sharing a task or sharing space? On the effect of the confederate in action coding in a detection task. *Cognition* 114, 348–355.

Heberlein, A. S., Adolphs, R., Tranel, D., and Damasio, H. (2004). Cortical regions for judgements of emotions and personality traits from point-light walkers. *J. Cogn. Neurosci.* 16, 1143–1158.

Hommel, B. (1993). The role of attention for the Simon effect. *Psychol. Res.* 55, 208–222.

Hommel, B. (1996). S-R compatibility effects without response uncertainty. *Q. J. Exp. Psychol.* 49A, 546–571.

Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends Cogn. Sci.* 8, 494–500.

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychol. Res.* 73, 512–526.

Hommel, B., Colzato, L. S., and van den Wildenberg, W. P. M. (2009). How social are task representations. *Psychol. Sci.* 20, 794–798.

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878.

Kammers, M. P. M., de Vignemont, F., Verhagen, L., and Dijkerman, H. C. (2009). The rubber hand illusion in action. *Neuropsychologia* 47, 204–211.

Knoblich, G., and Sebanz, N. (2006). The social nature of perception and action. *Curr. Dir. Psychol. Sci.* 15, 99–104.

Kornblum, S., Hasbroucq, T., and Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus–response compatibility-A model and taxonomy. *Psychol. Rev.* 97, 253–270.

Kuhbandner, C., Pekrun, R., and Maier, M. A. (2010). The role of positive and negative affect in the "mirroring" of other persons' actions. *Cogn. Emot.* 24, 1182–1190.

Lam, M. Y., and Chua, R. (2009). Influence of stimulus–response assignment on the joint-action correspondence effect. *Psychol. Res.* 74, 476–480.

Liepelt, R., and Brass, M. (2010). Top-down modulation of motor priming by belief about animacy. *Exp. Psychol.* 57, 221–227.

Liepelt, R., Wenke, D., Fischer, R., and Prinz, W. (2011). Trial-to-trial sequential dependencies in a social and non-social Simon task. *Psychol. Res.* doi: 10.1007/s00426-010-0314–313.

Lloyd, D. M. (2007). Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand. *Brain Cogn.* 64, 104–109.

Nicoletti, R., and Umiltà, C. (1989). Splitting visual space with attention. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 164–169.

Olfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9, 97–113.

Philipp, A. M., and Prinz, W. (2010). Evidence for a role of the responding agent in the joint compatibility effect. *Q. J. Exp. Psychol.* 28, 1–13.

Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition* 82, B51–B61.

Prinz, W. (1990). "A common coding approach to perception and action," in *Relationships between Perception and Action: Current Approaches*, eds O. Neumann and W. Prinz (Berlin: Springer), 167–201.

Ruys, K. I., and Aarts, H. (2010). When competition merges people's behavior: interdependency activates shared action representations. *J. Exp. Soc. Psychol.* 46, 1130–1133.

Schütz-Bosbach, S., Avenanti, A. Aglioti, S. M., and Haggard, P. (2008). Don't do it! Cortical inhibition and self-attribution during action observation. *J. Cogn. Neurosci.* 21, 1215–1227.

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint actions: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76.

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21.

Sebanz, N., Knoblich, G., Stumpf, L., and Prinz, W. (2005). Far from action blind: representation of others' action in individuals with autism. *Cogn. Neuropsychol.* 22, 433–454.

Simon, J. R. (1990). "The effects of an irrelevant directional cue on human information processing," in *Stimulus–Response Compatibility: An Integrated Perspective. Advances in Psychology*, Vol. 65, eds R. W. Proctor and T. G. Reeve (Amsterdam: North-Holland), 31–86.

Simon, J. R., and Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *J. Appl. Psychol.* 51, 300–304.

Treccani, B., Umiltà, C., and Tagliabue, M. (2006). Simon effect with and without awareness of the accessory stimulus. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 268–286.

Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 80–91.

van Steenbergen, H. (2007). *Neural Integration of Actions and Their Auditory Effects: An fMRI-Study*. Unpublished Master Thesis, Leiden University.

Vlainic, E., Liepelt, R., Colzato, L. S., Prinz, W., and Hommel, B. (2010). The virtual co-actor: the social Simon effect does not rely on online feedback from the other. *Front. Psychol.* 1:208. doi: 10.3389/fpsyg.2010.00208

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 September 2010; accepted: 21 April 2011; published online: 06 May 2011. Citation: Dolk T, Hommel B, Colzato LS, Schütz-Bosbach S, Prinz W and Liepelt R (2011) How "social" is the social Simon effect? Front. Psychology **2**:*84. doi: 10.3389/fpsyg.2011.00084*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz and Liepelt. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 03 December 2010
doi: 10.3389/fpsyg.2010.00208

The virtual co-actor: the social Simon effect does not rely on online feedback from the other

Elize Vlainic[1], Roman Liepelt[2,3], Lorenza S. Colzato[1,4], Wolfgang Prinz[2] and Bernhard Hommel[1,4]*

[1] *Cognitive Psychology Unit, Leiden University, Leiden, Netherlands*
[2] *Department of Psychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[3] *Junior Group "Neurocognition of joint action", Department of Psychology, Westfälische Wilhelms-University, Münster, Germany*
[4] *Leiden Institute for Brain and Cognition, Leiden, Netherlands*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Kirsten I. Ruys, Utrecht University, Netherlands
Aron Barbey, National Institutes of Health, USA

***Correspondence:**
Bernhard Hommel, Cognitive Psychology Unit, Department of Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, Netherlands.
e-mail: hommel@fsw.leidenuniv.nl

The social Simon effect (SSE) occurs if two participants share a Simon task by making a Go/No-Go response to one of two stimulus features. If the two participants perform this version of the Simon task together, a Simon effect occurs (i.e., performance is better with spatial stimulus–response correspondence), but no effect is observed if participants perform the task separately. The SSE has been attributed to the automatic co-representation of the co-actor's actions, which suggests that it relies on online information about the other's actions. To test this implication, we investigated whether the SSE varies with the presence and amount of online action-related feedback from the other person. Experiment 1 replicated the SSE with auditory stimuli. Experiment 2, in which participants were blindfolded, demonstrated that visual feedback from the other's actions is not necessary for the SSE to occur. Experiment 3 replicated Experiment 2 with a regular and a soundless keyboard. A comparable SSE was obtained in both conditions, suggesting that even auditory online input from the other's actions is not necessary. Taken together, our data suggest that the SSE does not rely on online information about the co-actor's actions but that *a priori* offline information about another actor's presence is sufficient to generate the effect.

Keywords: Simon effect, social Simon effect, action representation, task representation, grounded cognition

INTRODUCTION

Humans are active agents who organize their behavior according to their plans and action goals. However, where those plans and goals come from and how they are acquired is not very well understood. According to the ideomotor approach to voluntary action (Lotze, 1852; James, 1890; for an overview, see Stock and Stock, 2004), actions are cognitively represented in terms of their sensory consequences, so that the acquisition of action plans amounts to the experience-driven integration of motor patterns with codes of their sensory effects (Elsner and Hommel, 2001). Indeed, numerous studies have provided evidence that performing a movement creates associations between the underlying motor pattern and the sensory consequences that go along with executing this pattern (for an overview, see Hommel, 2009). This implies that our cognitive action representations are grounded in sensory experience, that is, in the perceptual consequences a given action was experienced to create. According to ideomotor theory, this perceptual grounding provides us with the means to carry out movements intentionally: we internally re-create the sensory experience of the action effects to some degree (in other words, we anticipate them) and thereby reactivate the associated motor pattern that will then produce the anticipated effects in the external world (Elsner and Hommel, 2001; Hommel, 2009).

Recent research has raised the possibility that action representations do not only comprise of information about the sensory consequences of one's own action but that information about other people's actions might also be considered. Most of this research stems from the so-called "Social Simon paradigm," in which two persons share a Simon task (Sebanz et al., 2003). In the standard Simon task, single participants carry out left and right responses to a non-spatial attribute of stimuli that appear randomly on the left or right side. The standard finding in this task is that participants perform better if the stimulus happens to appear on the side of the correct response than if it does not (Simon and Rudell, 1967). Sebanz et al. (2003) had two participants share this task, so that each participant responded to only one of the stimuli by pressing a single key, which from the perspective of each participant rendered the task a Go/No-Go task. While performing this Go/No-Go version alone did not elicit a Simon effect, working on the task together with a co-actor did. This shared-task effect has been called the social Simon effect (SSE; Sebanz et al., 2003).

The SSE suggests that action or task representations are grounded not only in the experience of one's own actions but that they can also include aspects of the current social or at least situational context (Hommel et al., 2009), which seems to imply that action planning is truly situated (Clancey, 1997). Given that we are social animals used to act in social context, which often requires the consideration of other people's activities, this may not come as a surprise. However, the cognitive mechanisms responsible for integrating information about the current action context are not very well understood. According to Sebanz et al. (2003), the SSE might suggest that people do not only create cognitive representations of their own actions but they may also automatically co-represent the actions of a co-actor. In particular, Sebanz et al. (2003) suggest that

the effect may arise at a representational level that does not distinguish between one's own and another person's actions. According to the ideomotor principle (Hommel, 2009), both types of actions are cognitively represented in terms of their sensory consequences, which might imply that sensory feedback from both one's own and the co-actor's actions is crucial for the SSE to occur. Alternatively, it might be that it is not the other person's action that matters the mere possibility of acting might suffice. If so, an actor should show a SSE even if he or she is unable to perceive the co-actors action and continuously monitor his or her presence.

In an auditory version of the Simon task, Ruys and Aarts (2010) provided actors with relatively constant (online) sensory information about the co-actor's presence by presenting them with colored-light flashes that signaled the co-actor's responses. Even though actors could not see their co-actor, a full-blown SSE was obtained. This outcome demonstrates that it is not the shared presence in the same room that is important for the SSE, but it fails to clarify whether the SSE was due to the sensory feedback about the co-actor's actions or the mere belief that one is collaborating with someone else.

One problem with comparing physical acting with virtual co-acting is that this comparison confounds a number of potentially important factors, such as instructions and the availability of sensory cues. In an attempt to control for the latter, Sebanz et al. (2003) had participants wear earplugs and prevented them from seeing the other person's hand, which did not reduce the SSE. However, the co-actor was still clearly visible as was his/her involvement in the task, which does not render this manipulation particularly strong.

Two recent studies investigated the contributions of online versus offline information more systematically by providing knowledge about a second actor who was said to work on the same task in a different room (Welsh et al., 2007; Tsai et al., 2008). However, while Tsai and colleagues showed clear evidence for effects of offline information (i.e., a SSE was obtained in the physical absence of the co-actor) Welsh and colleagues did not, which renders the evidence equivocal. For evaluating this discrepancy it is informative to consider the set-up of the tasks. Tsai et al. (2008) invited participants who were already acquainted with one another prior to the testing day and allowed them to communicate via intercom before the task and during the break. In contrast, in the study of Welsh et al. (2007) the experimenter was the co-actor, who did not remind the actor of their interaction after having left the room. In other words, the actor's belief that the co-actor would still collaborate with him/her was not updated. It could thus be that offline information about the co-actor is not sufficient to establish the SSE if it is not constantly updated by online information. Therefore it is still not clear what role online information of the co-actor plays in the SSE.

In the present study, we controlled for previous acquaintance with the co-actor and made an attempt to manipulate the availability of sensory feedback about the other in a more systematic fashion. To increase control over perceptual cues, we used an auditory version of the social Simon task. Experiment 1 established this auditory version and was expected to replicate the standard SSE in the auditory domain in accordance with Ruys and Aarts (2010). Experiment 2 included a blindfold condition that eliminated all action-related visual information about the other, to see whether this would reduce or even eliminate the effect. Experiment 3 went one step further by also eliminating auditory cues about the other's actions.

EXPERIMENT 1

METHOD

Subjects

Forty participants (20 male), aged 18–30 (average age: 24.8) were randomly selected from the database of the Max Planck Institute. All participants read and signed an informed consent form for behavioral experiments before being registered into the database. All subjects were right handed (tested according to Oldfield, 1971), had normal or corrected to normal vision and had normal hearing. The subjects were invited as pairs and were asked beforehand if they were already acquainted with one another before the testing day. Acquainted participants could not participate together, and were rescheduled with new co-actors in order to keep *a priori* knowledge of the task and the co-actor as constant as possible for all pairs. Each participant performed a Single Go/No-Go task, a Joint Go/No-Go task (i.e., the Social Simon task) and a standard (solo) Simon task. Each task comprised of the same auditory stimuli. Each participant received 10.50 € for their participation.

Materials

The auditory stimuli consisted of human vocal utterances without any semantic meaning in German, the testing language. The sounds were originally generated for a functional magnetic resonance imaging study (Henk van Steenbergen, unpublished). We used the reversed and compressed Dutch words "groen" (green) and "paars" (purple) spoken by different male actors and processed using Adobe Audition 2.0 – which resulted in stimuli sounding like "oerg" and "chap." The sounds were adjusted to equal lengths of 300 ms and presented with a loudness of approximately 60 dB. Two loudspeakers were placed 50 cm to the left and right from the middle of a computer screen. Response buttons were placed 25 cm away from the computer screen, 30 cm apart from each other.

Study design and procedure

A 2 (congruent, incongruent) × 2 (Go, No-Go) × 3 (Single, Joint, Standard) factorial design was used. There were 64 trials per design cell for the Single Go/No-Go and the Standard Simon task, and 128 trials per cell for the Joint Go/No-Go task. To keep track of the performance, a feedback screen was presented after half of the trials in each condition. The feedback showed the average reaction times (RTs) and percentage correct (PC), which in the Joint condition referred to the mean performance across both participants. The task was preceded by a training phase of 25 trials per cell.

The two auditory stimuli "oerg" and "chap" were assigned to the left and right button, respectively. In the Joint condition, one participant responded to the "oerg" sound with the left button and was thus seated on the left side while the other participant responded to the "chap" sound and was seated on the right side (see **Figure 1**).

Each trial began with a warning signal, a 300 ms beep presented through both loudspeakers (symbolized by the fixation mark in **Figure 2**). After a silent period of 700 ms, the stimulus tone appeared for 300 ms through the left or right loudspeaker.

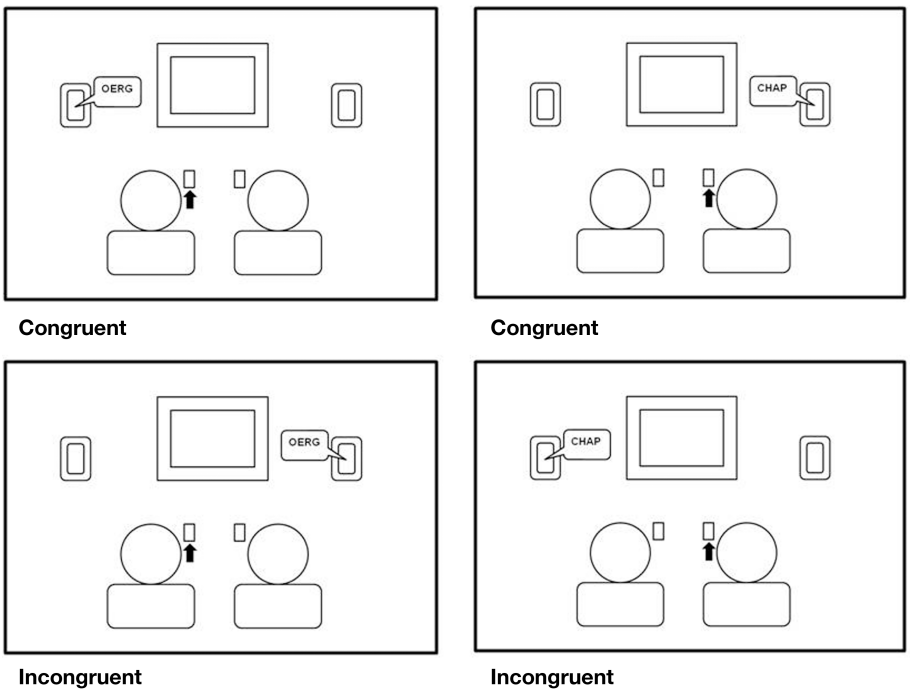


FIGURE 1 | Design of the social Simon task. The example shows a stimulus–location–congruent response (top panels) and an incongruent response (bottom panels) for the left and right located actor (left and right column, respectively).

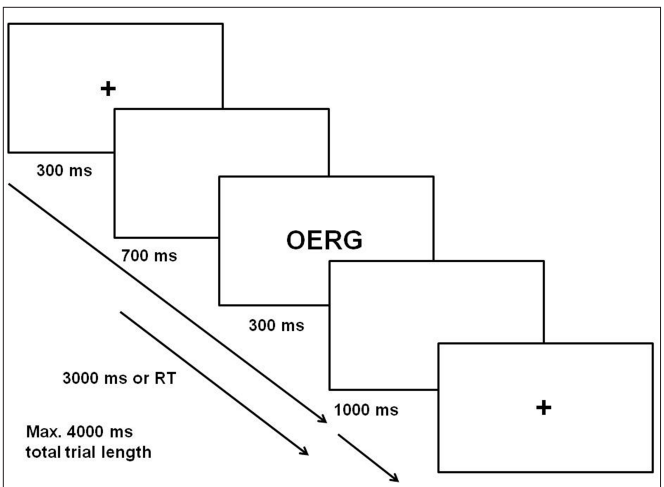


FIGURE 2 | Timing and sequence of events in Experiment 1.

The trial ended after the response was emitted but no later than 3000 ms after stimulus onset. The next trial began after another blank interval of 1000 ms.

The (social) Simon effect was measured by subtracting RTs for incongruent trials (no correspondence of stimulus location and response) from RTs for congruent trials (correspondence of stimulus location and response). Each participant performed the task under three conditions. In the Single condition, participants carried out the task alone in a separate room, sitting on one side and only responding to one sound. In the Joint condition, two participants sat in the same room, side by side, and on the same side as in the Single and Standard conditions. In the Joint condition, each participant responded to the same sound as in the Single condition. In the Standard condition, participants sat in separate rooms and responded to both sounds, but still sat on the same side as in the other two conditions. The order of Single and Joint condition was counterbalanced. The Standard task was presented last as a control condition.

RESULTS

All analyses were tested with an alpha of 0.05. The error rate was very low (Single = 0.5%, Joint = 0.6%, Standard = 3.8%) and error trials were excluded from analyses. The median RTs per participant for correct responses were entered into a two-way repeated measures ANOVA, with Type of task (Single Go/No-Go, Joint Go/No-Go, Standard Simon) and Congruency (congruent, incongruent) as independent factors (for average RTs see **Table 1**). There was a main effect of Congruency ($F(1,39) = 95.72$, $p < 0.001$, $\eta^2 = 0.711$); responses were slower in incongruent than in congruent trials ($M = 330$, SE = 9.8, and $M = 312$, SE = 9.0 respectively). The main effect of Type of task was not significant but the interaction between Congruency and Type of task was ($p < 0.001$). Paired-samples tests between congruent and incongruent trials revealed a significant congruency effect in the Standard ($t(39) = 14.48$, $p < 0.001$) and the Joint condition ($t(39) = 6.04$, $p < 0.001$), but not in the Single condition ($t(39) = -0.51$, $p = 0.61$). Given that the Joint condition comprised of twice as many trials as the other two conditions, we re-analyzed the data by considering only the first 64 trials per cell of the Joint condition, but the outcome was the same.

Table 1 | Experiment 1, mean RTs and Simon effect for the three conditions.

	Congruent	Incongruent	Simon effect
Single	323	321	–2
Joint	302	315	13
Standard	311	353	42

DISCUSSION

The outcome of Experiment 1 is straightforward: a Simon effect was obtained both in the standard and in the joint-action condition but not in the single condition. This replicates the basic findings of Sebanz et al. (2003) and extends it to auditory stimuli (in accordance with Ruys and Aarts, 2010).

EXPERIMENT 2

The aim of Experiment 2 was to eliminate visual action-related information about the co-actor without changing any other aspect of the experimental task, the context, and the instruction. We did that by having all participants wear goggles that in one group of participants were translucent, which would basically put them into the same situation as the participants of Experiment 1, but that in another group of participants were opaque. Thus, in this group, no visual online information was available, even though the participants were aware of the presence of their co-actor and heard him/her carry out the task. If visual online information would be relevant for the participant's continuous grounding of the task representation, the SSE should be weaker or absent in the blindfolded group. Alternatively, if *a priori* knowledge (offline information) is sufficient to establish the SSE, while online information is merely redundant, then we should find no reduction of the SSE in the blindfolded group.

METHOD

Forty-two participants (18 male), aged 18- to 30-years old (average age: 23.19), were selected according to the same criteria applied in Experiment 1. Each participant received 7.50 €. One pair of subjects violated the instructions not to talk during the experiment and their data were removed from analyses. The method was the same as in Experiment 1, with the following exceptions. Participants in the seeing group wore see-through glasses while participants in the blindfolded group wore opaque glasses (see **Figure 3**).

A Joint Go/No-Go task similar to the Joint condition of Experiment 1 was used. The task employed a 2 (Go/No-Go) × 2 (congruent, incongruent) × 2 (visual information present or absent) factorial design. Participants were presented with a feedback screen after half of the trials, blindfolded subjects were allowed to take off their goggles to see it. The task consisted of 128 trials per cell for each participant (in total 512 trials were presented). It was preceded by a training phase of 25 trials per cell (during the training phase the participants in the blindfolded condition were already blindfolded). Participants were instructed not to talk to each other during the experiment.

RESULTS

The error rate was again very low (1.2%). Median RTs for correct responses were entered into a two-way mixed ANOVA, with the independent variable Congruency (congruent, incongruent) as within-subjects factor and the independent variable Visual Feedback (present, absent) as between-subjects factor. There was a main effect of Congruency ($F(1,38) = 122.55$, $p < 0.001$, $\eta^2 = 0.76$); responses were slower in incongruent than in congruent trials ($M = 306$, SE = 7.9, and $M = 284$, SE = 7.2 respectively). There was neither a main effect of Visual Feedback ($F(1,38) = 0.07$, $p = 0.8$, $\eta^2 = 0.002$), nor a significant interaction ($F(1,38) = 1.04$, $p = 0.314$, $\eta^2 = 0.03$), suggesting that the Simon effects were equivalent in the two conditions (**Table 2**).



FIGURE 3 | Goggles used in the blindfolded (see left participant) and seeing conditions (see right participant).

DISCUSSION

There was no evidence whatsoever that eliminating visual online feedback about the co-actor reduced or eliminated the SSE – the numerical effect was even larger in the absence of visual information. Given that participants were blindfolded even during the training phase, each participant had only very little information about the co-actor's actions to improve on that during the task. This suggests that action and task representations do not rely on online information, but on *a priori* knowledge (offline information) to interact with a social, intentional interaction partner. However, in Experiment 2 auditory action-related online information from the button presses may have established the SSE in the blindfold condition, an issue that we addressed in Experiment 3.

EXPERIMENT 3

Although participants in Experiment 2 were prevented from processing visual online feedback, they did have access to auditory online feedback. Both co-actors were using buttons of a standard keyboard, which provided sensory cues about the other's continuous presence and responses. Experiment 3 aimed to assess the contribution from this auditory information by having pairs of seeing and blindfolded participants working either with a standard keyboard that did provide auditory feedback or with a noise-free keyboard that did not. If online auditory action-related feedback from the co-actor would play a role, the SSE should be reduced or disappear with a noise-free keyboard. Alternatively, if *a priori*

Table 2 | Experiment 2, mean RTs and Simon effect for the two conditions.

Visual feedback	Congruent	Incongruent	Simon effect
Present	283	303	20
Absent	285	309	24

knowledge (offline information) is sufficient to establish the SSE, then we should find no reduction of the SSE when eliminating visual and auditory online information.

METHOD

Forty participants (18 male), aged 18- to 30-years old (average age: 23.14), were selected according to the same criteria as in Experiment 1. The method was as in Experiment 2, with the following exceptions. In addition to the manipulation of the visual feedback between participants, the presence of auditory feedback (present, absent) was manipulated within participants. Each participant performed one block with a standard keyboard and another block with a noise-free keyboard, with the order being balanced across participants. To shorten the experiment, the length of each trial was reduced to a maximum of 2000 ms. Each participant worked through 32 trials per cell, 256 trials in total. The task was preceded by a training phase of eight trials, two per cell, during which the participants in the blindfold condition were again already blindfolded.

RESULTS

The error rate was again very low (0.8%). Median RTs for correct responses were entered into a three-way mixed ANOVA, with Congruency (Congruent, Incongruent) and Auditory Feedback (present, absent) as within-subjects factors and Visual Feedback (present, absent) as between-subjects factor. There was a main effect of Congruency ($F(1,38) = 40.99$, $p < 0.001$, $\eta^2 = 0.52$); Responses were slower in incongruent than in congruent trials ($M = 375$, SE = 12.5, and $M = 355$, SE = 11.7 respectively). There was neither a main effect of Visual Feedback ($F(1,38) = 2.84$, $p = 0.1$, $\eta^2 = 0.07$), nor of Auditory Feedback ($F(1,38) = 2.52$, $p = 0.12$, $\eta^2 = 0.06$), nor any significant interaction. The average Simon effect was similar for participants with both auditory and visual feedback (25 ms) and the participants without any feedback (26 ms; **Table 3**).

DISCUSSION

Despite having no online feedback about the other's actions, participants showed a full-blown SSE and there was not even a sign of a reduction of the effect in the absence of visual and auditory feedback. The only peculiarity in the numerical data pattern is the rather small effect in the condition with auditory but without visual feedback. However, given that, in Experiment 2, the same condition yielded a full-blown SSE comparable to the other conditions in Experiment 3, we consider this an accidental observation of no theoretical relevance. In any case, it seems clear that online visual or auditory feedback from the other is not required for the SSE to occur. Instead, the present findings suggest a central role of *a priori* knowledge (offline information) and the belief to interact with a social, intentional agent (Tsai and Brass, 2007) for the SSE.

Table 3 | Experiment 3, mean RTs and Simon effects as a function of visual and auditory feedback.

Auditory feedback	Visual feedback	Congruent	Incongruent	Simon effect
Present	Present	325	350	25
	Absent	374	382	8
Absent	Present	342	362	20
	Absent	379	405	26

CONCLUSIONS

The main aim of our study was to investigate the contribution of online visual and auditory information about a co-actor to the SSE. The very existence of the SSE suggests that action and task representations are grounded in the current situational context and consider cues about the presence and activities of co-actors. However, our present findings suggest that this grounding does not need to be continuous, in the sense that these representations can survive in the absence of ongoing visual and auditory feedback. After having established our auditory version of the social Simon task in Experiment 1 and replicated the basic findings reported by Sebanz et al. (2003), we tested the contribution of visual feedback from the other in Experiment 2 and the contribution of auditory feedback about the other's actions in Experiment 3. Even though our manipulation of auditory feedback does not rule out task-unrelated feedback from the co-actor, such as breathing noises or coughs, participants in the no-visual/no-auditory condition of Experiment 3 did not have any sensory cues about the action being performed by the other. And yet, a full-blown SSE was obtained.

What matters for the SSE does not seem to be online information about the social situation but the mere knowledge that a social, intentional co-actor is present. This conclusion does not support the concept of co-representation suggested by Sebanz et al. (2003). If the SSE would emerge at a representational level that does not distinguish between one's own actions and the actions of another person, and if that representational level would be fed by sensory feedback about both types of actions, one would expect the SSE to strongly rely on more or less continuous sensory action feedback. Eliminating this feedback should thus eliminate the SSE, which is not what our present findings show. Instead, what seems to matter is apparently the actor's belief that he/she is interacting with an intentional agent (Tsai and Brass, 2007), which is likely to rely on *a priori* knowledge about the intentional co-actor.

Hence, top-down effects (Liepelt and Brass, 2010) seem to be much more central to the SSE than previously thought. Top-down modulation may be even more important in the SSE than, for example, in automatic imitation research, where taking away the actor's intention reduces but does not eliminate stimulus–response priming (Liepelt et al., 2008).

This is likely to explain why the SSE is eliminated if the actor is led to believe to interact with an un-intentional agent (Tsai and Brass, 2007). It also provides some pointers to why Tsai et al. (2008) were able to produce an SSE but Welsh et al. (2007) were not. As discussed already, the participants of Tsai et al., but not

of Welsh et al., were repeatedly updated about the presence of their co-actor, which was likely to strengthen the actor's belief in this presence.

To summarize, we show that task representations can be grounded in offline information, but the contradictory findings between Welsh et al. (2007) and Tsai et al. (2008) remind us that this offline information may be kept active for only a limited amount of time. Note that our setup did not prevent participants from seeing and talking to each other upon arrival, before the actual experiment began, and it might be that experience (or the memory of it) that provides the sensory grounding for the task co-representations responsible for the SSE. In any case, we can conclude that task co-representations do not rely on online feedback about the other person's actions.

ACKNOWLEDGMENTS

We would like to thank Henrik Grunert, Andreas Romeyke, Thomas Pronk, and Kim Ouwehand for their help in programming and technical support, and Henk van Steenbergen from Leiden University for providing the sounds.

REFERENCES

Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representation*. New York: Cambridge University Press.

Elsner, B., and Hommel, B. (2001). Effect anticipation and action control. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 229–240.

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychol. Res.* 73, 512–526.

Hommel, B., Colzato, L. S., and Wery van den Wildenberg, P. M. (2009). How social are task representations? *Psychol. Sci.* 3, 1–5.

James, W. (1890). *The Principles of Psychology*, Vol. 2. New York: Dover Publications.

Liepelt, R., and Brass, M. (2010). Top-down modulation of motor priming by belief about animacy. *Exp. Psychol.* 57, 221–227.

Liepelt, R., von Cramon, D. Y., and Brass, M. (2008). What is matched in direct matching? Intention attribution modulates motor priming. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 578–591.

Lotze, R. H. (1852). *Medicinische Psychologie oder die Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Ruys, K. I., and Aarts, H. (2010). When competition merges people's behavior: interdependency activates shared action representations. *J. Exp. Soc. Psychol.* 46, 1130–1133.

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21.

Simon, J. R., and Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *J. Appl. Psychol.* 51, 300–304.

Stock, A., and Stock, C. (2004). A short history of ideo-motor action. *Psychol. Res.* 68, 176–188.

Tsai, C.-C., and Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Co-acting with a human hand versus a wooden hand in a dyadic interaction. *Psychol. Sci.* 18, 1058–1062.

Tsai, C. C., Kuo, W. J., Hung, D., and Tzeng, O. (2008). Action co-representation is tuned to other humans. *J. Cogn. Neurosci.* 20, 2015–2024.

Welsh, T. N., Higgins, L., Ray, M., and Weeks, D. J. (2007). Seeing vs. believing: is believing sufficient to activate the processes of response co-representation? *Hum. Mov. Sci.* 26, 853–866.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 July 2010; accepted: 02 November 2010; published online: 03 December 2010.

Citation: Vlainic E, Liepelt R, Colzato LS, Prinz W and Hommel B (2010) The virtual co-actor: the social Simon effect does not rely on online feedback from the other. Front. Psychology ***1****:208. doi: 10.3389/fpsyg.2010.00208*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Vlainic, Liepelt, Colzato, Prinz and Hommel. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 06 December 2010
doi: 10.3389/fpsyg.2010.00213

The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions

Klaus Kessler* and Hannah Rutherford

Institute of Neuroscience and Psychology, Social Interaction Research Centre, University of Glasgow, Glasgow, UK

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Paul D. Siakaluk, University of Northern British Columbia, Canada
Tad Brunye, US Army NSRDEC and Tufts University, USA

****Correspondence:***
Klaus Kessler, Department of Psychology, Room 610, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK.
e-mail: k.kessler@psy.gla.ac.uk

We set out to distinguish level 1 (VPT-1) and level 2 (VPT-2) perspective taking with respect to the embodied nature of the underlying processes as well as to investigate their dependence or independence of response modality (motor vs. verbal). While VPT-1 reflects understanding of what lies within someone else's line of sight, VPT-2 involves mentally adopting someone else's spatial point of view. Perspective taking is a high-level conscious and deliberate mental transformation that is crucially placed at the convergence of perception, mental imagery, communication, and even theory of mind in the case of VPT-2. The differences between VPT-1 and VPT-2 mark a qualitative boundary between humans and apes, with the latter being capable of VPT-1 but not of VPT-2. However, our recent data showed that VPT-2 is best conceptualized as the deliberate simulation or emulation of a movement, thus underpinning its embodied origins. In the work presented here we compared VPT-2 to VPT-1 and found that VPT-1 is not at all, or very differently embodied. In a second experiment we replicated the qualitatively different patterns for VPT-1 and VPT-2 with verbal responses that employed spatial prepositions. We conclude that VPT-1 is the cognitive process that subserves verbal localizations using "in front" and "behind," while VPT-2 subserves "left" and "right" from a perspective other than the egocentric. We further conclude that both processes are grounded and situated, but only VPT-2 is embodied in the form of a deliberate movement simulation that increases in mental effort with distance and incongruent proprioception. The differences in cognitive effort predict differences in the use of the associated prepositions. Our findings, therefore, shed light on the situated, grounded and embodied basis of spatial localizations and on the psychology of their use.

Keywords: grounding, embodiment, movement simulation, perspective taking, social cognition, spatial language

INTRODUCTION

In this study we set out to investigate the differences between two forms of visuo-spatial perspective taking in terms of embodied processing and regarding the consequences for the situated and grounded use of projective spatial prepositions (i.e., "left," "right," "in front," and "behind").

Flavell et al. (1986) have categorized the ability to understand someone else's visuo-spatial perspective into level 1 and level 2 perspective taking (VPT). While level 1 (VPT-1) reflects understanding of *what* lies within someone else's line-of-sight, i.e., which objects are visible and which occluded ("I know what you can see;" see **Figure 1**), level 2 (VPT-2) involves mentally adopting someone else's spatial point of view and understanding *how* the world is represented from this virtual perspective (i.e., "I see the world through your eyes") as shown in **Figure 1**. As another example please imagine we would like to tell a friend that she has an eyelash on her left cheek, which would require determining "left" and "right" from our friend's perspective – independently from our own point of view. Or think of way descriptions, where an instruction like "in front of the building turn left" assumes that the instructing and the instructed persons are aligned into the same virtual perspective, i.e., that they both either mentally face the entrance from the outside or imagine coming out of the building.

These examples point out the importance of VPT in communication, e.g., for establishing a common reference frame for understanding spatial localizations or more generally for establishing a shared view of the world (Frith and Frith, 2007). VPT-2 is regarded as the more complex process of the two, which is evidenced by a later ontogenetic development, specific difficulties experienced by autistic children, and by phylogenetic differences. VPT-1 develops around the age of 2 years and autistic children do not experience particular difficulties with this task (Leslie and Frith, 1988; Baron-Cohen, 1989). In contrast, VPT-2 develops around 4–5 years (Gzesh and Surber, 1985; Hamilton et al., 2009), but not in children diagnosed with autism spectrum disorder (ASD), and VPT-2 performance is predicted by theory of mind (ToM) score (Hamilton et al., 2009). Primates seem capable of certain forms of VPT-1 but not at all of VPT-2 (Tomasello et al., 2005). The latter conforms to their inability to perform simple ToM tasks (Call and Tomasello, 1999), which pose no problem for 5-year-old (non-autistic) children.

However, primates (Tomasello et al., 1998; Brauer et al., 2005) and other species (Scheumann and Call, 2004; Pack and Herman, 2006) have been reported to physically align their perspective with humans. Apes even deliberately change their position to be able to look around obstacles and share what a human experimenter can see (Tomasello et al., 1998; Brauer et al., 2005). This reflects

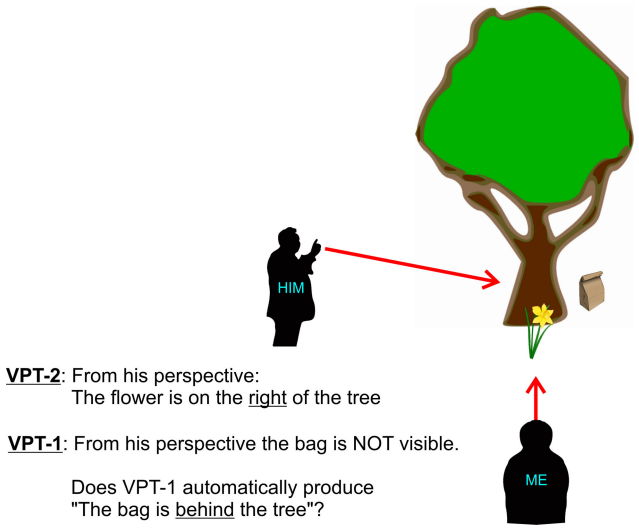


FIGURE 1 | Level 1 vs. level 2 perspective taking. According to Flavell et al. (1986) level 1 perspective taking (VPT-1) requires understanding of what lies within someone else's line-of-sight. Level 2 (VPT-2) involves mentally adopting someone else's spatial point of view. Determining that the flower is on the "right" of the tree from the other person's perspective requires a more complex transformation than VPT-1. VPT-2 has been generally related to tasks that require relative judgments and which prominently include verbal localizations that use "left of" and "right of" (Michelon and Zacks, 2006). In contrast, VPT-1 has been related so far only to visibility judgments, but we propose that VPT-1 also extends to the language domain and subserves verbal localizations that use "in front of" and "behind of." Further explanations in the text.

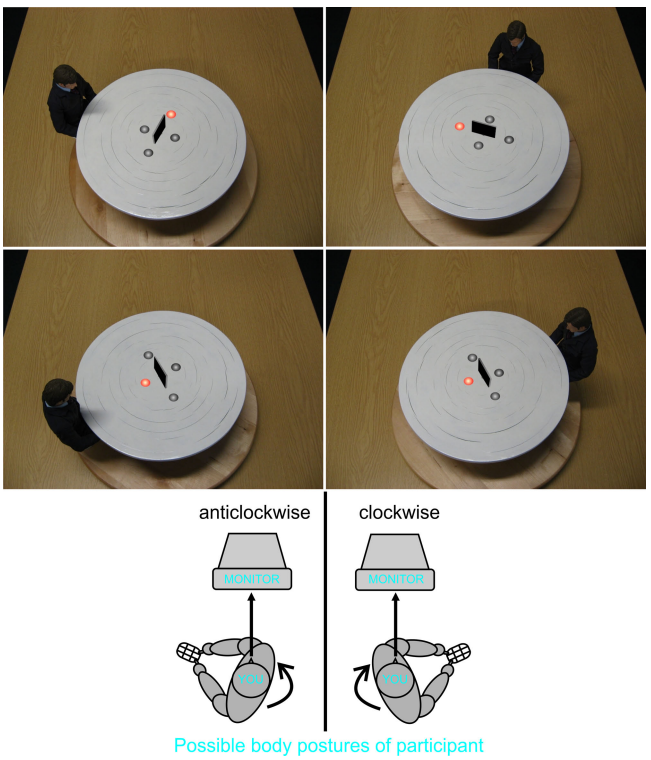


FIGURE 2 | Stimuli (top four images) and posture manipulations (drawings at the bottom). The top two images show examples for VPT-2, i.e., the target (red) is left (left image) and right (right image), respectively, from the avatar's perspective. The two images below show two examples for VPT-1: the target is either visible (left image) or occluded (right image) to the avatar. These stimuli were used in both Experiments. In Experiment 1, participants pressed a key to indicate whether the target was visible/occluded (VPT-1) or left/right (VPT-2). The same stimuli were used in Experiment 2 where participants responded verbally whether the target was "in front"/"behind" or "left"/"right" of the occlusion from the avatar's perspective. Conform to Kessler and Thomson (2010) we employed several angular disparities (60, 110, 160, 200, 250, 300) and a manipulation of the participant's body posture (congruent vs. incongruent to the direction of the avatar's location) as shown at the bottom. Further explanations in the text.

the basic understanding that a physical (apes) or mental effort (humans, i.e., VPT-2) is sometimes necessary in order to understand someone else's view of the world (Frith and Frith, 2007), which led us to hypothesize that VPT-2 might have originated from deliberate physical alignment of perspectives exhibited by apes (Kessler and Thomson, 2010).

We reasoned that if this was the case then VPT-2 would still be an "embodied" process in the sense that it relies on the posture and action repertoire of the body. Our recent findings (Kessler and Thomson, 2010) have indeed confirmed that although VPT-2 is a high-level cognitive process it is not a purely abstract transformation of a reference frame or coordinate system as had been the established view within linguistics/computational linguistics (Retz-Schmidt, 1988). Instead, we found substantial evidence that VPT-2 is relying on action-related and proprioceptive representations of the body. Specifically, we altered the body posture of the participants before each trial (cf. **Figure 2**) so that their body was either congruent or incongruent (in some experiments also neutral) with the direction of VPT-2. This simple manipulation had a dramatic effect on reaction times, where a congruent posture speeded up processing while an incongruent slowed it down. Based on the pattern of results across four experiments, we concluded that the embodiment of VPT-2 is best conceptualized as the deliberate emulation or simulation of a body rotation, supporting the notion of endogenous sensorimotor embodiment (Kessler and Thomson, 2010). This conforms to Wilson's sixth and most powerful meaning of embodied cognition: "6. Off-line cognition is body based. Even when decoupled from the environment, the activity of the mind is grounded in mechanisms that evolved for interaction with the environment – that is, mechanisms of sensory processing and motor control" (Wilson, 2002, p. 626).

In the experiments presented here we compared VPT-1 and VPT-2 in terms of their embodiment. While we expected to replicate the evidence for movement simulation/emulation subserving VPT-2 we did not know whether a similar process would also subserve VPT-1. In fact, the onto- and phylogenetic differences we have mentioned above suggested that the underlying processes could differ quite substantially and Michelon and Zacks (2006) provided conclusive evidence for a qualitative difference between VPT-1 and VPT-2: Congruent to our previous findings (Kessler and Thomson, 2010) and to results reported by others (Huttenlocher and Presson, 1973; Levine et al., 1982) VPT-2 showed an increase in reaction times with increasing angular disparity, while reaction times in the VPT-1 task remained flat across angles.

Michelon and Zacks (2006) concluded that VPT-1 is based on imagining the other person's line-of-sight which determines the relevant inter-object spatial relations, while VPT-2 requires some sort of mental rotation. A very similar distinction was suggested by

Kessler (2000) based on a connectionist network model for processing spatial prepositions: While a mental self-rotation (i.e., VPT-2) is necessary for understanding "left" and "right" from a different perspective, "in front" and "behind" from the same perspective are solved based on between-object relations that "compute" the line-of-sight. For the latter imagine someone telling you "the bag is behind the tree." For determining "behind" it is only necessary to draw a line between that person and the tree and the bag would be on the side of the tree that is occluded from the person's view (cf. **Figure 1**). Our expectation therefore was that VPT-1 subserves "in front" and "behind" judgments and that VPT-1 does not rely on the simulation of a body movement. Hence, VPT-1 would either not be "embodied" at all or very differently "embodied" than VPT-2.

Accordingly, we also expected VPT-1 and VPT-2 to rely on different neural substrates, although we could not directly test this hypothesis in our behavioral experiments. VPT-2 could either be a form of action simulation that involves action control areas in the posterior frontal cortex together with body schema representations in the parietal lobe, or VPT-2 could be a form of action emulation, where the perceptive and proprioceptive outcomes of the transformation are generated without the need for a full movement simulation that instantiates all the intermediate steps "to get there." The distinction between emulation and simulation is rather gradual in this context (and "simulation" will be used throughout the document), but in an extreme scenario emulation might not involve action control areas at all, while essentially relying on transformations within body schema and other proprioceptive areas. The body schema would be involved in any case, which is indeed supported by a growing number of findings where the temporo-parietal junction was identified as an essential substrate for VPT-2 (Zacks and Michelon, 2005; Arzy et al., 2006; Keehner et al., 2006).

In contrast, if VPT-1 solely relies on understanding spatial relations between a person and at least two objects, then the primary substrate of VPT-1 should be the dorsal between-object system in parietal cortex (e.g., Goodale and Milner, 1992; Ungerleider and Haxby, 1994). To our knowledge no neuroimaging data are yet available for VPT-1, hence, the behavioral results presented here will provide a first hint for whether motor simulation/emulation networks are a likely or unlikely neural substrate for VPT-1.

Kessler (2000) argued based on his connectionist model that qualitatively different processes are employed for the two dimensions of projective spatial prepositions ("in front"/"behind" vs. "left"/"right"), and explicitly related these processes to the different neural substrates mentioned above (i.e., to the between-object system and to the motor simulation/emulation systems, respectively). In psycholinguistics there have been suggestions to relate the use of "in front" and "behind" to the line-of-sight (Grabowski, 1999; Grabowski and Miller, 2000; Kessler, 2000), yet, to our knowledge no explicit link has been established so far to VPT-1 as the underlying cognitive mechanism. In the model by Kessler (2000) all spatial dimensioning starts with the extraction of a so-called "anchor"-direction, which is basically the line-of-sight of that perspective and which automatically produces the "in front" pole in relation to the relatum (the reference object, i.e., the tree in **Figure 1**). "Behind" requires an additional processing step for determining the opposite direction to "in front." In agreement with Grabowski (1999) this predicts faster production times for "in front" compared to "behind," which has been shown to be the case (Herrmann et al., 1987; Bryant et al., 1992). Furthermore, if the assumption is correct that VPT-1 is the underlying process, then response times for "visible" judgments should be faster than for "occluded."

Our first two steps to investigate the proposed link between the two horizontal dimensions of projective prepositions and the two levels of VPT were as follows. Firstly, we aimed to show that VPT-1 and VPT-2 are indeed differently embodied cognitive processes. Secondly, we aimed to replicate our effects using a verbal response (i.e., participants saying "left," "right," "in front," or "behind") in order to test our assumption that "left"/"right" is subserved by VPT-2 whereas "in front"/"behind" by VPT-1. So far we had exclusively employed spatially mapped key presses for VPT-2 (Kessler and Thomson, 2010). We expected that the embodiment effect observed in these experiments would persist as we believed it to be a defining characteristic of VPT-2 that did not depend on the response modality. However, a result to the contrary would be important, forcing us to adjust our theoretical considerations. It would mean that only in the case of a spatially mapped motor response a body rotation is fully simulated (cf. Kessler and Thomson, 2010), while in the case of a verbal response VPT-2 could rely on the transformation of a more abstract "disembodied" (e.g., geometric) representation of the egocentric perspective (e.g., Mahon and Caramazza, 2008). Qualitative differences in spatial representation updating after a physical or an imagined self-rotation have indeed been reported for motor (pointing) vs. verbal ("left"/"right") responses (de Vega and Rodrigo, 2001). In our case our primary interest was whether the effect of the participant's body posture on processing time depends on response modality or not.

To this end we conducted two experiments. The first experiment employed a motor response to indicate the target's location, conform to Michelon and Zacks' as well as Kessler and Thomson's procedures. This ensures comparability to previously reported results and allows for an optimal first comparison between VPT-1 and VPT-2 with respect to embodiment. The second experiment, however, employed a voice key to measure response time and to record verbal localizations of the target by means of the prepositions of interest. This would reveal whether our previous findings (Kessler and Thomson, 2010) regarding the embodiment of VPT-2 as well as our potential new findings from Experiment 1 would actually generalize to the language domain.

EXPERIMENT 1

The goal of this experiment was to investigate the differences between VPT-1 and VPT-2. Firstly, we aimed at replicating the difference reported by Michelon and Zacks (2006) regarding the dependence on angular disparity: While reaction times (RTs) strongly increased with angle for VPT-2, RTs for VPT-1 remained constant across angles. This predicted an interaction between task (VPT-1 vs. VPT-2) and angular disparity (60°, 110°, 160°). Secondly, we wanted to replicate our previous results (Kessler and Thomson, 2010) where we found a strong body posture effect for VPT-2 (congruent posture faster than incongruent) and compare this pattern of results to the one obtained for VPT-1 with the identical posture manipulation. We expected that VPT-1 would not be embodied in the same way as VPT-2. Hence, we also expected an interaction between task and body posture (congruent vs. incongruent).

MATERIALS AND METHODS

Participants

All procedures were in concordance with the declaration of Helsinki and approved by the local ethics committee. Participants were volunteers, right-handed, had normal or corrected-to-normal vision, were naive with respect to the purpose of the study, and received payment or course credit for participation. Sixteen females and 8 males took part in Experiment 1. Mean age was 23.1 years. One female participant had to be excluded due to excessive response times in the VPT-2 task (more than 3 standard deviations away from the sample mean). According to her self-report she had difficulties in general to determine left and right, even from her own (egocentric) perspective.

Stimuli and design

We employed a VPT-2 task congruent to the one originally employed by Kessler and Thomson (2010) revealing the embodied nature of VPT-2, and we added a VPT-1 task. The stimuli are shown in **Figure 2**, where an avatar was seated at one of six possible angular disparities (60°, 110°, 160° clockwise and anticlockwise) around a table. Pictures were taken from a vertical angle of 65°. Stimuli were colored bitmaps with a resolution of 1024 by 768 pixels corresponding to the graphic card settings during the experiment. Viewing distance was 100 cm and a chin rest was employed to ensure constancy.

In the center of the stimulus-table four gray hemispheres (potential targets) were arranged around an occluder (**Figure 2**). On each trial one of the hemispheres turned red indicating its status as the target. Hence, from the avatar's perspective the target on a given trial could be left, right (VPT-2), visible, or occluded (VPT-1). In this first experiment participants either pressed a key to indicate the target's left/right location from the avatar's point of view, or they pressed a key from another set of two to indicate whether the target was visible or occluded to the avatar. Due to the different pairs of response keys for the two tasks (VPT-1 vs. VPT-2), tasks were blocked into miniblocks of 24 trials each, to allow for optimal response preparation within each block (index fingers above the appropriate pair of keys).

Conform to Kessler and Thomson (2010) we also varied the body posture of the participants randomly across trials (**Figure 2**). The body in relation to the head/gaze direction could be turned clockwise or anticlockwise, hence, being either congruent or incongruent with respect to the direction of the avatar's sitting position (at either clockwise or anticlockwise angular disparities of 60°, 110°, 160°). Participants moved the response device (Targus® wireless number keypad) together with their body, while their head remained on the chin rest gazing ahead. Markings on the table indicated exactly were to place the numberpad each way to ensure a constant angle of ±60° (clockwise/anticlockwise) between body and gaze direction across trials. We administered a total of 336 trials with 28 trials in each cell of the $2 \times 3 \times 2$ design consisting of the factors "task" (VPT-1 or VPT-2), "angular disparity" (60°, 110°, or 160°; collapsed across clockwise and anticlockwise disparities), and "body posture" (congruent or incongruent to the clockwise or anticlockwise direction of the avatar's location). In separate analyses for VPT-1 and VPT-2 we also included the factor "response" (visible vs. occluded and left vs. right, respectively). Our dependent variable was response time (RT) of correct responses only and we employed individual medians for each condition and participant to reduce distortions by outlier RTs. Error rates are reported in **Table 1**.

Procedure

E-Prime® 2.0 was used for experimental control. The two tasks (VPT-1 vs. VPT-2) were presented in alternating miniblocks of 24 trials each (14 miniblocks in total; initial task balanced across participants). Every block started with an instruction about the given task. For VPT-2 participants were instructed to press the "4" key (colored yellow) with their left index for "left" and the "6" key (colored blue) with their right index for "right" on the wireless numberpad. For VPT-1 participants were instructed to press the "0" key (colored green) with their left index to indicate "visible" and the "." key (colored brown) with their right index to indicate "occluded." We did not choose a vertical key alignment in the latter case in order to avoid interference or congruence effects between target location and key location at 160° angular disparity. For instance if one would choose the top key for "occluded" responses and the bottom key for "visible" responses, then the key alignment and the target locations could mismatch at 160° (the "visible" target is above the "occluded" target, cf. **Figure 2**). In essence our response key mapping corresponded to the mapping employed by Michelon and Zacks (2006).

Every trial started with a picture displaying the posture instruction (cf. **Figure 2B**). When participants had assumed the correct posture they pressed both response buttons to proceed. A fixation cross was then shown for 500 ms and was automatically replaced by the experimental stimulus. Participants were instructed to respond as quickly and as accurately as possible. Twelve practice trials were administered in form of a miniblock of 6 VPT-2 trials and a miniblock of 6 VPT-1 trials.

RESULTS

In our first analysis we compared the two tasks (VPT-1 vs. VPT-2) together with the factors "body posture" (congruent vs. incongruent) and "angular disparity" (60°, 110°, 160°). We also conducted two separate analyses for each task, where we included the two possible responses as an additional factor: "left" vs. "right" for VPT-2 and "visible" vs. "occluded" for VPT-1. This allowed us to test for asymmetries within each task. For instance all participants were right-handed; it could therefore be that "right"-responses that

Table 1 | Mean error rates per condition for both experiments (a value of 1.00 would mean that one mistake was committed on average in a particular condition).

	60		110		160		
	Cong	Incong	Cong	Incong	Cong	Incong	Sum
EXPERIMENT 1							
VPT-1	0.87	0.70	0.91	0.57	0.78	0.39	4.22
VPT-2	0.70	0.57	0.26	0.83	0.83	1.13	4.30
EXPERIMENT 2							
VPT-1	0.00	0.13	0.04	0.00	0.04	0.00	0.22
VPT-2	0.00	0.17	0.30	0.26	0.39	0.35	1.48

required a keypress with the right index were faster than "left"-responses with the left index. With respect to VPT-1 we expected faster responses to visible targets than to occluded ones –due to fewer inter-object-relations required for processing. That is, determining visibility requires only the direct relation between the avatar's line-of-sight and the target, while determining occlusion also requires the occluder to be processed, involving three objects and their inter-relations (see Introduction).

Prior to each of these three multifactorial analyses we conducted Mauchly's sphericity tests and whenever the sphericity assumption was violated ($p < 0.05$) we conducted multivariate analyses of variance (MANOVA) conform to recommendations in the literature, as this test does not assume sphericity (Davidson, 1972; Obrien and Kaiser, 1985; Vasey and Thayer, 1987). Error Rates are shown in **Table 1** and were low and inconsistent with a general speed-accuracy-trade-off.

Combined analysis for VPT-1 and VPT-2

Sphericity was violated, so a 2 × 3 × 2 MANOVA was employed with "task," "angular disparity," and "body posture" as factors. The analysis revealed significant main effects of all three factors ($p < 0.01$), but most importantly the expected interactions between task and angular disparity ($F(2,21) = 26.1$, $p < 0.00001$, $\eta_p^2 = 0.658$) as well as between task and body posture ($F(1,22) = 30.7$, $p < 0.00001$, $\eta_p^2 = 0.583$) reached significance and strong effect sizes (compare **Figure 3A**). These interaction effects will be analyzed further in relation to each task separately in the next two sub-sections (Analysis for VPT-2 and Analysis for VPT-1).

Analysis for VPT-2

Sphericity was violated, so we employed a 3 × 2 × 2 MANOVA with the factors "angular disparity," "body posture," and "response" (left vs. right). The analysis revealed significant main effects of angle ($F(2,21) = 25.6$, $p < 0.00001$, $\eta_p^2 = 0.661$) and of body posture ($F(1,22) = 33.4$, $p < 0.00001$, $\eta_p^2 = 0.603$). Although all participants were right-handed we did not find significantly faster responses with the right index ($p = 0.62$). Somewhat in contrast to our previous results (Kessler and Thomson, 2010) we did not find a significant interaction between angle and body posture either ($p = 0.72$). As can be seen in **Figure 3A** (right graph) this is due a strong body posture effect at all angular disparities, even at the lowest of 60° (Newman–Keuls test $p = 0.0003$).

Analysis for VPT-1

Sphericity was not violated, so we employed a 3 × 2 × 2 ANOVA with the factors "angular disparity," "body posture," and "response" (visible vs. occluded). The only significant model term was the main effect of response ($F(1,22) = 7.1$, $p = 0.014$, $\eta_p^2 = 0.245$), with "visible" being consistently faster than "occluded" judgments as shown in **Figure 4** (left graph) and conform to our predictions (Grabowski, 1999; Kessler, 2000; see Introduction).

DISCUSSION OF EXPERIMENT 1

We expected VPT-1 and VPT-2 to be qualitatively different processes, and we expected this to be reflected by distinct response patterns in relation to angular disparities as well as in relation to body posture. This was confirmed by the two significant interactions (task × angle; task × body posture) in the combined analysis. Furthermore we did not find an asymmetry between spatially mapped left and right motor responses (VPT-2) while we did find that responses to visible targets were faster than to occluded targets (VPT-1). We replicated Michelon and Zacks' (2006) findings showing that response times for VPT-2 increased with angular disparity, whereas response times for VPT-1 were not affected by angular disparity and remained constant. Interestingly, we found faster responses for visible than for occluded targets conform to our hypothesis which was extrapolated from asymmetries reported for "in front"/"behind" judgments (Herrmann et al., 1987; Bryant et al., 1992). We are not aware of any previous reports of such a finding regarding visibility judgments. This finding is a first hint that VPT-1 might indeed subserve "in front" and "behind" localizations.

We have also replicated and extended our previous VPT-2 findings (Kessler and Thomson, 2010) regarding the effect of the participants' body posture by observing significantly faster RTs with a congruent than an incongruent posture at all angular disparities, even at 60°. We have confirmed that VPT-2 is embodied, in concordance with

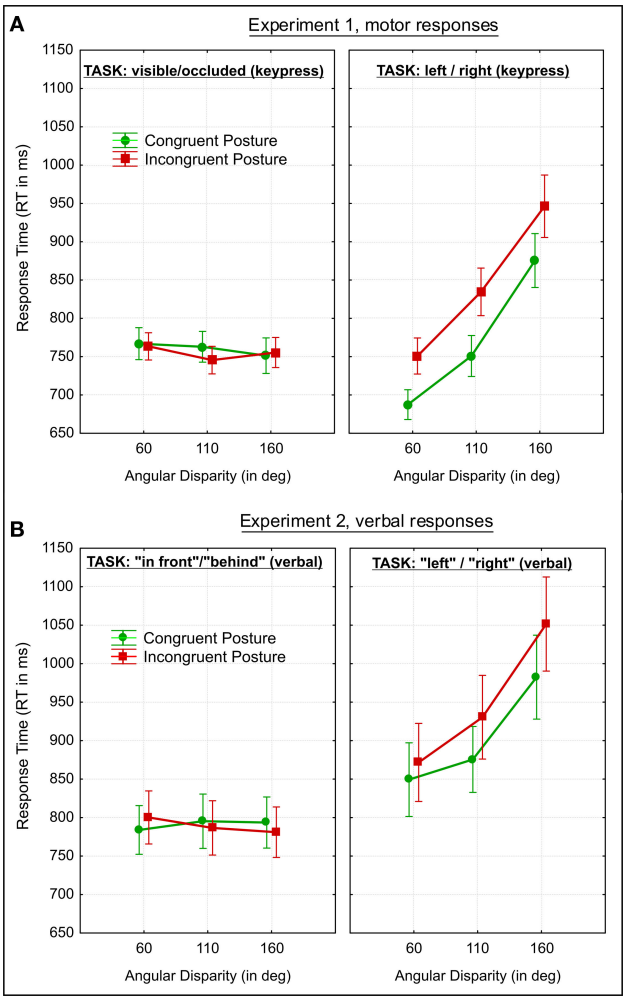


FIGURE 3 | Results for Experiment 1 and 2 in the combined analysis. Group mean response times (RT) for **(A)** Experiment 1 and **(B)** Experiment 2. Vertical bars denote the standard error of mean. Further explanations in the text.

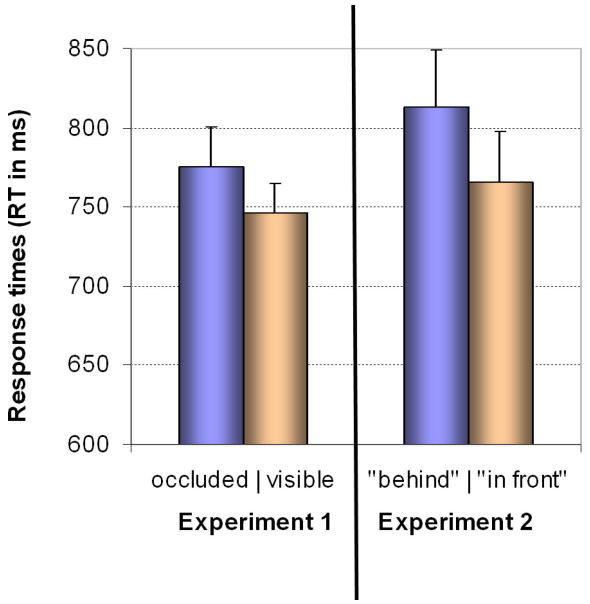


FIGURE 4 | Results for Experiment 1 and 2 in the level 1 analysis. Group mean response times (RT) for visible vs. occluded judgments with key presses in Experiment 1 (left graph) and for "in front" vs. "behind" judgments with verbal responses in Experiment 2 (right graph). Vertical bars denote the standard error of mean. Further explanations in the text.

our previous conclusion that it is the endogenous simulation of a body rotation (Kessler and Thomson, 2010). The novel finding in Experiment 1 was that VPT-1 was not affected by body posture and was therefore not at all embodied, or very differently than VPT-2. In Experiment 2 we set out to generalize these findings and conclusions to the language domain for establishing a direct link between the two levels of VPT and the two horizontal dimensions of projective prepositions ("in front"/"behind" vs. "left"/"right").

EXPERIMENT 2

MATERIALS AND METHODS

Participants

Fifteen female and 9 male volunteers, English native speakers, right-handed, with normal or corrected-to-normal vision, who were naive with respect to the purpose of the study, and who received payment or course credit took part in Experiment 2. Mean age was 22.3 years. One female participant had to be excluded from data analysis due to excessively slow response times. After the experiment the participant disclosed that she was diagnosed with a "slow processing" expression of dyslexia.

Stimuli, design, and procedure

Exactly the same stimuli and design were employed as in Experiment 1. Also the procedure was largely the same with the posture instruction at the beginning of each trial (cf. **Figure 2**), the 500 ms fixation cross, and the subsequent presentation of the experimental stimulus (cf. **Figure 2**). The main and essential difference was that we used a voice key (Logitech® headset in combination with DMDX software version 4 http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm) that provided us with voice onset times (RTs) as well as with a recording of what had been said on a given trial. The voice onset threshold was tuned for each participant individually and we verified/adjusted the voice onset time on every single trial before proceeding with the analysis. Specifically, the recordings enabled us to determine for each response whether it had been correct and whether the voice key had been indeed triggered by the onset of the verbal response and not by any other acoustic event (e.g., smacking of lips) or if the voice onset had been missed by the voice key. If the automatic voice onset detection had been incorrect we re-measured the voice onset time with audio editing software (Music Editor v8.2.5). This verification procedure was essential as the initial phonemes varied across the four prepositions. To further validate our procedure for extracting the response times (RTs) we sampled 184 response audio files (46 files per preposition, 2 from each participant) from the total pool of responses and asked 8 independent Raters to determine the voice onset times in each of these 184 response files. For each Rater the difference between their RT for a given file and our manually adjusted voice key RT was calculated. We then calculated average differences for each preposition to see whether there was a specific bias in our data that could have affected the outcomes (e.g., "right" being systematically delayed compared to "left"). The average differences for each preposition were as follows (with minimum and maximum across Raters in brackets): behind = −4.79 ms (min. = −12.22; max. = 15.48), in front = −1.78 ms (min. = −12.39; max. = 15.78), left = −5.68 ms (min. = −14.35; max. = 15.26), right = −2.68 ms (min. = −11.07; max. = 22.83). None of the Raters revealed specific distortions for a particular preposition. This validation reveals that our procedure worked well, but also, that it was not possible to determine the voice onset with millisecond accuracy, which we have to take into consideration when interpreting our results. Most importantly for the interpretation of our data, no bias was observed for the RTs of a particular preposition to be systematically over- or underestimated in relation to the others to an extent that could explain our effects.

Error rates were particularly low in this Experiment (see **Table 1**). We analyzed RTs only for the correct responses and we employed individual medians for each condition to reduce distortions by outlier RTs. The two tasks ("in front"/"behind" vs. "left"/"right") were presented again in alternating miniblocks of 24 trials each for maximum comparability to Experiment 1. At the beginning of each miniblock participants were instructed about the two verbal alternatives they were expected to use: "left"/"right" or "in front"/"behind." Again two miniblocks of 6 trials each were administered for practising VPT-2 and VPT-1, respectively, before the 336 experimental trials were presented.

RESULTS

As for Experiment 1 we conducted three analyses. In the first we compared the two tasks ("in front"/"behind" vs. "left"/"right") together with the factors "body posture" (congruent vs. incongruent) and "angular disparity" (60°, 110°, 160°). We then conducted two separate analyses for each task, where we included the two possible responses as an additional factor: "left" vs. "right" and "in front" vs. "behind," respectively. This allowed us to test for asymmetries between the poles of each dimension separately. Error Rates were particularly low and are shown in **Table 1**.

Combined analysis

Sphericity was violated, so we employed a 2 × 3 × 2 MANOVA with "task," "angular disparity," and "body posture" as factors. As in Experiment 1 the analysis revealed significant main effects of all three factors ($p < 0.01$), and the expected interactions between task and angular disparity ($F(2,21) = 18.1$, $p < 0.0001$, $\eta_p^2 = 0.595$) as well as between task and body posture ($F(1,22) = 21.33$, $p < 0.001$, $\eta_p^2 = 0.492$). In contrast to Experiment 1 also the three-way interaction between task, angular disparity, and body posture reached significance ($F(1,21) = 5.9$, $p < 0.01$, $\eta_p^2 = 0.197$). The latter was mainly due to a gradually increasing posture effect with angular disparity for "left"/"right": The body posture effect reached significance at 160° and 110° (Newman–Keuls test $p < 0.001$) but not quite at 60° ($p = 0.07$). "In front"/"behind," on the other hand, revealed a strikingly similar pattern to Experiment 1 with no RT increase with angular disparity and no posture effect (**Figure 3**, left graphs). The interaction effects will be analyzed further in relation to each task separately in the next two sub-sections (Analysis for "Left"/"Right" and Analysis for "in Front"/"Behind").

Analysis for "left"/"right"

Sphericity was violated, so we employed a 3 × 2 × 2 MANOVA with the factors "angular disparity," "body posture," and "response" ("left" vs. "right"). The analysis revealed significant main effects of angle ($F(2,21) = 22.1$, $p < 0.0001$, $\eta_p^2 = 0.619$) and of body posture ($F(1,22) = 19.1$, $p < 0.001$, $\eta_p^2 = 0.498$). Conform to Experiment 1 we did not observe significantly faster responses for either "left" or "right" responses ($p = 0.14$). However, in Experiment 2 RTs were determined based on vocal responses and although our procedures for determining voice onsets worked well on average (see Stimuli, Design, and Procedure), increased variability could have masked a true difference between left and right.

Analysis for "in front"/"behind"

Sphericity was not violated, so we employed a 3 × 2 × 2 ANOVA with the factors "angular disparity," "body posture," and "response" ("in front" vs. "behind"). The main effect of response reached significance ($F(1,22) = 16.7$, $p < 0.001$, $\eta_p^2 = 0.432$), as shown in **Figure 4**, right part of the graph. The interaction between response and angle also reached significance ($F(1,22) = 3.5$, $p < 0.05$, $\eta_p^2 = 0.137$) suggesting that the difference between "in front" and "behind" judgments was consistent but decreased across angular disparities (see **Figure 5**). "In front" judgments were overall faster than "behind" judgments conform to the findings in Experiment 1 for VPT-1 by means of key presses, with "visible" being faster than "occluded" judgments (**Figure 4**, left graph).

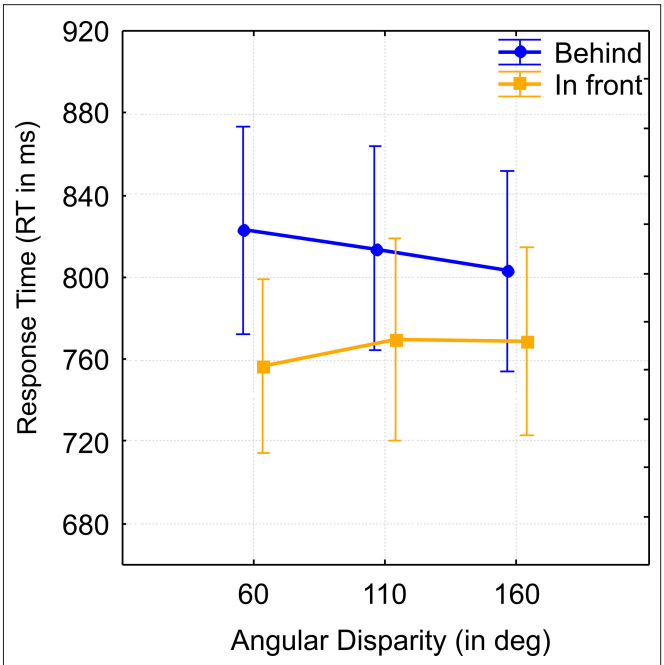


FIGURE 5 | Results Experiment 2 in the level 1 analysis. Group mean response times (RT) for "in front" vs. "behind" judgments in Experiment 2 at each angular disparity, which reflects the significant interaction between angular disparity and preposition. Vertical bars denote the standard error of mean. Further explanations in the text.

DISCUSSION OF EXPERIMENT 2

RTs for "in front" were consistently faster than for "behind" which corresponded to the asymmetry between "visible" and "occluded" responses in Experiment 1. Angular disparity also had an effect in form of an interaction but very differently compared to the VPT-2 tasks in both Experiments. Here RTs for "behind" decreased with angle while RTs for "in front" increased. This may be explained by taking into consideration that at 60° targets that are "in front" from the avatar's perspective are also closer to the observer than targets that are "behind," which are further away. This is reversed at 160° where targets that are "behind" are closer to the observer than targets that are "in front." This seems to suggest that the observer's fixed viewpoint plays a role in determining the inter-object relationships that lead to "in front"/"behind" judgments. This further underpins the qualitative difference to "left"/"right" judgments, where the observer's viewpoint is mentally shifted.

In concordance with Experiment 1 body posture did not modulate RTs for "in front"/"behind" judgments. In total we observed a strikingly similar pattern between "in front"/"behind" and visible/occluded judgments (**Figure 3** left column, **Figure 4**), which conforms to our prediction that VPT-1 subserves both types of judgments and is independent of response modality.

Verbal "left"/"right" judgments were significantly influenced by angular disparity conform to Experiment 1, our previous findings (Kessler and Thomson, 2010) and to reports by others (Kozhevnikov and Hegarty, 2001; Hegarty and Waller, 2004; Zacks and Michelon, 2005; Kozhevnikov et al., 2006). Most importantly, this task was also strongly affected by body posture. The pattern here also revealed a significant interaction between angle and posture conform to our previous findings (Kessler and Thomson, 2010).

We conclude that we accomplished our goal to generalize the data pattern obtained for the two VPT tasks with key presses (motor response) in Experiment 1 to verbal responses using spatial prepositions in Experiment 2. However, if one visually compares the RT patterns across Experiments (**Figure 3**) then it seems that RTs for VPT-2 ("left"/"right") are generally increased in Experiment 2 compared to Experiment 1 (**Figures 3A,B**, right graphs). In order to statistically substantiate this observation we conducted a direct comparison between the two Experiments.

COMPARISON BETWEEN EXPERIMENT 1 AND 2

We compared the two Experiments by means of a MANOVA (Sphericity was violated) employing the within subjects factors "task," "angular disparity," "body posture" and the between factor "experiment." The results reflected the experiment-specific findings with overall significant effects of angular disparity ($F(2,43) = 46.3$, $p < 0.00001$, $\eta_p^2 = 0.635$), body posture ($F(1,44) = 25.8$, $p < 0.00001$, $\eta_p^2 = 0.369$), task × angle ($F(2,43) = 44.8$, $p < 0.00001$, $\eta_p^2 = 0.626$), task × posture ($F(1,44) = 51.7$, $p < 0.00001$, $\eta_p^2 = 0.540$), and task × angle × posture ($F(2,43) = 4.5$, $p < 0.02$, $\eta_p^2 = 0.072$). The major difference between the two experiments was reflected in a significant interaction between task and experiment ($F(1,44) = 4.9$, $p < 0.05$, $\eta_p^2 = 0.101$) shown in **Figure 6**. The interaction was due to a larger difference in overall response times for VPT-2 between the two Experiments (Experiment 2 slower than Experiment 1) than for VPT-1.

We also compared each task (VPT-1 and VPT-2) separately between the two Experiments but did not find any significant effect of "experiment." It is, however, noteworthy that in the VPT-1 analysis the interaction between response ("in front" + "visible" vs. "behind" + "occluded") and angular disparity reached significance across both experiments ($F(2,88) = 6.2$, $p < 0.01$, $\eta_p^2 = 0.124$), suggesting that the corresponding interaction reported for Experiment 2 (see **Figure 5**) might generalize across modalities.

We compared Error Rates (ER) between the two Experiments and the two tasks in a 2 × 2 ANOVA with "task" as within and "experiment" as between subjects factors (as shown in **Table 1**, some conditions in Experiment 2 did not reveal any variance at all, so the full design could not be employed). The main effect of "experiment" reached significance ($F(1,44) = 20$, $p < 0.0001$, $\eta_p^2 = 0.313$) suggesting that error rates were lower with verbal responses (cf. **Table 1**).

Overall the comparison between Experiments further corroborated our conclusion that VPT-1 and VPT-2 were employed independently of response modality, yet, in addition we found an asymmetry between VPT-1 and VPT-2 across modalities that indicates that overall RTs are more strongly increased for VPT-2 than for VPT-1, when responses were made verbally (**Figure 6**). Error rate differences only reflected a main effect of experiment with verbal responses being more accurate, so the observed interaction with RTs is not likely to reflect a general speed-accuracy-trade-off.

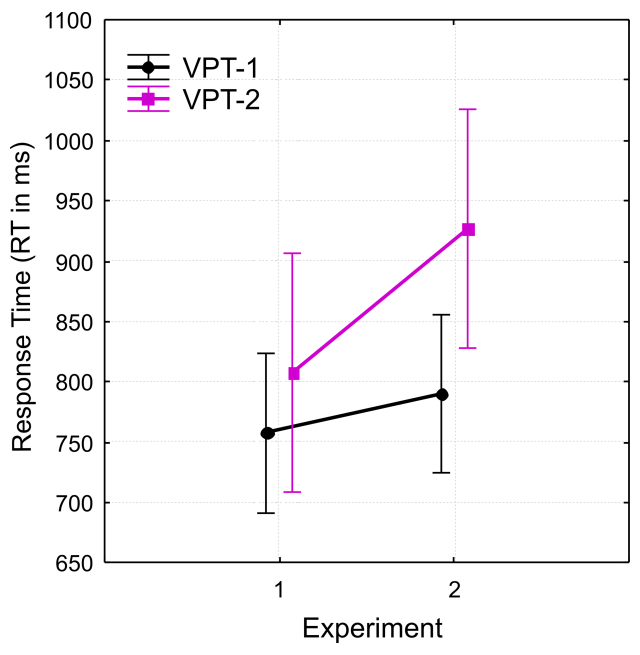


FIGURE 6 | Comparing the two Experiments. Group mean response times (RT) for VPT-1 vs. VPT-2 and for each Experiment (1 vs. 2) reflecting the significant interaction between "experiment" and "task". Vertical bars denote the standard error of mean. Further explanations in the text.

GENERAL DISCUSSION

We found revealing commonalities and differences across the two experiments. Disregarding response modality (motor vs. verbal), within each experiment we found a qualitatively distinct pattern between the two tasks, while the pattern was remarkably similar for each task across the two experiments (i.e., modalities). It is important to emphasize that these results confirm our hypothesis that the distinct patterns we observed for VPT-1 and VPT-2 regarding angular disparity and embodiment do not depend on response modality, which contrasts with some findings reported in the context of spatial updating (de Vega and Rodrigo, 2001). In the following we discuss the implications of our results in detail.

VPT-1 AND "IN FRONT"/"BEHIND"

One of our most striking results was that the response time patterns for visible/occluded and for "in front"/"behind" judgments were almost identical (cf. **Figure 3**, left column, **Figure 4**), considering that different individuals participated in the two experiments, that the task instructions differed (visibility vs. spatial location judgments), and that different response modalities were employed. The most prominent features in common were the absence of an increasing cost across angular disparities as well as the absence of an embodiment effect, i.e., of the participants' body posture (congruent vs. incongruent). These strong similarities support our hypothesis that identical processes subserve the two types of judgments, yet, our final piece of evidence is also the most convincing: In both experiments those trials, where the target was located closer to the avatar than the occluder, i.e., requiring "visible" or "in front" responses, respectively, were processed consistently faster than trials where the occluder was between the avatar and the target, i.e, in "occluded" or "behind" trials. In total we believe that this provides conclusive evidence that in scenarios where the spatial prepositions "in front" and "behind" have to be determined from the perspective of another person, level 1 perspective taking (VPT-1) is the employed cognitive process. Accordingly, our findings close a gap between cognitive and developmental psychology on the one hand and psycholinguistics on the other.

VPT-2 AND "LEFT"/"RIGHT"

With respect to VPT-2 we were able to replicate the embodiment effect in form of posture congruence effect with verbal responses, which we had previously observed with motor responses. This confirms that VPT-2 suberves "left"/"right" judgments (e.g., Michelon

and Zacks, 2006) and that disregarding response modality the default strategy consists of mentally simulating a body rotation (cf. Kessler and Thomson, 2010).

However, **Figure 6** also shows a systematic difference between Experiments 1 and 2. With verbal "left"/"right" responses VPT-2 take longer, i.e., RTs seem to be shifted up by more than 100 ms. It is puzzling that verbal "left"/"right" judgments come at such an increased cost compared to motor responses. "In front"/"behind" judgments do not seem to suffer such a cost compared to visibility judgments (visible/occluded). Also, a general speed-accuracy-trade-off does not seem to account for the particularly elevated costs of VPT-2 with a verbal response (see Comparison Between Experiment 1 and 2). We propose that for "left"/"right" judgments language processes, e.g., in form of lexical access, take longer compared to "in front"/"behind," possibly due to their larger ambiguity (e.g., Coventry and Garrod, 2004; e.g., Levelt, 1996; Grabowski, 1999). This ambiguity is reflected in larger RTs for verbal "left"/"right" responses even at low angular disparities (compare **Figure 3B**, right graph), and possibly even at 0° angular disparity as Grabowski (1999) suggested.

VPT-1 VS. VPT-2

We have replicated Michelon and Zacks' (2006) findings that VPT-1 and VPT-2 reveal qualitatively different response time patterns in relation to increasing angular disparity: While VPT-1 was not affected in any significant way, VPT-2 showed a significant increase in response time with increasing angular disparity. Our results in total corroborate the notion that the two types of perspective taking are based on two qualitatively different processes. While we replicated our previous findings (Kessler and Thomson, 2010) showing that VPT-2 strongly depends on the congruence of the participants' body posture with the direction in which the avatar is seated, we showed for the first time that no such pattern is observed for VPT-1. We conclude in concordance with Kessler and Thomson (2010) that VPT-2 is the endogenous (self-initiated) emulation or simulation of a body rotation into another perspective.

In contrast, VPT-1 seems to involve a process that determines object locations in relation to the line-of-sight of another person (Michelon and Zacks, 2006). This could be regarded as some sort of embodied representation in its own right: Gaze is a very strong social cue that has been related to processes of motor resonance (for a review Frischen et al., 2007). Perceiving someone's gaze is processed in a brain network that overlaps with gaze control (Grosbras et al., 2005). This would support the notion of embodiment of VPT-1 in the form of motor resonance, which, however, does not rely on deliberate movement simulation like VPT-2 (see Kessler and Thomson, 2010, Experiment 4 for details of this distinction). Michelon and Zacks (2006, Experiment 3) investigated VPT-1 in the absence of an avatar, that is, without an external "gaze"-anchor for establishing the line-of-sight. Humans can easily imagine a virtual line-of-sight and solve the task. The interesting question is whether such a process would be implemented in part by cortical gaze control areas. This could possibly extent the notion of embodiment of VPT-1 toward deliberate simulation in gaze coding areas. However, based on our current findings we conclude that even if VPT-1 involves gaze simulation, this form of simulation is very different from movement simulation during VPT-2.

Implications for the neural substrates of VPT-1 and VPT-2

We propose the following hypotheses regarding the neural substrates of VPT-1 and VPT-2, which seem most compatible with the currently available data. VPT-2 is either a form of action simulation that essentially involves action control areas in the posterior frontal cortex together with body schema representations in the parietal lobe and possibly vestibular information represented in the Insula, or VPT-2 is a form of action emulation, where the perceptive and proprioceptive outcomes of the transformation are generated without the need for a full movement simulation that contains all the intermediate steps "to get there." Although this distinction is rather gradual in our case an emulation process during VPT-2 might not require action control areas to be involved, while essentially relying on the emulation of representations in body schema and other proprioceptive areas (e.g., vestibular representations in the Insula).

Parietal areas and in particular the temporo-parietal junction (TPJ) indeed seem to be a prominent part of the physiological substrate of VPT-2 (Zacks and Michelon, 2005; Arzy et al., 2006; Keehner et al., 2006). For example Arzy et al. (2006) and Blanke et al. (2005) reported that the TPJ was related to disembodied processing (the self was imagined outside the body) and Samson et al. (2005) reported difficulties with VPT-2 after lesions in this area. Recent ERP mapping results corroborated the importance of the TPJ and further implicated posterior frontal areas (Schwabe et al., 2009). Zacks and Michelon (2005) also concluded that VPT-2 recruits posterior frontal areas that code for body movements but Wraga et al. (2005) questioned the involvement of full action simulation in VPT-2 compared to mental object rotation. Hence, for VPT-2 strong evidence exists in support of the temporo-parietal junction being part of the neural substrate, while posterior frontal and vestibular/insular (e.g., Blanke and Thut, 2007; Grabherr et al., 2007) contributions are still being debated.

To our knowledge no conclusive neuroimaging results exist for VPT-1, but in the light of our results simulation/emulation networks are an unlikely neural substrate. We predict that the dorsal between-objects pathway in parietal cortex would be the main processing substrate together with the ventral object/person recognition system in the temporal lobe (e.g., Goodale and Milner, 1992; Ungerleider and Haxby, 1994), but would not require additional processing input from action control or proprioceptive areas.

Implications for evolution and development

The qualitative distinction between VPT-1 and VPT-2 we observed in our data is reflected in onto- and phylogenetic differences. As we described in the Introduction, VPT-1 develops around the age of 2 years and autistic children do not seem to experience particular difficulties with this task (Leslie and Frith, 1988; Baron-Cohen, 1989). VPT-2 tends to develop around 4–5 years (Gzesh and Surber, 1985; Hamilton et al., 2009), but not in children diagnosed with autism spectrum disorder (ASD) (Hamilton et al., 2009). Primates seem capable of certain forms of VPT-1 but not at all of VPT-2 (Tomasello et al., 2005). This suggests that VPT-2 is the more advanced cognitive process, which is consistent with our RT results where VPT-2 was slower, and hence, the more demanding cognitive process than VPT-1.

We concluded (see VPT-1 vs. VPT-2, Implications for the Neural Substrates of VPT-1 and VPT-2) that VPT-2 is the endogenous (self-initiated) emulation or simulation of a body rotation into another perspective, while VPT-1 does not seem to rely on such a simulation

process. Instead VPT-1 seems to involve a process that determines object locations in relation to the line-of-sight of another person. Although movement simulation seems like a simple operation at first glance, it minimally requires the awareness of two pieces of information. Firstly, the knowledge that someone else may have a very different view of the world which requires some form of alignment before it can be fully understood, and that, secondly, one does not always have to physically change one's location in order to achieve such an alignment; instead, one can simply imagine it. While we seem to share the first step with apes who change their location to align themselves with humans or conspecifics (Tomasello et al., 1998; Brauer et al., 2005), the second stage seems to be uniquely human. The closer one looks the more impressive this latter achievement actually becomes: we use a skill that evolved millions of years ago, i.e., moving, decouple the actual execution of the movement from its planning, control, and sensory transformation and employ the abstract movement representation for a mental simulation of the representational consequences.

According to these considerations VPT-2 should develop after certain forms of mental simulation have been mastered by a given individual, specifically the skill and the awareness that one can actually imagine oneself deliberately in another location, i.e., outside our own body. The next step would be to employ such a mental operation to imagine someone else's perspective. This can be regarded as a prototype of theory of mind (ToM) where an individual infers the mental states of another person. In support of this claim Hamilton et al. (2009) found that only ToM score, but not verbal skills or mental object rotation performance, predicted VPT-2 ability in typically developing and in children diagnosed with autism. While mental object rotation performance was not impaired in the autism group, they had significantly lower ToM scores and were significantly impaired on the VPT-2 task.

So what might go wrong in autism? Can autistic individuals mentally simulate movements without actually executing them? Can they imagine themselves outside their bodies? If the answers are yes, is the transfer of this skill onto inferring someone else's mental states amiss or hampered? To our knowledge there is no report of adults diagnosed with autism that describes their ability or disability to conduct VPT-2 and the particular strategies they employ in case they have mastered this ability at some point of their individual development. A posture change like the one we employed here could reveal if autistic participants use movement simulation for VPT-2 in the same way as typical participants, or whether they have learned to employ a mental object rotation (OR) strategy instead, since OR does not seem to impose any difficulties for them (e.g., Hamilton et al., 2009). Kessler and Thomson (2010) showed that with typical participants VPT-2 as well as OR increased with angular disparity, yet, that the posture manipulation allowed for a qualitative distinction between VPT-2 and OR.

IMPLICATIONS FOR SPATIAL LANGUAGE: GROUNDED, SITUATED, AND EMBODIED PROCESSING

Both processes of perspective taking result in higher-level cognitive representations about other people's views of the world that are strongly grounded in the context of the specific situation (cf. Harnad, 1990; Barsalou, 2008; Myachykov et al., 2009). Specifically, the location and orientation of the avatar and the location of the target in relation to the avatar as well as to the occluder (or relatum in general) provide the situational input to the VPT processes. These transform the input into an abstract cognitive representation that is a direct outcome of the situational constraints and, hence, grounded in the situation.

However, based on our results we propose that for determining the situationally grounded meaning of projective spatial prepositions, "left" and "right" involve an effortful, embodied process of movement simulation that is equivalent to VPT-2, while "in front" and "behind" are determined quite effortlessly based on the generation of a "line-of-sight" and the resulting spatial relations; equivalent to VPT-1. Grabowski (1999) suggested an anthropomorphological definition of the semantics of these four spatial prepositions proposing that the human line-of-sight would determine the "front" and as a result the "behind" from a given perspective[1]. This is in full agreement with the notion we propose here (also Kessler, 2000), namely that both spatial dimensions and the related four prepositions are not only grounded in the situational constraints but are also related to the human perceptual apparatus and body. According to this notion front/behind are related to asymmetries of the human body (eyes define the front), while left/right relate to the symmetrical sides of our bodies. As a consequence the front of any other perspective can be easily determined or imagined, which automatically provides a line-of-sight and the "front" pole within the visible area in relation to a relatum (in the experiments presented here this was the occluder in the middle of the table, and in **Figure 1** it is the tree). Next, "behind" can be directly determined as the opposite pole, which is occluded from the view of the target perspective. Hence, the first horizontal dimension can be easily determined based on features and inter-object relations within the presented scene resulting in a grounded and situated "meaning" of the respective prepositions. Furthermore, our results suggest that this process is identical to VPT-1.

Due to the symmetry of our bodies "left" and "right" cannot be determined based on visual features, but have to be determined with more effort, namely, by means of a perspective transformation into the target perspective. Our results have confirmed that this transformation process is identical to level 2 perspective taking (VPT-2), which we now know is the simulation or emulation of a body rotation.

In terms of embodiment and grounding we propose that "in front"/"behind" are essentially grounded in relation to the visible features of the human body but that the extent to which embodiment – in a more powerful sense, conform to Wilson's sixth meaning of embodied cognition (2002) – is still part of the ongoing processing (e.g., simulation within gaze control areas) is an open empirical question as pointed out in a previous section (VPT-1 vs. VPT-2). "Left"/"right," in contrast, essentially rely on movement simulation/emulation when determined from a perspective that is misaligned with the egocentric viewpoint. These distinctions between embodied, grounded, and situated processing are in concordance with Myachykov et al.'s (2009) notion proposed in the context of number representations. Our care in distinguishing the "when," "how much," and "in what way" of embodiment of VPT-1 and VPT-2 and in finding different answers for the two forms of perspective taking is further in agreement with Chatterjee's (2010) recent call for more rigorous distinctions when claiming embodiment of cognitive processes.

[1]Grabowski suggests several important distinctions, e.g., "inside perspective" vs. "outside perspective" Here we simply focus on the main aspect of this notion, namely the relation to the human body as an anchor for spatial dimensioning.

Cognitive effort: implications for selecting a frame of reference

In congruence with the considerations in the previous section (Implications for Spatial Language: Grounded, Situated, and Embodied Processing), Kessler (2000) devised a model for the interpretation of spatial prepositions in complex situations with up to three competing frames of reference (FOR). For each FOR (or potential perspective) a basic direction, so-called "anchor" direction, is extracted based on the visual features of that perspective (e.g., face of another person or front of an intrinsically directed object like a car; see Grabowski, 1999 and Levelt, 1996, for discussions). This anchor direction is equivalent to "in front" (i.e., visible) while "behind" (i.e., occluded) requires a single additional step in the opposite direction. This prioritization of the front pole predicts an advantage in speed over the opposite pole, which is exactly what we found disregarding response modality (i.e., "visible" and "in front" are processed faster than "occluded" and "behind"). Furthermore such an "anchor" direction can then be employed to determine the left and right poles in the orthogonal direction: the egocentric gaze direction (egocentric anchor) is rotated into the target anchor and the egocentric left and right are mapped onto the target left and right of that perspective. This is essentially a model of perspective transformation and Kessler (2000, pp. 135, 208–210) pointed out the similarity between the model features and a gradual transformation of a body representation for action control. Our recent findings (Kessler and Thomson, 2010) and the results presented here support this hypothesis.

Hence, while the situated meaning of all four projective prepositions is strongly grounded within the spatial configuration of the situation, only localizations employing "left" or "right" necessitate a movement simulation. While one could point out that a line-of-sight could also be regarded as an embodied basis that subserves VPT-1 (see VPT-1 vs. VPT-2), our findings reveal that only the movement simulation for VPT-2 comes at a cost that increases with disparity, i.e., the distance of the simulated movement, and is modulated by the congruence of proprioceptive information (i.e., body posture). Such a qualitative distinction especially in terms of cognitive effort would have strong implications for the use of spatial prepositions in specific situations, assuming that cognitive systems tend to minimize their effort. While there is no increased effort involved in determining "in front" or "behind" from the perspective of someone sitting opposite to ourselves, it implies a mental self-rotation effort to adopt this person's perspective for "left" or "right." Therefore we might willingly adopt the frame of reference (FOR) of the other person for the use of "in front" and "behind" while we might be reluctant to do the same for "left" and "right" due to the involved cognitive effort (Kessler, 2000). Additional factors within the situational context play a role in our "willingness" to take on the effort or not, as has been shown for socio-emotional factors (Graf, 1994; Levelt, 1996; Grabowski and Miller, 2000; Kessler, 2000; Coventry and Garrod, 2004), but also for more implicit influences such as an action-related topic of the situation (Tversky and Hard, 2009).

In contrast, even the most situated and grounded psycholinguistic accounts (e.g., Grabowski, 1999; Grabowski and Miller, 2000) still implicitly assume that first and foremost a frame of reference (FOR) is chosen (e.g., egocentric, partner-centered, or intrinsic), and then all relative dimensions of space (i.e., "left"/"right" AND "in front"/"behind") are determined based on this particular FOR, thus, implying that the identical cognitive transformation into that FOR (i.e., "origo projection," Grabowski, 1999; Grabowski and Miller, 2000) underlies all dimensional relations (i.e., "in front"/"behind" AND "left"/"right"). In contrast, we have confirmed the hypothesis of Kessler (2000) that two qualitatively different cognitive processes (VPT-1 vs. VPT-2) are involved. Accordingly, we predict that different FORs may be chosen for the two horizontal dimensions ("in front"/"behind" vs. "left"/"right") in situations, where the two underlying processes (VPT-1 and VPT-2, respectively) significantly differ in terms of cognitive effort. In essence we propose a dissociation between semantic conceptions of spatial prepositions and the psychology of their situated, grounded, and embodied use.

CONCLUSIONS

Based on our findings we draw the following three major conclusions. Firstly, level 1 (VPT-1) and level 2 (VPT-2) perspective taking are qualitatively different cognitive processes, especially with respect to embodiment. VPT-2 is the mental simulation of a body movement where the effort increases with mental distance to cover (angular disparity effect) and with the level of body posture incongruence. VPT-1 relies on determining visibility of the target by imagining the line-of-sight from a given perspective. This process does not depend on mental distance (no angular disparity effect) or on movement simulation (no body posture effect). The finding that visibility is faster to judge than occlusion further supports the assumed process. VPT-2 is the more effortful and, hence, the more sophisticated cognitive process, which is consistent with the available developmental, comparative, and neuroimaging data.

Secondly, both VPT processes are applied in their essential form with key presses as well as with verbal responses in form of prepositions. We therefore conclude that VPT-1 is the cognitive process that subserves verbal localizations using "in front" and "behind" from a perspective other than the egocentric and that VPT-2 is the cognitive process that subserves verbal localizations using "left" and "right" from a perspective other than the egocentric. In the latter case, however, verbal responses come at an additional cost compared to key presses, which might reflect higher ambiguity. Both VPT processes result in a grounded and situated "meaning" of the prepositions they subserve, but only VPT-2 is embodied in the form of movement simulation.

Thirdly, the difference in cognitive effort associated with VPT-2 compared to VPT-1 implies that in specific situations language users will prefer a different, less "effortful," frame of reference for "left"/"right" than for "in front"/"behind." This prediction is at odds with the assumption of a single and general "origo projection" process for the psychological use of spatial prepositions.

ACKNOWLEDGMENTS

We would like to thank Anna Borghi and the two Referees for their very helpful comments for improving our manuscript. This research was supported by ESRC/MRC funding (RES-060-25-0010) awarded to Klaus Kessler. Finally we would like to thank the eight Raters for their help with validating our procedures.

REFERENCES

Arzy, S., Thut, G., Mohr, C., Michel, C. M., and Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *J. Neurosci.* 26, 8074–8081.

Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *Br. J. Dev. Psychol.* 7, 113–127.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., Landis, T., and Thut, G. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *J. Neurosci.* 25, 550–557.

Blanke, O., and Thut, G. (2007). "Inducing out-of-body experiences," in *Tall Tales: Popular Myths About the Mind and Brain*, ed. S. Della Sala (Oxford: Oxford University Press), 425–439.

Brauer, J., Call, J., and Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *J. Comp. Psychol.* 119, 145–154.

Bryant, D. J., Tversky, B., and Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *J. Mem. Lang.* 31, 74–98.

Call, J., and Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Dev.* 70, 381–395.

Chatterjee, A. (2010). Disembodying cognition. *Lang. Cogn.* 2, 79–116.

Coventry, K. R., and Garrod, S. C. (2004). *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions.* Hove, New York: Psychology Press.

Davidson, M. L. (1972). Univariate versus multivariate tests in repeated-measures experiments. *Psychol. Bull.* 77, 446–452.

de Vega, M., and Rodrigo, M. J. (2001). Updating spatial layouts mediated by pointing and labelling under physical and imaginary rotation. *Eur. J. Cogn. Psychol.* 13, 369–393.

Flavell, J. H., Green, F. L., and Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monogr. Soc. Res. Child Dev.* 51, i–v, 1–87.

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694–724.

Frith, C. D., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.

Grabherr, L., Karmali, F., Bach, S., Indermaur, K., Metzler, S., and Mast, F. W. (2007). Mental own-body and body-part transformations in microgravity. *J. Vestib. Res.* 17, 279–287.

Grabowski, J. (1999). A uniform anthropomorphological approach to the human conception of dimensional relations. *Spat. Cogn. Comput.* 1, 349–363.

Grabowski, J., and Miller, G. A. (2000). Factors affecting the use of dimensional prepositions in German and American English: object orientation, social context, and prepositional pattern. *J. Psycholinguist. Res.* 29, 517–553.

Graf, R. (1994). *Selbstrotation und Raumreferenz: Zur Psychologie partnerbezogenen Lokalisierens (Self-rotation and Spatial Reference: The Psychology of Partner-Centred Localisations)*. Frankfurt: Peter Lang.

Grosbras, M. N., Laird, A. R., and Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Hum. Brain Mapp.* 25, 140–154.

Gzesh, S. M., and Surber, C. F. (1985). Visual perspective-taking skills in children. *Child Dev.* 56, 1204–1213.

Hamilton, A. F. D., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Hegarty, M., and Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 175–191.

Herrmann, T., Bürkle, B., and Nirmaier, H. (1987). Zur hörerbezogenen Raumreferenz: Hörerposition und Lokalisationsaufwand (Listener-oriented spatial reference: Listener position and localisation effort). *Sprache Kogn.* 6, 126–137.

Huttenlocher, J., and Presson, C. C. (1973). Mental rotation and the perspective problem. *Cogn. Psychol.* 4, 277–299.

Keehner, M., Guerin, S. A., Miller, M. B., Turk, D. J., and Hegarty, M. (2006). Modulation of neural activity by angle of rotation during imagined spatial transformations. *Neuroimage* 33, 391–398.

Kessler, K. (2000). *Spatial Cognition and Verbal Localisations: A Connectionist Model for the Interpretation of Spatial Prepositions.* Wiesbaden: Deutscher Universitäts-Verlag.

Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88.

Kozhevnikov, M., and Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* 29, 745–756.

Kozhevnikov, M., Motes, M. A., Rasch, B., and Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Appl. Cogn. Psychol.* 20, 397–417.

Leslie, A. M., and Frith, U. (1988). Autistic childrens understanding of seeing, knowing and believing. *Br. J. Dev. Psychol.* 6, 315–324.

Levelt, W. J. M. (1996). "Perspective taking and ellipsis in spatial descriptions," in *Language and Space*, eds. P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garret (Cambridge, MA: A Bradford Book), 77–108.

Levine, M., Jankovic, I. N., and Palij, M. (1982). Principles of spatial problem solving. *J. Exp. Psychol. Gen.* 111, 157–175.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Michelon, P., and Zacks, J. M. (2006). Two kinds of visual perspective taking. *Percept. Psychophys.* 68, 327–337.

Myachykov, A., Platenburg, W. P. A., and Fischer, M. H. (2009). Non-abstractness as mental simulation in the representation of number. *Behav. Brain Sci.* 32, 343–344.

Obrien, R. G., and Kaiser, M. K. (1985). Manova method for analyzing repeated measures designs – an extensive primer. *Psychol. Bull.* 97, 316–333.

Pack, A. A., and Herman, L. M. (2006). Dolphin social cognition and joint attention: our current understanding. *Aquat. Mamm.* 32, 443–460.

Retz-Schmidt, G. (1988). Various views on spatial prepositions. *AI Mag.* 9, 95–105.

Samson, D., Apperly, I. A., Kathirgamanathan, U., and Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain* 128(Pt 5), 1102–1111.

Scheumann, M., and Call, J. (2004). The use of experimenter-given cues by South African fur seals (*Arctocephalus pusillus*). *Anim. Cogn.* 7, 224–230.

Schwabe, L., Lenggenhager, B., and Blanke, O. (2009). The timing of temporoparietal and frontal activations during mental own body transformations from different visuospatial perspectives. *Hum. Brain Mapp.* 30, 1801–1812.

Tomasello, M., Call, J., and Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Anim. Behav.* 55, 1063–1069.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735.

Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129.

Ungerleider, L. G., and Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165.

Vasey, M. W., and Thayer, J. F. (1987). The continuing problem of false positives in repeated measures anova in psychophysiology – a multivariate solution. *Psychophysiology* 24, 479–486.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Wraga, M., Shephard, J. M., Church, J. A., Inati, S., and Kosslyn, S. M. (2005). Imagined rotations of self versus objects: an fMRI study. *Neuropsychologia* 43, 1351–1361.

Zacks, J. M., and Michelon, P. (2005). Transformations of visuospatial images. *Behav. Cogn. Neurosci. Rev.* 4, 96–118.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 September 2010; accepted: 10 November 2010; published online: 06 December 2010.

Citation: Kessler K and Rutherford H (2010) The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. Front. Psychology ***1****:213. doi: 10.3389/fpsyg.2010.00213*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Kessler and Rutherford. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

PERSPECTIVE ARTICLE
published: 27 December 2010
doi: 10.3389/fpsyg.2010.00234

Embodied language comprehension requires an enactivist paradigm of cognition

Michiel van Elk[1,2]*, Marc Slors[3] and Harold Bekkering[1]

[1] *Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands*
[2] *Laboratory of Cognitive Neuroscience, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[3] *Department of Philosophy, Radboud University Nijmegen, Nijmegen, Netherlands*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Martin H. Fischer, University of Dundee, UK
Guy Dove, University of Louisville, USA

****Correspondence:***
Michiel van Elk, Laboratory of Cognitive Neuroscience, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Station 19, AI 2101, 1015 Lausanne, Switzerland.
e-mail: michiel.vanelk@epfl.ch

Two recurrent concerns in discussions on an embodied view of cognition are the "necessity question" (i.e., is activation in modality-specific brain areas necessary for language comprehension?) and the "simulation constraint" (i.e., how do we understand language for which we lack the relevant experiences?). In the present paper we argue that the criticisms encountered by the embodied approach hinge on a cognitivist interpretation of embodiment. We argue that the data relating sensorimotor activation to language comprehension can best be interpreted as supporting a non-representationalist, enactivist model of language comprehension, according to which language comprehension can be described as procedural knowledge – knowledge how, not knowledge that – that enables us to interact with others in a shared physical world. The enactivist view implies that the activation of modality-specific brain areas during language processing reflects the employment of sensorimotor skills and that language comprehension is a context-bound phenomenon. Importantly, an enactivist view provides an embodied approach of language, while avoiding the problems encountered by a cognitivist interpretation of embodiment.

Keywords: embodied cognition, enactivism, simulation, language comprehension

INTRODUCTION

One of the most exciting discoveries in cognitive neuroscience over the last decades is certainly the finding that our brain "resonates" to certain classes of stimuli. Observing the actions of others, for instance, activates brain areas comparable to the areas that are activated when one would perform these actions oneself (Rizzolatti et al., 1996; Calvo-Merino et al., 2006; Grafton, 2009). But neural resonance is not limited to action perception and social interaction. In this paper we will focus on neural resonance in language comprehension. Following the notion of communicative motor resonance during speech perception (the motor theory of speech perception, for overview, see Galantucci et al., 2006), several studies have shown that reading verbs referring to concrete action results in the recruitment of effector-specific regions of primary motor and premotor cortex, comparable to the activation observed when moving the effector most strongly associated with these actions (Hauk and Pulvermuller, 2004; Pulvermuller et al., 2005; Aziz-Zadeh et al., 2006; Boulenger et al., 2009).

The theoretical significance of these findings is a matter of ongoing debate. Those who incline toward an embodied approach to cognition claim that resonance mechanisms support language comprehension by providing an internal representation of described actions or events (Barsalou, 1999; Glenberg, 2010). Parallel to the simulationist interpretation of neural resonance in social cognition (Gallese and Goldman, 1998; Rizzolatti et al., 2001; Goldman, 2006) such representations may consist in the re-enactment of these actions (Zwaan and Madden, 2005; Fischer and Zwaan, 2008; Zwaan, 2009). However, just like some have pointed out that resonance mechanisms are neither necessary nor sufficient for action understanding (Jacob and Jeannerod, 2005; Jacob, 2009) and that neural resonance does not amount to simulation (Gallagher, 2007; Zahavi, 2008), skeptics of an embodied approach to language comprehension argue that neural resonance is neither necessary nor sufficient for language comprehension (Bedny et al., 2008; Mahon and Caramazza, 2008; Postle et al., 2008; Toni et al., 2008; Kemmerer and Gonzalez-Castillo, 2010). In addition, it is unclear how an embodied approach to cognition can account for the understanding of concepts or actions that we have never experienced ourselves.

We state that these problems for an embodied approach to language comprehension hinge on a cognitivist, representationalist understanding of embodied cognition. Tackling them requires switching to an enactivist paradigm of cognition. Cognitivism is here defined as the theoretical approach that attempts to explain cognition in terms of the manipulation of discrete internal representations. Although on this account cognition may be used for the purpose of guiding actions, the cognitive process *as such* is thought of in terms that do not essentially involve the actions that it may help to guide. Cognition understood as the manipulation of internal representations is supposed to mediate between perception and action. But although it is enabled by perception and used for action, neither perception, nor action is constitutive of it. By contrast, enactivism can be defined as the view that cognition emerges in the interaction between an organism and the environment, such that perception and action are co-constitutive of it. Cognition is manifested in the kind of appropriate, dynamic perception–action coupling that allows us to cope effectively with our physical and social environment. On the enactivist view it is misleading to think of such coupling as requiring discrete

representations of one's environment: effectively dealing with one's environment does not *presuppose* awareness of features of one's environment, rather it *reflects* such awareness. Enactivism implies that cognition is essentially tied to action and that cognition is always context-bound.

In the present paper, drawing a parallel with the recent enactivist criticism of the simulation interpretation of the mirror neuron system, we argue that the data relating sensorimotor activation to language comprehension can best be interpreted as supporting a non-representationalist, enactivist model of language comprehension. We will start by outlining evidence from cognitive neuroscience in favor of an embodied approach to language comprehension and by elaborating on the problems such an approach faces. We will then briefly turn to a parallel with the debate on the function of the mirror neuron system and highlight the recent enactivist move made in that debate. After having contrasted the enactivist cognition paradigm with the current cognitivist paradigm in cognitive neuroscience, we will introduce a similar move in the present context and discuss how an enactivist approach to embodied language comprehension can deal with the objections to an embodied approach to language comprehension. We will conclude with a discussion on the perspectives and limitations of the enactive approach to language and discuss the prospective for future research on language and embodiment.

EMBODIMENT AND LANGUAGE COMPREHENSION

What exactly does it mean to say that "cognition is embodied"? In cognitive neuroscience, theories of embodied cognition often seem to imply that cognition is embodied, because it recruits neural resources comparable to those used in perception and action. For instance, according to Barsalou's (1999) perceptual symbol systems, concepts have a perceptual basis and the recruitment of concepts involves the re-enactment of perceptual experiences in sensorimotor areas of the brain (see also Prinz, 2002). Similarly, in social cognition it has been argued that action understanding employs a process of motor simulation, involving brain structures comparable to those involved when one would perform the observed action oneself (Gallese and Goldman, 1998; Rizzolatti et al., 2001; Goldman, 2006). Central to these embodied theories is the idea that cognition is grounded in relevant perceptual or motor simulations.

The notion that the processing of concepts is accompanied by activation in modality-specific brain areas is supported by a number of studies (for review, see Fischer and Zwaan, 2008; Borghi and Cimatti, 2010). Behavioral studies support the idea that semantic processing recruits modality-specific resources. In a property verification study, for instance, participants were found to respond faster if the preceding word represented properties from the same instead of a different modality (e.g., gustatory, auditory, or visual; Pecher et al., 2003). In addition, the recruitment of modality-specific systems is found to extend beyond the word level, to the processing of semantics at the sentence-level. For instance, after reading sentences describing an agent viewing an object in a specific context, participants were faster in identifying pictures that were congruent with the situation described in the sentence (e.g., faster verification of visually degraded pictures when the context refers to fog; Yaxley and Zwaan, 2007). Neuroimaging studies have provided further support for the idea that the processing of language is accompanied by the activation of modality-specific brain regions. For instance, reading action verbs or sentences describing action-related language consistently results in the activation of motor-related brain areas (Hauk and Pulvermuller, 2004; Pulvermuller et al., 2005; Aziz-Zadeh et al., 2006; Boulenger et al., 2009), whereas the reading of words referring to concrete semantic concepts (e.g., animals or fruits) has been associated with increased activation in visual areas (Martin et al., 1996; van Schie et al., 2003). In sum, these studies support the assumption that the processing of linguistic concepts recruits modality-specific brain areas and on the basis of these findings it has been suggested that concepts are represented in brain areas comparable to those used for perception and action (Barsalou, 2008).

Two recurrent problems for the embodied approach to language are the "necessity question" and the "simulation constraint." The "necessity question" is the question whether activation in modality-specific brain areas during language processing is necessary for language comprehension or whether it should be considered as an epiphenomenon (e.g., post-lexical simulation). In other words: if language is truly grounded in sensorimotor areas, we should expect language processing to break down if the activation of sensorimotor areas is disrupted. Unfortunately, the data from studies with patients showing category-specific deficits in association with damage to sensorimotor areas is still inconclusive (for review, see Mahon and Caramazza, 2009). Similarly, data from TMS studies that have attempted to disrupt processing in motor areas during the reading of action verbs has provided only mixed results, with some studies showing early effects in the motor system during reading action-related sentences (Oliveri et al., 2004; Buccino et al., 2005), while other studies only observed effects during the later stages of word processing or during explicit motor imagery (Tomasino et al., 2008; Papeo et al., 2009). In addition, although some studies have reported somatotopic-specific effects in the motor system when reading verbs referring to specific effectors (Hauk and Pulvermuller, 2004; Tettamanti et al., 2005; Aziz-Zadeh et al., 2006; Boulenger et al., 2009), other studies have pointed out that a strict overlap between the activation of motor areas during action execution and the reading of actions verbs has never been directly demonstrated (Postle et al., 2008; Kemmerer and Gonzalez-Castillo, 2010). Thus, although there is some evidence for the involvement of motor resonance in language comprehension, at present it is unclear how and when exactly activation in motor areas supports language comprehension.

The "simulation constraint" poses another, more principled problem for embodied approaches to language. If language comprehension involves the re-enactment of our own sensorimotor experiences, it remains unclear how we can understand language for which we lack the relevant simulations. For instance, how do we understand actions that are beyond our own motor repertoire, such as animal actions or how do we understand language that is unrelated to the concrete sensorimotor domain, such as abstract words like "love," "war," or "justice"? Although several attempts have been made to provide an embodied account of the representation of abstract concepts (Barsalou, 1999; Glenberg et al., 2008; Glenberg, 2010), most research supporting the embodied approach of language has focused selectively on the processing of language referring to concrete actions or objects (Hauk and Pulvermuller, 2004; Pulvermuller et al., 2005; Aziz-Zadeh et al., 2006; Boulenger et al., 2009).

Our aim in the remainder of this paper is to sketch a way of giving up on the simulation constraint, while retaining an embodied approach to language comprehension. In addition, we will speculate on the consequences for the necessity question.

INTERLUDE: A PARALLEL WITH THE MIRROR NEURON DEBATE

In order to see how we can reject the simulation constraint while retaining an embodied approach to language comprehension, it is helpful to look at recent developments in an adjacent debate, the debate in social cognition on the function of the mirror neuron system. Mirror neuron activity has often been interpreted as representing simulations of perceived goal-directed actions for the purpose of grasping the intentions and emotions "behind" those actions (Gallese and Goldman, 1998; Gallese and Lakoff, 2005; Goldman, 2006; Gallese, 2007). The simulation interpretation of mirror neurons is controversial. One line of criticism is put forward by critics of embodied approaches to social cognition. It is argued that mirror neuron based simulation is at best sufficient to retrace motor intentions, while attribution of higher-level intentions (so-called "prior intentions"; Searle, 1983) requires much more elaborate cognitive activity (Jacob and Jeannerod, 2005; Saxe, 2005, 2009; Jacob, 2008). The point is that one type of movement may be recruited to carry out various higher-level intentions. It is also argued that mirror neuron activity is not necessary for the attribution of intentions. People attribute intentions, for instance, to moving geometric shapes in the famous Heider and Simmel (1944) movies and it is difficult to imagine how body-specific motor simulations could underlie this intention attribution (see also Castelli et al., 2002).

These arguments are intended to downplay the role of neural resonance in social cognition and hence to oppose embodied approaches to social cognition. However, they can also be taken seriously without abandoning an embodied view. Recently a number of philosophers have argued that mirror neurons may be part of larger neural processes underlying social perception, i.e., the direct pick-up of basic intentions and emotions in the conduct of other people (Gallagher, 2007; Gallagher and Zahavi, 2008; Zahavi, 2008; Hutto, 2009). Mirror neurons, according to these philosophers, need not be interpreted as coding for the re-enactment of the initiating stages of the other's action. Rather, they should be interpreted as contributing to the processing of the perceived behavior of others for the direct purpose of social *interaction*. The idea here is to think of social perception as an enactive process involving sensorimotor skills and not as mere sensory input processing. This idea is borrowed from enactive theories of perception according to which perception involves active engagement with the world rather than mere passive reception of information from the environment (cf. Hurley, 1998; Noë, 2004). The enactivist interpretation of neural resonance in social cognition fits well with the fact that many mirror neurons are broadly congruent to an observed action, rather than strictly congruent (Fogassi and Gallese, 2002; Csibra, 2005) and with the finding that mirror neurons fire during cooperative tasks in which one's own movements need to be complementary rather than imitative relative to the actions of the person one needs to cooperate with (Newman-Norlund et al., 2007). Thus according to an enactivist account, rather than reflecting a simulation process involving the mapping of observed actions onto one's own motor system, mirror neuron activation should be conceived of as reflecting the employment of sensorimotor skills. More specifically, activation of mirror neurons should be considered an integral part of the process of perceiving and responding to other's actions. In some cases this may require a covert response (e.g., perceiving other's action goals), in other cases a more overt reaction may be required (e.g., catching a team player's ball). What these cases have in common and what is a central notion of the enactive paradigm is that perceiving is an active process (Noë, 2004).

What is interesting about this recent enactivist move, in the context of our present discussion, is the fact that a simulationist interpretation of the function of mirror neurons is rejected (see, however, Slors, 2009) while their contribution to social cognition is still viewed from an embodied perspective. In order to see whether a similar move can be made with respect to resonance phenomena in language comprehension, we need to turn to the dominant cognitivism in current embodied approaches to language comprehension and the possible enactivist alternative.

COGNITIVISM VERSUS ENACTIVISM

In philosophy, embodied cognition is usually conceived of as an alternative for cognitivism, where "cognitivism" stands for an approach to cognition in terms of the rule- or algorithm-based manipulation of discrete internal representations of the world (Brooks, 1991; Clark, 1991, 1997; Gallagher, 2005; Gibbs, 2005; Rowlands, 2006; Chemero, 2009). In cognitive neuroscience, however, embodied approaches to cognition are in an important sense still fully cognitivist. Their main quarrel with traditional approaches to cognition is not about whether cognition should be thought of in terms of representations, but about how we should think of these representations. Contrary to traditional cognitivism, the embodied approach argues that the vehicle used for representing concepts is sensorimotor in nature (cf. Barsalou, 1999; Zwaan and Madden, 2005; Fischer and Zwaan, 2008; Mahon and Caramazza, 2008; Glenberg, 2010). In cognitive neuroscience the notion that concepts are embodied primarily means that there is a correspondence between the brain activations associated with processing the referent of a concept and the processing of the concept itself. For instance, seeing a car and thinking or reading about a car involves the activation in comparable visual areas. Thus, the dispute between modal and amodal theories of language comprehension is basically a discussion about the representational vehicle of concepts (i.e., whether the representational vehicle of concepts is shared with neural resources used for perception and action). Both modal and amodal theories of language thus share a cognitivist notion of cognition in terms of discrete internal representations of the world.

This often applied representationalist notion of embodiment in cognitive neuroscience implies an important break from philosophical approaches to embodied cognition, which emphasize that cognition should be understood in terms of the dynamical interaction between an organism and its environment (Varela et al., 1991; Hurley, 1998; O'Regan and Noe, 2001; Noë, 2004; Gallagher, 2005; Thompson, 2007; Chemero, 2009). We refer to these diverse approaches as "enactivist." A defining feature of the enactivist paradigm of cognition is that it challenges the representationalism of the traditional cognitivist paradigm by taking cognition to be based on "knowing *how*" instead of "knowing *that*." That is, an

organism's knowledge of its environment is not taken to consist in the adequate representation or internal modeling of environmental features. Rather, knowledge consists in the way sensory information is linked to motor output. The structuring and restructuring of sensorimotor links in the recursive interaction of an organism with its environment, by means of which the organism adapts to it, implies or specifies knowledge of the world. Thus, in the enactivist paradigm, the fact that knowledge is essentially embodied and embedded involves its being non-representational (see, however, Hutto, 2005). Knowledge – cognition – as the American naturalist Dewey (1896) pointed out, cannot be understood by breaking it into parts; it always exists at the level of the situated organism as a whole (Ryle, 1949; Dennett, 1969). With its roots in Gibsonian ecological psychology (Gibson, 1979) an important branch of enactivism focuses on a non-representationalist account of perception based on so-called "sensorimotor contingencies" (Hurley, 1998; O'Regan and Noe, 2001; Noë, 2004). There are interesting connections here with earlier developments in Robotics. Brooks (1991), for instance, showed that robots without a central processor or an internal map of the environment can successfully move around due to independent "perception–action modules" that act directly on the incoming information. These approaches to cognition essentially highlight the direct coupling between perception and action, without invoking representations as an explanatory variable. Thus, the enactivist view rejects the notion of "shared representations" between language processing and sensorimotor processing.

Another branch of enactivism focuses on the continuity between mind and life by arguing that living is itself a cognitive process. A living being creates and maintains its own domain of meaningfulness by generating and maintaining its own self-identity as an embodied organism (Thompson, 2007). Again, the embodiment of cognition is taken to imply a non-representationalist notion of cognition. The mind is not seen as a complex system of cognitive cogs and levers, but rather as unified whole, an organism, whose cognitive feats can be described in terms of the non-linear dynamics of dynamic systems theory (Varela et al., 1991; see for applications in cognitive neuroscience: Thelen, 1994; Beer, 2000). Dynamical systems theory provides a model of cognition that consists of "a set of quantifiable variables changing continually, concurrently and interdependently over time in accordance with dynamical laws that can, in principle, be described by some set of equations" (Chemero, 2009). Initially, dynamical systems theory was applied to model relatively simple motor behaviors, such as walking (Thelen, 1994), finger wagging (Haken et al., 1985; Schoner and Kelso, 1988), or the social coupling of motor behavior (Schmidt et al., 1990; Richardson et al., 2007). In addition, dynamical approaches have been applied to model higher-level cognition as well, such as the A-not-B error (Thelen et al., 2001), categorical perception (Beer, 2000) and mathematical problem solving (Stephen et al., 2009a,b). Central to dynamical models is the assumption that seemingly complex behavior can be accurately described with relatively simple mathematical models, such as coupled oscillators or dynamic fields.

Several authors have argued for an approach to language comprehension that fits the enactive paradigm of cognition, broadly conceived (Barwise and Perry, 1981, 1983; Clark, 2006; Beckner et al., 2009) or have applied dynamical systems modeling to language perception (Pollack, 1991; Port et al., 1995; Port, 2003), and production (Elman, 1990; Port, 2003). Surprisingly, these approaches are largely ignored in recent discussions on the embodiment of language in cognitive neuroscience. Vice versa, recent findings in cognitive neuroscience showing the involvement of modality-specific brain areas during language processing have hardly been incorporated by enactivists or in dynamical models. This lack of cross talk is probably related to the incommensurable paradigms in the respective fields of research. Embodied cognition in cognitive neuroscience uses the cognitivist paradigm and has thus been concerned primarily with explaining how meaning is represented in the brain (Barsalou, 1999; Fischer and Zwaan, 2008; Mahon and Caramazza, 2008; Zwaan, 2009). By contrast, approaches to language that fit the enactivist paradigm are typically anti-representationalist and focus primarily on those aspects of language that allow for a dynamical explanation, such as speech rhythms (Port, 2003), syntax (Elman, 1990, 1995) or the functioning of language at an inter-individual level (Clark, 2006; Beckner et al., 2009).

AN ENACTIVIST APPROACH TO LANGUAGE COMPREHENSION

In this section we shall briefly sketch the contours of an enactivist conception of language comprehension. We will then argue that this conception fits the neuroscientific data on embodiment and language better than a cognitivist embodied cognition approach in terms of modal representations and motor simulations. Finally and most importantly, we will argue that the enactivist conception of language comprehension provides an embodied approach to language comprehension that avoids the necessity question and the simulation constraint.

An enactivist approach to language comprehension implies that language, ultimately, is used for action and social interaction. This means that linguistic utterances acquire their meaning in context and not merely as a function of syntax and semantics. When you are sitting in a restaurant and your partner asks you "Can you give me the salt?" you do not reply by saying "yes," although that would be the correct answer if syntax and semantics were all that matters. The speech act of your partner directs you to perform a certain action (Searle, 1969). Instead of asking for the salt, your partner could have pointed toward the salt as well to make the same request. Or suppose you are sitting in the restaurant again and the waiter asks you whether you would like anything for desert. You respond by saying that you are fine and that you would like to pay the bill. In this case, your response to the waiter's request follows a linguistic convention in a script-like fashion (Schank and Abelson, 1977). In both examples, language comprehension can be accurately described as the procedural knowledge how to respond in certain situations to specific utterances. On the enactivist account this notion of language comprehension is paradigmatic; it can be extended to cover many or even most instances of language comprehension. Learning to understand language is learning how to couple specific linguistic inputs to specific actions. These actions may be immediate but they may also be in the more distant future (e.g., as in understanding the sentence "the election will be on May the 5th"). They may also be only "virtual" in the sense that understanding an utterance only involves being disposed to act in certain ways given certain circumstances. Of course in many instances responding appropriately to an utterance is responding linguistically. But linguistic practice is not free floating – it is a practice of embodied beings in

a physical world. As Wittgenstein (1953) held, understanding the meaning of a word is knowing how it can be used. And this use always takes place within a social context involving the pragmatics of interacting embodied persons. In short, on an enactivist account, language comprehension can be described as procedural knowledge – knowledge *how*, not knowledge *that* – that enables us to interact with others in a shared physical world.

The enactivist view implies that language comprehension should be studied in relation to its potential for action. Thus, the brain activations associated with language processing do not mirror a representation-based inference process. In contrast, the activation of modality-specific brain areas during language processing should be conceived of as reflecting the employment of sensorimotor skills. On this account, the motor activation that has been found in association with the processing of action verbs or words referring to manipulable objects likely supports action prediction or anticipation. For instance, in the sentence "Can you *give* me the salt?" the motor activation in relation to the processing of the word "give" may prepare the listener for a subsequent grasping action (Zwaan and Kaschak, 2009) and the motor activation in association with the processing of words like "cup," "scissors," or "hammer" may reflect the retrieval of conceptual knowledge to enable the subsequent (virtual) interaction with the object (cf. van Elk et al., 2009a; Rueschemeyer et al., 2010). Similarly, perceptual resonance during language processing may reflect a pattern completion inference process used for prediction (see also Barsalou, 2009). For instance, the activation in visual areas that accompanies the processing of words referring to concrete concepts may support the categorical perception of behaviorally relevant categories (Ward, 2009), such as in "look, there's an eagle up in the sky" (Zwaan et al., 2002) or may facilitate the retrieval of relevant contextual information that allows one to make inferences and predictions about objects and situations (Barsalou et al., 2003; Barsalou, 2009).

An advantage of an enactive approach is that it allows for the fact that language comprehension is a context-bound phenomenon that is dependent on the relation between the organism and the context in which the organism is acting. Cognitivist embodied approaches often make the implicit assumption that there is a core meaning of words that can be specified in terms of a specific representational vehicle. More specifically, cognitivist embodied approaches to language processing seem to imply that the sensorimotor representations that are activated in association with the processing of words occur relatively fast, automatic, and in a bottom-up fashion (Pulvermuller, 2005). The idea is that word reading results in the spreading of activation throughout a network of associated sensorimotor features, thereby constituting the meaning of the word. However, in one context, the motor activation associated with the processing of, e.g., the word "pass" may specify a specific action tendency, such as with the speech act "please pass me the salt," whereas, in another context a different motor activation will be involved, such as in the utterance "pass me the ball" in a soccer game. In line with the idea that meaning is context-bound, recent studies indicate that the sensorimotor features that are co-activated in association with the processing of words are indeed dependent on the context in which the word is presented (Hoenig et al., 2008; van Dam et al., 2010). For instance, the word "tennis ball" primarily activates visual features when presented in a visual context, whereas motor features are more strongly activated when the word is presented in an action context. Similarly, in another study we found that a word's long-term semantic associations can be selectively overruled when the word is used in a different context (van Elk et al., 2009b). For instance, whereas the concept "cup" is strongly associated to the word "mouth," this semantic association can be overruled if one intends to use the cup in an unusual fashion (e.g., bring the cup toward the eye), thereby underlining the flexibility and context-dependence of language use. Moreover, these findings argue against a cognitivist interpretation of embodiment, according to which sensorimotor activation during language processing reflects the activation of representations, specifying the core meaning of concrete words. Thereby the enactive paradigm to language differs in important ways from previous theories that have argued that language is primarily for action (Glenberg, 1997; see also Borghi and Cimatti, 2010), but that still maintained the notion of internal simulation processes underlying language understanding. As pointed out, these approaches run into the simulation constraint and the necessity question that the enactivist paradigm tries to avoid, by avoiding the notion of internal simulations.

Another important advantage of an enactivist approach to embodied language comprehension over a cognitivist approach is that it accounts for a broad range of action-related effects during language processing that need not be restricted to simulation, re-enactment, or pre-enactment. Thus it can accommodate findings that are harder to interpret in cognitivist terms. For instance, in a recent study we found a stronger motor resonance for verbs describing animal actions compared to human actions (van Elk et al., 2010). If motor resonance is primarily related to the familiarity of the action, we should have expected a stronger motor activation for human actions, as the way in which most animals move is clearly different from the way in which humans move. In contrast, animals only have a very limited action repertoire (e.g., a duck can "swim," "squeak," or "fly"), whereas humans can perform many different actions. Accordingly, actions are easier to predict for animals than for humans and the stronger motor resonance for animal actions fits well with the idea that motor resonance is used for action prediction (van Elk et al., 2010). In another study it was found that making a lexical decision about verbs and imagining the actions described by these verbs are two neurally dissociable processes, involving activation in different regions of premotor cortex (Willems et al., 2010). This finding argues against a strict simulationist interpretation of motor resonance as well, but goes well with the enactivist view: making a decision about whether a string of letters represents a word and imaging the action described by a word are two different skills that involve different regions of premotor cortex. These studies underline the importance of sensorimotor activation for language processing but they cannot be accounted for merely in terms of simulation or re-enactment.

In an enactivist account of language comprehension, the simulation constraint mentioned in Section "Embodiment and Language Comprehension" is absent. That is, on an enactivist account the idea that language comprehension is embodied is not exhausted by the idea that the processing of action words involves re-enactment or pre-enactment of the described action. Thus, the fact that many instances of language comprehension are hard to conceive of in terms of simulation – either because utterances involve actions

that are beyond our own motor repertoire, or because they are unrelated to the concrete sensorimotor domain – need not be an impediment to an embodied account of them. Hence, switching from a cognitivist to an enactivist paradigm of cognition effectively deals with an important objection to an embodied approach to language comprehension.

In the case of the necessity question a solution can be conceived of along similar lines. The necessity question starts from the implicit assumption that a core meaning of words can be specified and the critical issue is whether this core meaning is instantiated in sensorimotor areas. However, as indicated above, according to an enactivist view, language comprehension consists in the context-bound employment of sensorimotor skills, rather than in the search for cognitivist representations. On an enactivist account, blocking activation of motor or premotor areas associated with the specific action mentioned in an utterance thus need not impede understanding when language comprehension is subserved by sensorimotor activation other than re-enactment or pre-enactment.

It is important to stress that the emphasis on context sensitivity is not intended to simply replace standard accounts of context sensitive language understanding, such as Grice's theory of conversational implicatures (Grice, 1989). Grice's account of context sensitive language use identifies principles and maxims that describe the various ways in which context is taken into account when uttering and understanding sentences. Cognitivist approaches to language understanding, specifically those of a non-embodied kind, typically take these principles and maxims to be implemented in our cognitive architecture. But nothing in Grice's theory implies this. On our view, we should be very careful in distinguishing levels of description here (cf. Dennett, 1969; Bennet and Hacker, 2003) and resist the tendency to explain personal-level cognitive phenomena in terms of isomorphic brain-level processes. On an enactive view on context sensitive language understanding, Grice's principles and maxims that describe conversational implicatures *model* real life linguistic interaction. That is, such interactions are not *governed* by these principles and maxims. Rather they are informatively *described* by them, possibly in a slightly idealized way.

A related issue is the question whether a context needs to be represented, in order to be effective. The question "what would you like for dinner?" has different implications for action and is hence understood differently when sitting in a restaurant and when walking in a supermarket with your partner. When a given context is taken into account in understanding a sentence, can it not be said that the hearer somehow represents this context? On an enactive view, taking a context into account means that the context is relevant to specific perception–action couplings. In an innocent but uninformative way, this means that if one responds appropriately in a given situation, the context is represented accurately. But that does not mean, enactivists would stress, that such couplings are co-determined by a discrete representation of the context that is causally operative in bringing about one type of coupling rather than another. In fact, the situation itself is already part of the enactive process of perceiving and acting in the world, and thus there is no need to suppose the representation of the context. Context sensitivity, then, need not imply the kind of representationalism that is characteristic of cognitivism.

SCOPE AND LIMITATIONS OF THE ENACTIVE APPROACH

An important question is to what extent the enactive paradigm can scale up the requirements of a full-blown theory of high-level cognition. It has repeatedly been argued that the enactive model works relatively well when it comes to explaining lower level sensorimotor processes, but that so-called "representation-hungry problems" (Clark, 1997) are more difficult to explain within an enactive framework (see, however, van Rooij et al., 2002; Chemero, 2009).

In the present context, it is especially relevant to consider if and how the enactive paradigm can account for the processing of abstract words. First of all, we would like to point out that our aim was primarily to show how the enactive approach circumvents the problems associated with a cognitivist interpretation of sensorimotor activation during the processing of concrete words. When it comes to the processing of abstract concepts we have to be more careful. As discussed before, several attempts have been made to provide an embodied account of abstract word meaning (Barsalou, 1999; Lakoff and Johnson, 1999; Glenberg et al., 2008; Borghi and Cimatti, 2009; Glenberg, 2010. It has been repeatedly pointed out, for instance, that many abstract concepts bear a direct relation to the concrete domain, such as words referring to divine concepts (Meier et al., 2007), words describing power relations (Schubert, 2005) and even words referring to numbers (Dehaene et al., 1993). Typically, the relation between the abstract and the concrete domain is conceived as abstract words being linked to concrete sensorimotor representations. For instance, it has been argued that numbers are spatially represented in the brain along a "mental number line" (Moyer and Landauer, 1967; Dehaene et al., 1993). Although the enactive view acknowledges the idea that many abstract concepts are related to concrete sensorimotor experiences, it rejects the view that this relation can be identified at a brain level, in the form of specific neural representations. Similar to the enactive approach to concrete words, the processing of abstract words should always be considered in relation to its potential for action. For instance, the observed relation between number words and space may be part of a common magnitude system that is used both for perception and action (Walsh, 2003). In line with this suggestion, it has been found that number processing influences action planning, such that large numbers facilitate power grips and small numbers facilitate precision grips (Lindemann et al., 2007). Similarly, action compatibility effects in association with the processing of words with a positive or negative valence may reflect approach and avoidance tendencies (see for instance van Dantzig et al., 2008). In sum, although the enactive paradigm proposed in the present paper is primarily intended as an alternative to a cognitivist interpretation of neural resonance during language processing, one could conceive a similar approach in considering abstract concepts in relation to their potential for action (see for instance Borghi and Cimatti, 2009).

Another possible limitation of the enactive paradigm concerns the costs associated with abandoning sensorimotor simulations in language processing. It has been argued that perceptual symbols and sensorimotor simulations allow for the systematicity and productivity of thought (Barsalou et al., 2003). For instance, it has been argued that simulations allow one to make inferences beyond the information that is directly available. In addition, concepts can be combined into more complex concepts, via a selective process of merging existing simulations (e.g., Prinz, 2002).

Although an in-depth discussion of these concerns is beyond the scope of the present paper, we would like to point out that a cognitivist embodied account of systematicity and productivity runs into the same problems as mentioned before. With respect to the simulation constraint, it remains unclear how it is possible to make inferences about concepts for which we lack the relevant simulations. In addition, when it comes to conceptual combination it remains unclear how combined concepts are understood, whose sensorimotor properties cannot be inferred on the basis of their constituent concepts (e.g., a "wooden spoon" is typically conceived as big, whereas neither the concept "wood" or "spoon" implies this property).

IMPLICATIONS FOR FUTURE RESEARCH

In the final section of this paper the implications of an extended approach to language for future research will be discussed. As argued before, the enactive view can accommodate research findings that are difficult to reconcile with a simulationist interpretation of embodiment. In addition, the enactive view provides an important break from previous attempts aimed at determining the necessity of neural resonance for language understanding. Rather than focusing on the nature of linguistic representations, research should consider under what conditions and in which contexts language processing is accompanied by activation in modality-specific brain areas. We would like to suggest possible directions for future research on the functional role of neural resonance in language processing.

First, according to the enactive view language is primarily used for action and accordingly, motor activation in association with language processing should be considered in relation to its potential for action. In line with this suggestion, several studies have shown direct effects of language processing on motor performance (Boulenger et al., 2008; Nazir et al., 2008; Frak et al., 2010) or from action preparation on language processing (Lindemann et al., 2006; van Elk et al., 2008; van Elk et al., 2009b). Moreover, the enactive approach predicts that interactions between language and action are not restricted to relatively simple reaching and grasping movements, but extend to naturalistic action settings as well. One intriguing possibility would be to investigate the functional role of effects of language on action in a communicative setting for instance (e.g., such as when someone asks you to pass the salt across the table).

Second, the enactive view implies that the coupling between language and action is flexible and context-dependent. In contrast, embodied accounts of language processing have suggested that the coupling between language and action is obligatory and that the motor system is activated within the first few 100 ms after word onset (e.g., Pulvermuller et al., 2005). According to an enactive view, rather than being automatic, the activation of motor-related areas should be dependent on the context in which a word is presented. Thus, the word "pass" may be associated with the movement of different effectors, depending on the context. Similarly, whereas in some instances a word like "apple" may prime a power grip (Glover and Dixon, 2002; Glover et al., 2004), when presented in a different context it may prime a precision grip (e.g., after hearing a sentence like "when only the core was left, he threw away the apple").

Third, as indicated in the previous section, motor activation in relation to language processing may support action prediction or anticipation. Thus, motor activation during language processing may prepare the listener for subsequent actions, as in the sentence "please pass me the salt." Interestingly, studies on action observation suggest that violations of an expected action result in a stronger motor activation, likely reflecting the updating of a forward model (Koelewijn et al., 2008; Stapel et al., 2010). Similarly, if motor resonance in language processing is related to prediction we should expect a stronger motor activation if the actions described in a sentence do not match one's expectations. In sum, these examples illustrate that the enactivist view on language generates testable predictions that should be addressed more broadly in future research.

CONCLUSION

We conclude that an embodied approach to language comprehension in cognitive neuroscience requires an enactivist rather than a cognitivist conception of embodied cognition. An enactivist paradigm allows us to make sense of more of the cognitive neuroscientific data relating language comprehension to action effects or modality specific neural processing than a cognitivist paradigm by including sensorimotor activations that cannot be subsumed under the heading of (p)re-enactment. Also, the enactivist paradigm more easily allows for the context-dependence of language comprehension. Finally and most importantly, an enactivist conception allows us to answer two of the most serious objections to an embodied account of language comprehension, the necessity question and the simulation constraint. In conclusion, the multidisciplinary evidence relating language comprehension to sensorimotor activity, argues for an enactivist conception of language. Language comprehension reflects the employment of sensorimotor skills and is a context-bound phenomenon that is dependent on the relation between the organism and the context in which the organism is acting.

ACKNOWLEDGMENTS

This paper was written during fellowships supported by The International Human Frontier Science Program Organization (grant ST000078/2009 to Michiel van Elk) and the Marie Curie Intra European Fellowship within the Seventh European Community Framework Program (IEF grant 252713 to Michiel van Elk) and by the Netherlands Organization for Scientific Research (grant 453-05- 001 awarded to Harold Bekkering).

REFERENCES

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609; discussion 610–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans R Soc. Lond. B Biol. Sci.* 364, 1281–1289.

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91.

Barwise, J., and Perry, J. (1981). Situations and attitudes. *J. Philos.* 77, 668–691.

Barwise, J., and Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N.,

Holland, J., Ke, J., Larsen-Freeman, D., and Schoenemann, T. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59, 1–27.

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., and Saxe, R. (2008). Concepts are more than percepts: the case of action verbs. *J. Neurosci.* 28, 11347–11353.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends Cogn. Sci.* 4, 91–99.

Bennet, M. R., and Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Borghi, A. M., and Cimatti, F. (2009). "Words as tools and the problem of abstract word meaning," in *Paper Presented at the Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Amsterdam.

Borghi, A. M., and Cimatti, F. (2010). Embodied cognition and beyond: acting and sensing the body. *Neuropsychologia* 48, 763–773.

Boulenger, V., Hauk, O., and Pulvermuller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Boulenger, V., Silber, B. Y., Roy, A. C., Paulignan, Y., Jeannerod, M., and Nazir, T. A. (2008). Subliminal display of action words interferes with motor planning: a combined EEG and kinematic study. *J. Physiol. Paris* 102, 130–136.

Brooks, R. (1991). Intelligence without representation. *Artif. Intell.* 47, 139–159.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Brain Res. Cogn. Brain Res.* 24, 355–363.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* 16, 1905–1910.

Castelli, F., Frith, C., Happe, F., and Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125, 1839–1849.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Clark, A. (1991). An embodied cognitive science? *Trends Cogn. Sci.* 3, 345–351.

Clark, A. (1997). *Being There*. Cambridge, MA: MIT Press.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends Cogn. Sci.* 10, 370–374.

Csibra, G. (2005). Mirror neurons and action observation. Is simulation involved? *Interdisciplines*. Available at: http://www.interdisciplines.org/mirror/papers/4

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396.

Dennett, D. C. (1969). *Content and Consciousness*. London: Routledge.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychol. Rev.* 3, 357–370.

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211.

Elman, J. L. (1995). "Language as a dynamical system," in *Mind as Motion*, eds R. F. Port and T. van Gelder (Cambridge, MA: MIT Press), 195–225.

Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol.* 61, 825–850.

Fogassi, L., and Gallese, V. (2002). "The neural correlates of action understanding in non-human primates," in *Mirror Neurons and the Evolution of Brain and Language*, eds M. I. Stamenov and V. Gallese (Amsterdam: John Benjamins Publishers), 13–35.

Frak, V., Nazir, T., Goyette, M., Cohen, H., and Jeannerod, M. (2010). Grip force is part of the semantic representation of manual action verbs. *PLoS ONE* 5, e9728. doi: 10.1371/journal.pone.0009728

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377.

Gallagher, S. (2005). *How the Body Shapes the Mind*. New York: Oxford University Press.

Gallagher, S. (2007). Simulation trouble. *Soc. Neurosci.* 2, 353–365.

Gallagher, S., and Zahavi, D. (2008). *The Phenomenological Mind*. London: Routledge.

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 362, 659–669.

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 12, 493–501.

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479.

Gibbs, R. (2005). *Embodiment and Cognitive Science*. New York: Cambridge University Press.

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 586–596.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61, 905–919.

Glover, S., and Dixon, P. (2002). Semantics affect the planning but not control of grasping. *Exp. Brain Res.* 146, 383–387.

Glover, S., Rosenbaum, D. A., Graham, J., and Dixon, P. (2004). Grasping the meaning of words. *Exp. Brain Res.* 154, 103–108.

Goldman, A. I. (2006). *Simulating Minds. The Philosophy, Psychology and Neuroscience of Mindreading*. New York: Oxford University Press.

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N.Y. Acad. Sci.* 1156, 97–117.

Grice, P. (1989). *Studies in the Way of Words*. Harvard: Harvard University Press.

Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in humand hand movements. *Biol. Cybern.* 51, 347–356.

Hauk, O., and Pulvermuller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Hum. Brain Mapp.* 21, 191–201.

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259.

Hoenig, K., Sim, E. J., Bochev, V., Herrnberger, B., and Kiefer, M. (2008). Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *J. Cogn. Neurosci.* 20, 1799–1814.

Hurley, S. L. (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Hutto, D. D. (2005). Knowing what? Radical versus conservative enactivism. *Phenomenol. Cogn. Sci.* 4, 389–405.

Hutto, D. D. (2009). Interaction? Yes. But of what kind and on what basis? *Conscious. Cogn.* 18, 543–546.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind Lang.* 23, 190–223.

Jacob, P. (2009). The tuning-fork model of human social cognition: a critique. *Conscious. Cogn.* 18, 229–243.

Jacob, P., and Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends Cogn. Sci.* 9, 21–25.

Kemmerer, D., and Gonzalez-Castillo, J. (2010). The two-level theory of verb meaning: an approach to integrating the semantics of action with the mirror neuron system. *Brain Lang.* 112, 54–76.

Koelewijn, T., van Schie, H. T., Bekkering, H., Oostenveld, R., and Jensen, O. (2008). Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage* 40, 767–775.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Lindemann, O., Abolafia, J. M., Girardi, G., and Bekkering, H. (2007). Getting a grip on numbers: numerical magnitude priming in object grasping. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1400–1409.

Lindemann, O., Stenneken, P., van Schie, H. T., and Bekkering, H. (2006). Semantic activation in action planning. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 633–643.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Mahon, B. Z., and Caramazza, A. (2009). Concepts and categories: a cognitive neuropsychological perspective. *Annu. Rev. Psychol.* 60, 27–51.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature* 379, 649–652.

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., and Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *J. Pers. Soc. Psychol.* 93, 699–710.

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature* 215, 1519–1520.

Nazir, T. A., Boulenger, V., Roy, A., Silber, B., Jeannerod, M., and Paulignan, Y. (2008). Language-induced motor perturbations during the execution of a reaching movement. *Q. J. Exp. Psychol.* 61, 933–943.

Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M., and Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nat. Neurosci.* 10, 817–818.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

O'Regan, J. K., and Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973; discussion 973–1031.

Oliveri, M., Finocchiaro, C., Shapiro, K., Gangitano, M., Caramazza, A., and Pascual-Leone, A. (2004). All talk and no action: a transcranial magnetic stimulation study of motor cortex activation during action word production. *J. Cogn. Neurosci.* 16, 374–381.

Papeo, L., Vallesi, A., Isaja, A., and Rumiati, R. I. (2009). Effects of TMS on different stages of motor and non-motor verb processing in the primary motor cortex. *PLoS ONE* 4, e4508. doi: 10.1371/journal.pone.0004508

Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychol. Sci.* 14, 119–124.

Pollack, J. B. (1991). The induction of dynamical recognizers. *Mach. Learn.* 7, 227–252.

Port, R. F. (2003). Meter and speech. *J. Phon.* 31, 599–611.

Port, R. F., Cummins, F., and McAuley, J. D. (1995). "Naive time, temporal patterns and human audition," in *Mind as Motion*, eds R. F. Port and T. van Gelder (Cambridge, MA: MIT Press), 339–371.

Postle, N., McMahon, K. L., Ashton, R., Meredith, M., and de Zubicaray, G. I. (2008). Action word meaning representations in cytoarchitectonically defined primary and premotor cortices. *Neuroimage* 43, 634–644.

Prinz, J. (2002). *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge, MA: MIT Press.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.

Pulvermuller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797.

Pulvermuller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. C. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Move.Sci.* 26, 867–891.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.

Rowlands, M. (2006). *Body Language*. Cambridge, MA: MIT Press.

Rueschemeyer, S. A., Pfeiffer, C., and Bekkering, H. (2010). Body schematics: on the role of the body schema in embodied lexical-semantic representations. *Neuropsychologia* 48, 774–781.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson House.

Saxe, R. (2005). Against simulation: the argument from error. *Trends Cogn. Sci.* 9, 174–179.

Saxe, R. (2009). The neural evidence for simulation is weaker than I think you think it is. *Philos. Stud.* 144, 447–456.

Schank, R. C., and Abelson, R. J. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 227–247.

Schoner, G., and Kelso, J. A. (1988). Dynamic pattern generation in behavioral and neural systems. *Science* 239, 1513–1520.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *J. Pers. Soc. Psychol.* 89, 1–21.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Slors, M. V. P. (2009). Neural resonance: between implicit simulation and social perception. *Phenomenol. Cogn. Sci.* 9, 437–458.

Stapel, J. C., Hunnius, S., van Elk, M., and Bekkering, H. (2010). Motor activation during observation of unusual vs. ordinary actions in infancy. *Soc. Neurosci.* 5, 451–460.

Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., and Dixon, J. A. (2009a). The dynamics of insight: mathematical discovery as a phase transition. *Mem Cogn.* 37, 1132–1149.

Stephen, D. G., Dixon, J. A., and Isenhower, R. W. (2009b). Dynamics of representational change: entropy, action, and cognition. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1811–1832.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–281.

Thelen, E. (1994). Three-month old infants can learn task-specific patterns of interlimb coordination. *Psychol. Sci.* 5, 280–285.

Thelen, E., Schoner, G., Scheier, C., and Smith, L. B. (2001). The dynamics of embodiment: a field theory of infant perseverative reaching. *Behav. Brain Sci.* 24, 1–34; discussion 34–86.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Tomasino, B., Fink, G. R., Sparing, R., Dafotakis, M., and Weiss, P. H. (2008). Action verbs and the primary motor cortex: a comparative TMS study of silent reading, frequency judgments, and motor imagery. *Neuropsychologia* 46, 1915–1926.

Toni, I., de Lange, F. P., Noordzij, M. L., and Hagoort, P. (2008). Language beyond action. *J. Physiol. Paris* 102, 71–79.

van Dam, W. O., Rueschemeyer, S.-A., and Bekkering, H. (2010). Context effects in embodied lexical-semantic processing. *Front. Psychol.* 1:150. doi: 10.3389/fpsyg.2010.00150

van Dantzig, S., Pecher, D., and Zwaan, R. A. (2008). Approach and avoidance as action effects. *Q. J. Exp. Psychol.* 61, 1298–1306.

van Elk, M., van Schie, H. T., and Bekkering, H. (2008). Semantics in action: an electrophysiological study on the use of semantic knowledge for action. *J. Physiol. Paris* 102, 95–100.

van Elk, M., van Schie, H. T., and Bekkering, H. (2009a). Action semantic knowledge about objects is supported by functional motor activation. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1118–1128.

van Elk, M., van Schie, H. T., and Bekkering, H. (2009b). Short-term action intentions overrule long-term semantic knowledge. *Cognition* 111, 72–83.

van Elk, M., van Schie, H. T., Zwaan, R. A., and Bekkering, H. (2010). The functional role of motor activation in language processing: motor cortical oscillations support lexical-semantic retrieval. *Neuroimage* 50, 665–677.

van Rooij, I., Bongers, R. M., and Haselager, W. F. G. (2002). A non-representational approach to imagined action. *Cogn. Sci.* 26, 345–375.

van Schie, H. T., Wijers, A. A., Kellenbach, M. L., and Stowe, L. A. (2003). An event-related potential investigation of the relationship between semantic and perceptual levels of representation. *Brain Lang.* 86, 300–325.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488.

Ward, R. (2009). Representation in dynamical agents. *Neural Netw.* 22, 258–266.

Willems, R. M., Toni, I., Hagoort, P., and Casasanto, D. (2010). Neural dissociations between action verb understanding and motor imagery. *J. Cogn. Neurosci.* 22, 2387–2400.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Yaxley, R. H., and Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition* 105, 229–236.

Zahavi, D. (2008). Simulation, projection and empathy. *Conscious Cogn.* 17, 514–522.

Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *Eur. J. Soc. Psychol.* 39, 1142–1150.

Zwaan, R. A., and Kaschak, M. P. (2009). "Language comprehension as a means of 're-situating' oneself," in *The Cambridge Handbook of Situated Cognition*, eds M. P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 368–381.

Zwaan, R. A., and Madden, C. J. (2005). "Embodied sentence comprehension," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 224–245.

Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychol Sci.* 13, 168–171.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2010; accepted: 09 December 2010; published online: 27 December 2010.

Citation: van Elk M, Slors M and Bekkering H (2010) Embodied language comprehension requires an enactivist paradigm of cognition. Front. Psychology ***1****:234. doi: 10.3389/fpsyg.2010.00234*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 van Elk, Slors and Bekkering. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 25 January 2011
doi: 10.3389/fpsyg.2010.00242

On the need for embodied and dis-embodied cognition

Guy Dove*

Department of Philosophy, University of Louisville, Louisville, KY, USA

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Bradford Mahon, University of Rochester, USA
Jesse Prinz, City University of New York, USA

****Correspondence:***
Guy Dove, Department of Philosophy, University of Louisville, 313B Humanities Building, Louisville, KY 40292, USA.
e-mail: guy.dove@louisville.edu

This essay proposes and defends a pluralistic theory of conceptual embodiment. Our concepts are represented in at least two ways: (i) through sensorimotor simulations of our interactions with objects and events and (ii) through sensorimotor simulations of natural language processing. Linguistic representations are "dis-embodied" in the sense that they are dynamic and multimodal but, in contrast to other forms of embodied cognition, do not inherit semantic content from this embodiment. The capacity to store information in the associations and inferential relationships among linguistic representations extends our cognitive reach and provides an explanation of our ability to abstract and generalize. This theory is supported by a number of empirical considerations, including the large body of evidence from cognitive neuroscience and neuropsychology supporting a multiple semantic code explanation of imageability effects.

Keywords: concepts, imageability, embodied cognition, language, representation, symbol grounding

INTRODUCTION

In this essay, I propose and defend a new take on a familiar idea. The familiar idea is that our concepts are encoded in at least two general types of semantic representations: one type that is perception and motor based and another that is language based (Paivio, 1971, 1986). Although most concepts employ both types of representations, abstract concepts tend to depend more on linguistic representations than concrete concepts do. What separates my version of this idea from most previous ones is that I develop it within an embodied approach to cognition (although see Barsalou et al., 2008; Louwerse and Jeuniaux, 2008 for related yet distinct views).

My defense of this new take has three parts. The first part outlines and motivates an embodied approach to concepts based on simulation. The second part examines a challenge that faces any form of embodied cognition: the problem of abstraction. After making the observation that the symbolic structure of language is well suited to solving this problem, I propose that language should be seen as a form of what I refer to as "dis-embodied" cognition. What I mean by this is that linguistic representations are embodied in the neurophysiological sense that they rely on sensorimotor simulation but, unlike other embodied forms of cognition, they do not inherit semantic content from this fact. They do, however, accrue semantic content through their associations and inferential relationships with other linguistic representations. The third part surveys empirical evidence that supports the existence of separate semantic codes.

EMBODIED CONCEPTS

Historically, cognitive scientists have presumed that higher cognitive processes are carried out by computations involving amodal mental representations (i.e., representations that are not located within a sensorimotor modality). The precise nature of these representations was a matter of some debate. For instance, a great deal of controversy has surrounded the issue of how language-like they might be (Fodor, 1975). The presumption of amodality, however, went largely unquestioned. The strength of this presumption was clearly demonstrated by the heated nature of the debate concerning the possibility that analog perceptual representations might be employed in mental imagery tasks (Pylyshyn, 1973, 1981; Kosslyn and Shwartz, 1977). Now, there is general agreement that behavioral and neural evidence suggests that mental imagery (Kosslyn, 1994) and motor imagery (Jeannerod, 1995; Grèzes and Decety, 2001) depend on sensory and motor representations respectively.

Within the last two decades, a growing number of researchers and philosophers have argued that cognitive science needs to reorient itself with respect to its fundamental assumptions about the nature of mind and cognition. These researchers and philosophers contend that cognitive processes need to be viewed as fundamentally based in our bodily interactions with the world. Clark (1998, p. 506) expresses this view clearly in his economical assertion that, "Biological brains are first and foremost the control systems for biological bodies." The idea is that we cannot hope to understand the functioning of the brain without appreciating the central role it plays in guiding perception and action. This view has lead to a robust and diverse research program in which investigators examine the possible ways in which thinking, remembering, and understanding language are shaped by the fact that we dynamically interact with our complex physical and social environment by means of perceptual and motor capacities (Wilson, 2002). Embodied theories of cognition often suggest that concepts are understood via sensorimotor simulations. Neural systems that are involved in understanding real objects, actions, and events in the world are used to internally simulate those objects, actions, and events at later points in time.

THE THEORETICAL PROMISE OF EMBODIED CONCEPTS

Within cognitive science, the orthodox approach to concepts views them as containing amodal representations. This approach posits mental symbols that are manipulated solely based on their syntactic properties. By assumption, there is no intrinsic connection between these symbols and what they represent. This approach faces a well-known challenge: the symbol grounding problem. Harnad (1990,

p. 335) summarizes this problem with the question, "How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?" Perhaps the easiest way to think of this problem is to imagine trying to learn a foreign language from a dictionary in that language. Each word would be defined in terms of its connections to other words. In order to avoid this problem, the meaning of at least some mental symbols must be grounded in something other than their syntactic properties.

A key impetus for the hypothesis that concepts are couched in sensorimotor representations is the belief that this will help with the symbol grounding problem. In order to see just how it might help, we need to have a clear conception of what an embodied account of concepts might look like. For that reason, I am going to briefly sketch what I take to be the strongest and most developed embodied accounts of concepts: the perceptual symbol theory (Barsalou, 1999). I should emphasize, though, that many of the points made in this essay extend beyond this particular theory and do not depend on its ultimate success. A core tenet of perceptual symbol system theory is that sensorimotor simulations of experience are of central importance to our concepts. Intuitively, the idea is that our conceptualization of a category consists of simulating the experience of perceiving and/or acting on exemplars of that category. Such simulations are the result of a kind of neurophysiological re-enactment: information concerning the neural activation patterns associated with perception or action, which has been captured and stored by conjunctive neurons in neighboring association areas or convergence zones (Damasio and Damasio, 1994), is used later in absence of relevant input to generate a partial reactivation of the sensorimotor representations.

Perceptual symbols have a number of properties that make them well suited to serve as conceptual representations (Barsalou, 1999, 2003). First, simulations need not be conscious – that is, they may contain unconscious perceptual representations (for evidence to this effect see Pecher et al., 2009). This property removes some of the traditional objections to imagistic theories of cognition that turn on the unreliability or vagueness of introspection. Second, simulations will often be schematic in the sense they contain only some of the sensorimotor representations involved in the experience being simulated. For instance, a simulation in the visual modality of the concept DOG might involve shape representations but not color representations. Third, they will typically be multi-modal in the sense that they involve the reactivation of perceptual representations in several modalities. Fourth, perceptual symbols provide a novel means of drawing the type/token distinction (Barsalou, 1999, 2003). This is achieved through distinguishing *simulators* and *simulations*. A simulator is a distributed system spanning association and sensorimotor areas. To possess a concept, such as DOG, is to have a skill or ability to generate appropriate perceptual representations of dogs in a given situation. An innovative aspect of Barsalou's account is that it holds that these simulations are context-sensitive: simulations for a given concept vary depending on the context and the speaker's goals. For example, they might represent objects from a particular perspective. Typically, simulations will involve only a small subset of the information stored in memory.

Although I believe that simulation-based accounts of embodiment have the most empirical promise, I should acknowledge that there are other theoretical conceptions of embodiment. Borghi (2005) identifies two distinct strains of embodied cognition – one that focuses on affordances and situated action and the other that focuses on simulation – and argues that both are true depending on the context. I am going to limit myself to the simulation framework here for a couple of reasons. The first is that I believe that this framework is more flexible than critics assume. An unfortunate consequence of Barsalou's use of the term *perceptual symbol* is that it gives the false impression that simulations are based in perception and not in action mechanisms. However, nothing in the theory prevents purely motor-based simulations. Perceptual symbols are thus compatible with, for example, action schemas (Glenberg and Robertson, 1999). The second reason is that I am committed to a representational approach to concepts. One of the issues that separate different views of embodiment is the status of representations. Many proponents of affordances and situated action embrace non-representational accounts of cognition – often appealing to the promise of dynamical systems theory. Based largely on this issue, Clark (1997) distinguishes between embodied cognitive science and radical embodied cognitive science. Clark (1997, 2008) defends the former while theorists such as Chemero (2009) defend the latter. Siding with Clark, I assume that the notion of representation is too useful to give up and, furthermore, that an empirically successful theory of concepts will involve an appeal to representations (Markman and Dietrich, 2000).

A perceptual symbol consists of a neurophysiological re-enactment of a collection of sensorimotor representations. It can be thought of as having perceptual content because there are certain states of affairs in the world that would be likely to elicit these representations under normal conditions. Barsalou argues that this perceptual content can provide a leg up with regard to intentional content. He writes (Barsalou, 1999, p. 597; emphasis in the original):

> Where perceptual symbols do have an advantage [over amodal symbols] is in the ability of their content to play a *heuristic* role in establishing reference. Although perceptual content is rarely definitive for intentionality, it may provide a major source of constraint and assistance in determining what a symbol is about.

The general idea is that perceptual symbols help us refer to objects and events because they are already causally connected with those objects and events. This causal connection does not fully determine the conceptual content of a perceptual symbol but it can help secure that content.

Although embodied cognition has promise with respect to helping with the symbol grounding problem, it seems too early to declare victory for two reasons. The first is that it is not clear that the problem has been fully solved (Taddeo and Floridi, 2005). The second is that other approaches may have the conceptual resources to address the problem. Instead of proclaiming that embodiment solves this longstanding problem, I am going to make a weaker and hopefully less controversial claim: the heuristic role identified by Barsalou is an attractive design feature of perceptual symbols. A conceptual system containing perceptual symbols can benefit from the role that sensorimotor representations play in guiding action and perception. To be more precise, I am going to claim that this design feature is more beneficial with some concepts than it is with others.

EMPIRICAL EVIDENCE

There is little question that embodied cognition has been a productive research program. New research seems to emerge daily. Due to this abundance, I am only going to offer a selective review. My purpose is not to be comprehensive but, instead, to provide general motivation for an embodied approach to concepts.

A number of behavioral experiments support the notion that perceptual representations are central to some cognitive tasks. For instance, Pecher et al. (2003) found a modality-switching cost in a linguistic task. Participants verified verbally expressed facts involving one modality, such as the fact that *leaves rustle*, more rapidly after verifying a fact involving the same modality, such as the fact that *blenders make noise*, than after verifying a fact involving a different modality, such as the fact that *cranberries are tart*. More recently, van Dantzig et al. (2008) found a similar modality-switching cost between a perceptual detection task and a property verification task. Stanfield and Zwaan (2001) asked participants to affirm whether or not pictures depicted the actions described in previously presented sentences. The actions had either a vertical or horizontal orientation (such as driving a nail into a floor or into a wall). Participants responded more quickly to the pictures that had the same orientation as the action described. Stanfield and Zwaan (2001) suggest that the subjects generated a perceptual image of the action described in the sentence and then used this image to carry out the affirmation task.

Other behavioral studies demonstrate the degree to which cognitive tasks can be interwoven with action. For instance, Borghi et al. (2004) found a compatibility effect associated with language processing and action. Participants were instructed to decide whether or not a word that followed a sentence named a part of the object mentioned in the sentence. Half of the selected parts were found in the upper-portion of the object and half were found in the lower-portion. The experimenters found that responses were faster when the direction of the key press movement (upward or downward) matched the part location (upper or lower). Further studies indicate that the motor representations elicited by the cognitive tasks can exhibit somatotopic specificity. For instance, Scorolli and Borghi (2007) asked their participants to judge whether or not simple sentences containing a verb and a noun were sensible or not and respond either by pressing a pedal or speaking into a microphone. The verbs in the sentences referred to actions that were typically performed with the mouth, hands, or the feet. Response times with the microphone were fastest with "mouth-sentences" and response times with the pedal were fastest with "foot-sentences" (see also Scorolli et al., 2009).

Researchers have produced evidence using transcranial magnetic stimulation (TMS) that provides compelling support for the behavioral findings. Pulvermüller et al. (2005) carried out a TMS study in which they found that stimulation over motor areas affects action word processing. They weakly stimulated different parts of the motor system while participants performed a lexical decision task on arm- and leg-related action words. Weak stimulation of left hemisphere areas associated with arm-movement led to an increased response time with arm-related words in comparison with leg-related words, and the reverse pattern occurred with weak stimulation of motor areas associated with leg-movement. Response times were not modulated in a control condition with a sham stimulation. Using a different experimental paradigm, Buccino et al. (2005) found that listening to action-related sentences modulated activity in the motor system. Motor evoked potentials (MEPs) recorded from hand and foot muscles were specifically modulated by hand-related and foot-related action sentences respectively.

More support is provided by the fact that lesions can lead to the loss of multiple categories that share perceptual properties (Simmons and Barsalou, 2003). For instance, Adolphs et al. (2000) found that damage to the somatosensory cortex was correlated with deficits in the visual recognition of facial expressions. The authors propose that simulation of producing facial expressions is involved in the recognition of facial expressions in others. A selective deficit in action word processing has been found patients with motor neuron disease (Bak et al., 2001). A word of caution is needed, though, because modality-specific damage does not explain the category-specific deficits of all patients (Caramazza and Mahon, 2006).

A body of brain imaging data support an embodied approach to concepts. Martin et al. (1996), for example, found increased activation in visual areas with categories that appear to rely heavily on visual information for identification. Simmons et al. (2007) find evidence of a common neural substrate for color perception and verification of object-associated color (e.g., *taxi-yellow*). Using a visual naming task, Chao and Martin (2000) found increased activity in motor areas with highly manipulable objects when compared to less manipulable objects. Hauk et al. (2004) had participants read individual words that referred to actions involving leg, arm, and head movements such as *lick*, *pick*, and *kick*. They found that reading each type of action word produced increased activation in successively in the regions of M1 associated with performing the relevant movements. In a task where participants listened to action-related sentences, Tettamanti et al. (2005) found increased activation in effector-specific premotor and motor areas.

In sum, a number of studies using different experimental paradigms and techniques implicate sensorimotor representations in various cognitive tasks. Positing perceptual symbols provides an economical and robust explanation for a diverse set of observed phenomena, including reaction times, the functional character of some neuropathologies, and neural activation patterns in response to certain cognitive tasks.

CHALLENGES TO THE EVIDENCE

Aside from the problem of abstraction, which will be discussed in the next section, the inference to embodied cognition from the available evidence faces two major challenges. The first concerns how the debate is framed. Machery (2007, 2010) argues that amodal theories are not monolithic, and there are conceivable amodal systems that would fit with the available evidence. In a similar vein, Mahon and Caramazza (2008) contend that the activity in sensoriomotor areas observed in many experiments could be the result of spreading activation from amodal representations. The ability to offer amodal explanations for the available evidence undermines some of the hyperbolic rhetoric used by supporters of embodied cognition. Too often, such supporters claim that the empirical predictions of embodied and amodal approaches sharply diverge. What Machery and Mahon and Caramazza demonstrate is that the empirical decision between the embodied and amodal approaches may be more difficult than some have advertised. This point seems

well taken; the issue will ultimately be decided by which approach is best supported by the evidence. The defeasible position of this paper is that the available evidence favors an embodied approach.

The second challenge is that the neuroimaging evidence does not exclude the presence of amodal representations. Indeed, many of the cited imaging experiments find modulation of activity in multiple brain areas. Several commentators (e.g., Weiskopf, 2007; Chatterjee, 2010; Machery, 2010) point out that a number of the neuroimaging studies cited in support of embodied cognition actually find modulated activity in brain areas that are near – but not identical to – areas used for perceptual and motor processing. This is a serious challenge to a philosophical position known as neo-empiricism (Prinz, 2002). A core tenet of this position is that *all* conceptual representations are modality-specific (Machery, 2010). Against this universal claim, evidence suggesting that some conceptual representations are located within areas outside of areas used for perceptual and motor processing is damning. It is not clear, though, that such evidence undermines a simulation-based embodied approach.

On some level, the distributed activation patterns found in the literature fit with the theory of perceptual symbols. Barsalou (2003) proposes that long-term memory integration processes underlie the ability create appropriate simulations. Such processes are needed to explain our ability to generalize and abstract away from particular exemplars and generate the right simulations on a given occasion. This move offloads significant aspects of conceptualization into non-perceptual association areas or convergence zones (Damasio and Damasio, 1994). It also raises the question of whether or not these areas contain amodal symbols. Barsalou et al. (2003, p. 87) concede that "…conjunctive neurons in convergence zones constitute a somewhat amodal mechanism for capturing and re-enacting modality-specific states" but then go on to point out that alternative explanations of the activity of these neurons are available that do not require amodal symbols. They then suggest that we should pragmatically assume that convergence zones do not contain amodal symbols until evidence suggests otherwise.

This is not a satisfying solution to the challenge posed by activation in convergence zones because it is provisional and *ad hoc*. Fortunately, there is a better way to meet this challenge: we can adopt a more liberal definition of an embodied concept. The fundamental intuition behind the embodied approach is that cognition is fundamentally integrated with perceptual and motor systems. Such integration does not in and of itself exclude supramodal or even amodal representations as long as the function of these representations is to engage appropriate simulations and not to act as independent conceptual representations. I would even go further and suggest that the very modal/amodal distinction fits poorly with an integrated embodied perspective because it presupposes a clean distinction between cognition and perception. From an embodied perspective, no such clean distinction exists. If I am right, then evidence of relevant neural activity in areas near to, but not directly associated with, a particular sensorimotor modality is not unequivocally incompatible with an embodied approach.

THE PROBLEM OF ABSTRACTION

A well-known limitation of the evidence for embodied concepts is that it primarily involves concrete or highly imageable concepts (Pezzulo and Castelfranchi, 2007; Louwerse and Jeuniaux, 2008; Dove, 2009). This is problematic because, although it is not difficult to imagine how embodiment might help us acquire concrete concepts, it is difficult to see how it can be anything but a hindrance with abstract concepts such as DEMOCRACY, ELECTRON, ENTROPY, JUSTICE, NUMBER, PATIENCE, and TRUTH. Representations grounded in sensorimotor systems do not seem to be well suited to representing abstract intentional contents. For this reason, abstract concepts remain a critical issue for embodied cognition. More is at stake than simply the reach of this approach. For instance, Mahon and Caramazza (2008, p. 60) use the challenge posed by abstract concepts to support a parsimony argument in support of an amodal approach to concepts:

> Given that an embodied theory of cognition would have to admit 'disembodied' cognitive processes in order to account for the representation of abstract concepts, why have a special theory just for concepts of concrete objects and actions?

While I am not convinced that such parsimony arguments have much force (the history of psychology is rich with highly economical failed theories), the core premise of this argument – i.e., that abstract concepts require disembodied cognition – needs to be examined.

THREE EMBODIED APPROACHES TO ABSTRACT CONCEPTS

Supporters of embodied concepts have begun to address the problem of abstraction. Three main approaches exist in the literature (for a review see Glenberg et al., 2008). Although each approach has some empirical support, there are reasons to believe that these approaches do not provide a full solution to the problem of abstraction.

The first and most well established approach involves metaphoric extension. This approach originally emerged from work in cognitive linguistics (Lakoff and Johnson, 1980; Lakoff, 1987). The core idea is that we often understand one conceptual domain metaphorically in terms of another. Often, these metaphors are shaped by image schemas formed from our bodily interactions, linguistic experience, and historical context. For instance, the concept of ARGUMENT may be understood in terms of the concept of WAR. The primary evidence for this approach is our use of linguistic metaphors. Some recent behavioral studies, though, provide evidence of the metaphorical use of space to represent abstract concepts. For instance, Boroditsky and colleagues (Boroditsky and Ramscar, 2002; Casasanto and Boroditsky, 2008) provide evidence that some temporal judgments rely on spatial representations. Richardson et al. (2003) attempted to ascertain whether or not comprehending abstract verbs, such as *argue* and *respect*, automatically activates spatial image schemas with a specific orientation (horizontal for *argue* and vertical for *respect*). Participants listened to short sentences while engaged in either a visual discrimination task or a picture memory task. Reaction times suggest that there was an interaction between the horizontal/vertical orientation of the image schema and the horizontal/vertical orientation of the visual stimuli.

The second approach is similar in spirit to the first but focuses on the importance of action schemas (Glenberg and Robertson, 1999). The core idea of this approach is that some abstract language is grounded in motor processes. A primary source of evidence is

the action–sentence compatibility effect or ACE (Glenberg and Kaschak, 2002). Glenberg and Kaschak found that reaction times decreased when response direction (a button press either away/toward the body) and the implied direction of either concrete action sentences (e.g., *Andy gave you the pizza/You gave Andy the pizza*) or abstract transfer sentences (e.g., *Liz told you a story/You told Liz a story*) matched. They suggest that the ACE is the result of competition for resources by the motor planning associated with the action and the language processing associated with the sentence. Adding to the behavioral research, Glenberg et al. (2008) recently provide neurophysiological evidence that comprehension of both object-transfer and abstract-transfer sentences modulates motor system activity.

The third approach proposes that, contrary to our intuitions, some abstract concepts involve situated simulations (Barsalou, 1999). This approach is supported by evidence from feature generation experiments. In a preliminary study, Barsalou and Wiemer-Hastings (2005) asked participants to generate typical properties for three abstract concepts (TRUTH, FREEDOM, and INVENTION), three concrete concepts (BIRD, CAR, and SOFA) and three intermediate concepts (COOKING, FARMING, and CARPETING). The authors report two core findings: that participants generated situational properties with both concrete and abstract concepts and that participants tended to generate more event and introspective properties with abstract concepts. They propose that abstract and concrete concepts are generally associated with different aspects of situations: abstract concepts tend to focus on social aspects while concrete concepts tend to focus on physical entities and actions. In a more fully realized experiment employing similar methodology, Wiemer-Hastings and Xu (2005) found that participants tended to produce fewer entity properties, more introspective properties, and more relational properties with abstract concepts than with concrete concepts.

How promising are these approaches? Let us consider each in turn. There are a number of reasons to be skeptical of metaphorical projection solution to the problem of abstraction. First, there are reasons to question the force of the linguistic evidence supporting this approach. It is just not clear that such linguistic patterns directly reflect conceptual structure. Indeed, alternative explanations of metaphors that do not require positing metaphoric representations are available (Murphy, 1997). Another problem is that this proposal seems developmentally implausible (Murphy, 1996). For example, it seems unlikely that an understanding of the complexities of war is required for the acquisition of the concept of an argument. Furthermore, evidence suggests that children's understanding of metaphor remains quite poor before the ages of 8–10 (Winner et al., 1976). Finally, there is an inherent difficulty faced by the attempt to capture conceptual content in terms of metaphor: while a metaphor enables us to highlight the similarities between two concepts, it cannot capture the important differences. Arguments, after all, are not really wars. Recognizing the appropriate connections between a perceptual experience and what it is being metaphorically extended to cover seems to require a prior understanding of the concept. Without such an understanding, it is difficult to see how one can arrive at a correct interpretation of a metaphor. The very ubiquity of spatial metaphor undermines its potential for representing a specific abstract concept such as RESPECT. This ubiquity raises the question of whether a non-metaphoric understanding of the target concept is needed to anchor these metaphoric uses.

Although the action schema approach is similar in spirit to the metaphorical projection approach, it enjoys some advantages over the metaphorical projection approach. For one, the evidence offered in support of this approach seems more substantial and less equivocal. Second, the developmental picture behind this approach seems more plausible. It fits with the developmental evidence suggesting that concrete or highly imageable event words are easier for young children to acquire than abstract ones (Maguire and Dove, 2008). Despite these advantages, the action schema approach faces some of the same challenges as the metaphoric projection approach. For instance, the apparent representational flexibility of action schemas raises the question of how it is possible to acquire the relevant abstract concepts. If the same action schema underlies various concepts, how are the differences between these concepts represented? Another problem is that it is difficult to imagine how action schemas can account for all abstract concepts. For instance, it is not clear how they might handle concepts such as ELECTRON, NUMBER, and TRUTH.

Finally, consider the situated simulation approach. The body of evidence cited in support of it is admittedly quite thin. More importantly, this evidence may not resolve the issue of the embodiment of conceptual representations. A supporter of amodal symbols could well argue that disembodied symbols are needed to account for our ability to represent the social and relational aspects of situations. In the end, the most serious problem facing the situated simulations proposal is that a particular abstract concept such as DEMOCRACY is not likely to be associated with a simple set of sensorimotor experiences (Dove, 2009).

In sum, current attempts to offer an embodied solution to the problem of abstraction appear suffer from two weaknesses: insufficiency and incompleteness. The approaches appear to be insufficient because they do not provide a full explanation of the concepts to which they apply. They appear incomplete because they do not seem to capture all abstract concepts. This is not to say that these proposals have no merit. Instead, I suggest that each has some promise and empirical support, but, ultimately, more is needed to explain our ability to abstract and generalize.

DIS-EMBODIMENT

Supporters of an embodied approach to concepts tend to treat the problem of abstraction as a collection of exceptions. The task then becomes to explain a subset of these exceptions using the theoretical techniques and experimental designs of the research program of embodied cognition. This effort ignores the fact that abstraction represents a general problem for embodied concepts. What we need to explain is our ability to go beyond embodied experience. Earlier we emphasized how grounding our concepts in action and perception systems may help us acquire conceptual content. Now, we need to acknowledge that such grounding has potential costs associated with it. In particular, sensorimotor simulations seem ill-suited for representing conceptual content that is not closely tied to particular experiences. The problem is that some concepts appear to require what we might call ungrounded representations.

The orthodox position within cognitive science, clearly expressed in the quote from Mahon and Caramazza given above, is that such "disembodied" concepts require amodal representations. If we look at the general features of the proposed embodied solutions to the problem of abstraction – particularly the metaphor projection and action schema approaches – a different theoretical possibility emerges. Each of these approaches proposes ways in which embodied representations associated with a certain experiential/cognitive domain can be used to refer to objects and events outside of that domain. To capture this idea, I am going coin a new term: *dis-embodiment*. A mental symbol is dis-embodied if (1) it is embodied but (2) this embodiment is arbitrarily related to its semantic content. In other words, a mental symbol is dis-embodied if it involves sensorimotor simulations of experiences that are not associated with its semantic content. The dash in the middle of this term is intended to distinguish this notion from the more general notion of disembodiment to which Mahon and Caramazza appeal. What I want to suggest is that the proposals outlined above are on the right track, but they fail to provide a general solution to the problem of abstraction. Below, I argue that natural language itself serves as a form of dis-embodied cognition and plays an extensive role in enabling us to acquire and use abstract concepts.

LANGUAGE AS A FORM OF DIS-EMBODIED COGNITION

One way to approach the problem of abstraction is to scrutinize the abstract/concrete distinction (Scorolli, 2009). A number of researchers suggest that there are qualitative differences between abstract and concrete concepts. For example, Barr and Caplan (1987) propose that a meaningful distinction can be drawn between categories that are primarily represented by "extrinsic" features (those associated with relations between two or more entities) and those that are represented by "intrinsic" features (those associated with individual entities). Based on property generation studies, Wiemer-Hastings and Xu (2005) propose a two-factor account in which abstract concepts are both less contextually specific and predominately associated with social aspects of situations. Crutch and Warrington (2005) propose a qualitative distinction in which concrete concepts are organized primarily around similarity and abstract concepts are organized around semantic association. A recent eye-tracking experiment suggests that these representational differences emerge during on-line word-recognition (Deñabeitia et al., 2009). Participants were presented with visual displays that included a target picture of item that was a semantic associate of an abstract or concrete word. Their eye-movements were recorded as they listened to the relevant words. They tended to fixate more (and earlier) on depicted objects that were associates of abstract words than associates of concrete words. Overall, evidence of a qualitative distinction between abstract and concrete concepts is growing. What is the source of this distinction? I propose that it arises from an asymmetry between the types of representations employed by abstract and concrete concepts. While concrete concepts generally depend on both linguistic and non-linguistic perceptual symbols, abstract concepts tend to rely primarily on linguistic perceptual symbols.

Natural language has a number of design features commonly associated with amodal symbol systems that make it well suited to representing abstract concepts. Indeed, natural language is often held up as a paradigmatic example of an amodal symbol system. Three design features are particularly important. The first is the inherent representational arbitrariness of words and morphemes. There is, for example, no intrinsic similarity or other extralinguistic connection of the English word *cat* to the category of cats. Indeed, other languages associate phonetically and graphemically different words with the same category. Furthermore, the phonemic similarity of *cat* to *cap* carries no weight with respect to the contents of these words. The second is its stimulus-independence (Chomsky, 1966). Competent speakers are able to produce linguistic utterances in a self-generated fashion that is not an immediate response to proximal environmental stimulation. The third is its systematicity (Fodor, 1975; Fodor and Pylyshyn, 1988; Pinker, 1994). The ability to produce a sentence such as *Joni loves Chachi* seems to come hand in hand with the ability to form other sentences such as *Chachi loves Joni* and *Jenny loves Chachi*, etc. A common explanation of these design features is that natural language amounts to a syntactically recombinable symbol system. While there are disagreements concerning the cognitive architecture that underlies our linguistic competence, a large body of linguistic research suggests that the morphosyntactic structure of language is at least characterizable in terms of a productive grammar.

Now the mere fact that natural language is stimulus-independent and systematic does not sufficiently distinguish it from garden variety perceptual symbols. One of the achievements of perceptual symbol theory is that it demonstrates how a simulation-based symbol system might have these properties (Barsalou and Prinz, 1997). Stimulus-independence and systematicity alone cannot establish an advantage of verbal over non-verbal representations with respect to abstract contents. Natural language must bring something else to the table. In a philosophical exploration of possible conceptions of animal and human cognition, Camp (2009) suggests that we should view stimulus-independence and recombinability as degree properties. She then argues that natural language enhances these features in at least four ways. First, natural language is likely to increase the range of thoughts that any one individual may entertain because it enables one to hear the thoughts of others. Second, natural language makes it easier to reproduce the same thought in different situations because of its lack of context-sensitivity. Third, the manifest syntactic structure of natural language highlights the potential recombinality of thoughts and thus encourages us to entertain a wider of thoughts. Finally, natural language provides a sufficiently rich expressive medium to allow one to represent truth-values and inferential relations among thoughts. These enhancements mean that a creature with language is likely to enjoy a general cognitive advantage over a creature that does not.

A primary benefit afforded by a natural language is that it provides a representational system that can play the integrative role traditionally associated with amodal symbols. Consider the following argument for the necessity of amodal representations. After recognizing the existence of independent sensorimotor codes, Jackendoff (1992, p. 3) contends that amodal representations are necessary because "…none of these forms of input and output information suffices to explain the way that we understand the world in terms of objects, their motions, our actions on them, and so forth." The general idea is that amodal representations are needed to capture generalizations about entities and events that

go beyond the information contained within specific modalities. Amodal representations provide a means of gathering and integrating information from different modalities as well as transferring information between distinct sensorimotor codes. Because linguistic representations have the design features outlined above, they can also carry out this function (Carruthers, 2002).

I propose that when an individual acquires a natural language, she acquires a representational system that is different in some important respects from the multimodal, context-sensitive embodied symbol systems that exist independently of language. The acquisition of natural language, in other words, enhances and extends her representational abilities by giving her access to a context-free and arbitrary symbol system. This symbol system is independent of, and yet interacts with, other embodied symbols.

This proposal requires a revisionist conception of linguistic competence. Standard theories of linguistic competence are thoroughly amodal. Linguists have identified structural regularities at several levels of analysis, including phonology, morphology, syntax, and to some degree logical form or semantics. Knowledge relating to these levels is thought to be contained with language-specific functional modules (Fodor, 1983) and is generally thought to be couched in amodal codes. Comprehension involves translating perceptual information into these codes and production involves translating information in these codes into motor representations. The revisionist approach taken in this essay is that the process of achieving competence in a specific natural language involves acquiring the ability to generate appropriate simulations of linguistic experience. To be successful, these simulations must comport with the structural regularities at the different levels of analysis. They will not, however, depend on knowledge contained with an amodal symbol system. Three points about this revisionist proposal are especially important. The first is that it is neutral with respect to the issue of the degree to which linguistic competence is innate or learned. This proposal has to do with the format of the representations associated with this competence and not how it is acquired. The second is that, despite superficial appearances, this is not an inner speech view. The claim is that linguistic competence is contained within a system for generating perceptual symbols. These symbols consist of neurophysiological simulations that can be partial, selective, and unconscious. The third important point is that there is no independent lexical semantic code. The core thesis of this paper is that concepts are couched in two types of simulation-based representations: those associated with non-linguistic experience of the world and those associated with experience of language. Because simulations are detailed and often complex, linguistic perceptual symbols may exhibit structure at the various levels of analysis (phonology, morphology, syntax, etc.).

THINKING IN WORDS

Despite the clear differences between embodied and orthodox approaches to cognition, both adopt a similar view of the relationship between language and thought. Both see language as a medium of communication rather than a medium of thought. According to both, language expresses underlying thoughts that are encoded in some other semantic code. Within traditional cognitive science, this code is typically taken to be a language-like amodal symbol system (Fodor, 1975). Within embodied cognition, this code is thought to consist of embodied representations grounded in action and perception mechanisms. Glenberg et al. (2008, p. 4) offer the following summary of what researchers mean when they say that language is embodied:

> Linguistic symbols are embodied to the extent that: (a) the meaning of the symbol (the interpretant) to the agent depends on activity in systems also used for perception, action, and emotion, and (b) reasoning about meaning, including combinatorial processes of sentence understanding requires use of those systems.

The idea is that linguistic symbols have meaning because they dynamically activate sensorimotor representations associated with interacting with the world. On this account, linguistic symbols are intermediaries that do not directly have meaning or participate in reasoning about meaning.

I suggest that language plays two roles in our cognitive lives. One role is to engage sensorimotor simulations of interacting with the world. In this role, language serves primarily as a medium of communication. A second role is to elicit and engage symbolically mediated associations and inferences. Our concepts are not merely couched in sensorimotor representations but also in linguistic representations (words, phrases, sentences). Conceptual content is captured in part by the relationships of linguistic representations with other linguistic representations. These relationships may be merely associative or they may be inferential. On this view, a concept such as DOG will, not only be represented on a given occasion by multimodal simulations associated with interacting with dogs, but will also be represented in terms of related linguistic words, phrases, or sentences. This idea has a clear affinity with inferential role or conceptual role semantics (Harman, 1982; Block, 1986). This philosophical theory of mental content holds that the meaning of a concept is determined by its functional role within the cognitive life of an individual. My proposal is distinct from this theory because it adds the further requirement that the associative and inferential relationships be couched in language-based simulations.

One source of evidence for the view that internalized natural language can itself serve as a symbolic form of cognition is the effectiveness of statistical models that derive the meaning of words through statistical computations applied to large corpuses of text (Louwerse and Jeuniaux, 2008). A prominent example of this type of model is Latent Semantic Analysis or LSA (Landauer and Dumais, 1997). The idea behind LSA is that the aggregate of all the linguistic contexts in which a given word does and does not appear constrains semantic-relatedness. LSA has shown some effectiveness with respect to modeling a variety of linguistic tasks (Landauer et al., 1998). For example, an LSA model performed at a comparable level on the vocabulary portion of the Test of English as a Foreign Language to a large sample of students applying for college entrance in the United States from non-English speaking countries (Landauer and Dumais, 1997). Even if we grant that this particular model is psychologically implausible, it demonstrates the potential of a language-based representational system.

THEORETICAL INFLUENCES

I propose that our concepts are encoded in at least two types of semantic representations: one type employing embodied sensorimotor representations associated with our experience of the world

and the other type employing dis-embodied sensorimotor representations associated with our experience of language. Other types may exist. Gesture, for instance, might form an independent semantic representational system (Goldin-Meadow, 2003). This pluralistic embodied proposal has clear similarities with some previous theories. Highlighting the similarities – and the differences – between it and these theories should help clarify its central claims.

This proposal overlaps somewhat with another recent attempt to offer an embodied solution to the problem of abstraction. Borghi and Cimatti (2009) argue that supporters of embodied cognition have paid too little attention to the embodied social experience associated with language. They propose that there is a qualitative distinction to be made, not between two different mental processes, but rather between two different cognitive sources of grounding: one that depends crucially on direct sensorimotor experience and another that depends crucially on linguistic experience. Both of these sources can be useful in the acquisition of any concept but the acquisition of concrete concepts is likely to depend more on direct sensorimotor experience and the acquisition of abstract concepts is more likely to depend on linguistic experience. This distinction seems important and necessary. I suggest that it falls short, however, because it does not appropriately emphasize the importance of the computational properties of natural language. While I agree that linguistic experience is an important source of socially derived information about the world, I maintain that the structural properties of natural language contribute to its effectiveness in representing abstract concepts. My account differs from Borghi and Cimatti's because it holds that the acquisition of language creates a new dis-embodied semantic system, one that has many of the properties usually associated with the amodal symbol systems favored by traditional cognitive science. In other words, natural language on my view is not merely another source of information about the world but is also another way of thinking about the world.

My core thesis is that language is an internalized amodal symbol system that is built on an embodied substrate. As such, it extends our cognitive reach and helps us overcome the problem of abstraction. This idea is inspired in part by Andy Clark's view of language as a kind of cognitive scaffolding that provides cognitive benefits that would not otherwise be available to us. Clark (2008, p. 47) summarizes these benefits in the following passage:

> The computational value of a public system of essentially context-free, arbitrary symbols, lies… in the way such a system can push, pull, tweak, cajole, and eventually cooperate with various non-arbitrary, modality-rich, context-sensitive forms of biologically basic encoding.

Clark's claim is that natural language augments the cognitive abilities of an embodied mind. The core idea is that natural language is a cognitively useful symbol system, not because it mirrors the structure of our underlying thoughts, but because it does not. Clark makes much of the arbitrariness of linguistic symbols. Although the arbitrariness of the relationship between words and their semantic contents is well known, one might think that "forms of biologically basic encoding" are equally arbitrary. However, as we saw above in the context of the symbol grounding problem, there is a sense in which perceptual symbols are not arbitrary because they contain sensorimotor representations that enjoy a non-cognitive causal relationship with objects and events. Clark (2008) argues that language helps extend our cognitive abilities in at least three distinct but related ways: first, the very act of labeling objects and events provides a means of discovering increasingly abstract patterns in nature; second, the ability to recall and react to structured sentences enables us to acquire new skills and capacities, and third, our language abilities partially underwrite our ability to reflect on and influence the contents of our own thinking. Because he is primarily interested in simply establishing that language can in fact extend our cognitive abilities, Clark focuses on a collection of empirically based examples that seem to demonstrate cognitive extension. One of the most established of these is the apparent way in which verbal counting helps children acquire an understanding of positive integers (Dehaene, 1999; Carey and Sarnecka, 2006).

Where my account diverges from Clark's is with respect to scope. I contend that the sort of scaffolding he discusses is not limited to specific concepts or cognitive domains. Instead, acquiring a natural language extends our abilities to acquire concepts across the board. This is not simply because it offers a means of accessing socially derived information but also because it offers new representational powers. I suggest that most concepts depend to some significant degree on information represented in internalized natural language.

Clark may or may not be sympathetic with this general point, but there is no indication that he connects this scaffolding effect to the qualitative distinction between abstract and concrete concepts.

This brings us to perhaps the single greatest influence of the theory outlined in this essay: Dual Coding Theory or DCT (Paivio, 1986). This theory posits two independent cognitive subsystems, one employing symbolic verbal representations and the other employing analog non-verbal representations. Sadoski and Paivio (2004, p. 1340) write:

> A basic premise of DCT is that all mental representations retain some of the concrete qualities of the external experiences from which they derive. These experiences can be linguistic or non-linguistic. Their different characteristics develop into two separate mental systems, or codes, one specialized for representing and processing language (the verbal code) and one for processing non-linguistic objects and events (the non-verbal code).

The focus in DCT on the dynamic relationship between experience and mental representations seems to be in keeping with the basic tenets of embodied cognition. One might even reasonably see DCT as a precursor to the embodied cognition movement. However, an important aspect of DCT, i.e., its emphasis on language as an independent symbol system, has not generally been taken up by embodied cognition. To a certain degree, my account can be seen as an attempt to recapture an important insight from DCT within an embodied framework. It is important, however, to recognize that the result of this effort is not simply a recapitulation of DCT. There are some important differences between the account developed here and DCT. First, DCT claims that mental images are the basic constituents of the verbal and non-verbal systems. My account views perceptual symbols as the basic units. This is significant because perceptual symbol system theory represents an explicit attempt to avoid the weaknesses associated with image-based theories of concepts. Perceptual symbols differ from

mental images in a number of important ways: for instance, they need not be conscious, they can be schematic, and they are often multi-modal. Second, DCT and my theory differ with respect to the nature of the mental representations associated with language. According to DCT, they are a special class of mental images that are made up from different basic elements (*logogens*) than the basic elements of non-verbal representations (*imagens*). On my account, all conceptual representations consist of perceptual symbols. Linguistic representations are distinguished form non-linguistic ones by the fact that they are an internalization of an external symbol system.

In the end, the view advocated in this essay brings together ideas from a number of different theories and combines them in a novel way. While it clearly owes a debt to these previous views, it stands or falls on its own.

EMPIRICAL EVIDENCE

We began the last section with the acknowledgment of the seriousness of the problem of abstraction. We now have a theoretical picture of how language might help explain this ability: language might extend our cognitive abilities in such a way that enables us to have some of the benefits of an amodal symbol system. This theoretical picture rests on two independent hypotheses: (1) that language processing involves sensorimotor simulation and (2) that linguistic representations play an important role in our ability to abstract and generalize.

LANGUAGE PROCESSING INVOLVES PERCEPTUAL SYMBOLS

Given the dynamic nature of linguistic communication, the idea that language processing involves perceptual symbols seems attractive. After all, most linguistic communication is time-constrained and would seem to require the integration of action, perception, and cognition. Below, I survey some of the evidence favoring this hypothesis.

The first reason to think language processing might involve sensorimotor simulations is a negative one: the project to locate self-contained language areas of the brain has not succeeded. Ever since the work of Broca and Wernicke in the late nineteenth century (Finger, 1994), the classical localizationist position has been that subcomponents of language are represented and processed in bounded and specialized cortical regions (Geschwind, 1970). One of the primary sources of evidence for this perspective has been the study of aphasic syndromes resulting from focal brain injuries (for a review see Saffran, 2000). Researchers, however, have begun to move away from strict localization and toward the view that language requires the activity of a number of spatially distinct brain regions. This shift has occurred in response to several forms of evidence. For one, neuroimaging studies indicate that widely distributed brain areas are active in language processing (Posner and Raichle, 1994). Another reason for this shift is the fact that the association of grammatical processing with Broca's area has broken down to a large degree (Grodzinsky, 2000). For instance, there is evidence of some retained grammatical knowledge in Broca's aphasics (Bates and Wulfeck, 1989; Bates et al., 1991). In addition, grammatical deficits have been found in Wernicke's aphasics and other clinical populations (Dick et al., 2001). It also appears that grammatical deficits are associated with damage throughout the left perisylvian cortex (Caplan et al., 1996). Finally, recent evidence suggests that Broca's area itself might have multiple functions. For example, a number have studies have implicated in action-related tasks (Thoenissen et al., 2002; Nishitani et al., 2005). In sum, evidence from cognitive neuroscience and neuropsychology suggests that language processing is widely distributed in the brain and involves a number of sensorimotor areas. Although this distribution is not logically incompatible with an amodal approach, it fits well with the idea that language processing involves sensorimotor simulations.

A second, more direct reason to think that language processing might involve perceptual symbols is that there is evidence of functional links between motor and perception circuits with the left perisylvian cortex (Pulvermüller, 2005). For example, there is evidence that listening to speech modulates tongue muscle responses (Fadiga et al., 2002). This sort of evidence is often seen as supporting the motor theory of speech perception (Liberman and Whalen, 2000) or the direct realist theory (Fowler et al., 2003). Critics of these theories argue that auditory areas alone might be sufficient for perceiving speech (e.g., Toni et al., 2008). If true, this would rule out a strongly action-based account of speech perception in which speech perception necessarily involves motor processing. However, it does not rule out a weaker view that speech recognition generally involves multimodal perceptual symbols.

A third reason to suppose that language processing involves perceptual symbols is that several studies implicate active integration of multimodal information in on-line language processing. It is well established that visual input can influence phonemic speech-processing (McGurk and MacDonald, 1976). A large body of eye-tracking experiments shows the manifold ways in which visual information can be continuously integrated with auditory information during the processing of speech (Spivey and Richardson, 2009). Visual information has been shown to influence language comprehension at various levels of linguistic analysis, including word-recognition (Allopenna et al., 1998), syntactic processing (Tanenhaus et al., 1995), and thematic role assignment (Altmann and Kamide, 1999). Consider a study involving syntactic ambiguity (Spivey et al., 2002; Spivey and Richardson, 2009). Participants were presented with a four-quadrant display of real objects and instructed to carry out actions. The display on one condition contained (going clockwise from the upper left quadrant) a spoon on a napkin, a bare napkin, a bowl, and a pen. The participants were instructed to "Put the spoon on the napkin in the bowl." Eye-tracking evidence indicates that subjects often fixate on the irrelevant bare napkin before fixating on the bowl and carrying out the action. This suggests that they initially misparse the initial prepositional phrase as syntactically attached to the verb. This effect did not occur with a similar display in which two spoons appear, one on a napkin and one not on a napkin (replacing the pen in the earlier display).

A fourth reason to think that language processing might involve perceptual symbols is the employment of perceptual areas in language processing among people with congenital perceptual deficits. For example, neuroimaging studies find increased activation in auditory areas when congenitally deaf individuals view signs

(Petito et al., 2000). Similarly, some primary visual areas show increased activation when congenitally blind individuals read Braille (Sadato et al., 1996).

Taken together, these various bodies of evidence suggest that language processing is much more integrated with action and perception systems than was previously assumed by researchers. It should be acknowledged, however, that this evidence is only suggestive and not conclusive. One could maintain that this evidence does not falsify the hypothesis that language processing is handled by amodal symbols since the implicated activity in sensorimotor systems could be associated with spreading activation and not be constitutive of language processing. As I mentioned earlier in the essay, this is a general problem faced by any embodied hypothesis. Ultimately, the issue is an empirical one, and unfortunately the evidence currently available does not completely settle matters.

Given this uncertainty, it seems worthwhile to consider what would happen if it turns out that language processing is indeed handled by an amodal symbol system of the sort posited by the current orthodoxy. This would turn the hypothesis that language is a form of dis-embodied cognition into the hypothesis that language is a form of disembodied cognition (non-hyphenated). It would result in a different kind of hybrid theory, one in which concepts are represented by both multimodal perceptual symbols and amodal linguistic symbols. Although I am promoting the dis-embodied view in this essay, the second view is an intriguing and compelling alternative (for general arguments in favor of a hybrid approach see Dove, 2009; Kemmerer, 2010).

IMAGEABILITY RECONSIDERED

Imageability effects provide support for the account developed in this essay. Typically, imageability is defined as the ease with which a word gives rise to a sensory-motor mental image (Paivio, 1971). Imageability is a broader concept than concreteness because it includes sensory images of bodily states and motor images. It is generally recognized that imageability better captures the relevant phenomena and supports broader generalizations. Highly reliable imageability ratings on number scales have been gathered for linguistic concepts by number of researchers (Toglia and Battig, 1978; Bird et al., 2001). Traditionally, cognitive scientists examined imageability in terms of processing advantages for high imageable concepts over low imageable ones in several cognitive tasks. For instance, lexical access has been shown to be quicker for highly imageable words than for abstract ones (Coltheart et al., 1980) and highly imageable words are recalled more quickly in memory tasks than abstract words (Paivio, 1986; Wattenmaker and Shoben, 1987).

Two major theories dominate the literature: the DCT (Paivio, 1971, 1986) and the context-availability theory (Schwanenflugel and Shoben, 1983). According to the DCT, words with low imageability are associated primarily with verbal representations while highly imageable words are associated with both verbal representations and perceptual ones. Imageability effects are then explained in terms of the greater availability of perceptually encoded information. According to the context-availability theory, highly imageable words are more closely linked to relevant contextual knowledge in semantic networks than less imageable concepts. In other words, highly imageable words have greater contextual information stored in semantic memory, and imageability effects are to be explained by the facilitation of processing associated with increased activation in these networks. On this approach, the reason that participants respond more quickly in a lexical decision task to a word such as "fingertip" than to one such as "idea" is that the former has more semantic associations than the latter.

Evidence suggests that both theories are right, depending on the task. I am going to focus on the evidence for the DCT because this evidence has more relevance to the claims in this essay.

Consider first neuropsychological case studies. Several research teams describe aphasic patients with significant left hemisphere damage who exhibit a selective semantic impairment for high imageable words (Berndt et al., 2002; Bird et al., 2003; Crepaldi et al., 2006). Patients with a selective semantic impairment for low imageable words are less common but have also been found (Marshall et al., 1996; Luzzatti et al., 2002). This double dissociation suggests that, at least at some level, the semantic processing of concepts with low imageability is functionally independent from the semantic processing of concepts with high imageability.

A number of event-related potential (ERP) experiments support a neuroanatomical distinction between concepts of high and low imageability. For instance, Holcomb et al. (1999) created a task that involved manipulations of both context and concreteness. ERP recordings were time-locked to sentence-final words in a word-by-word reading task in which participants made semantic congruency judgments (e.g., Armed robbery implies that the thief used a *weapon* vs. armed robbery implies that the thief used a *rose*). They found that sentence-final concrete words generated a larger and more anterior N400 than sentence-final abstract words in both contexts (see also Kounios and Holcomb, 1994; West and Holcomb, 2000). Further studies have found context-independent topographic effects associated with imageability in single-word presentations (Kellenbach et al., 2002; Swaab et al., 2002). Using two-word stimuli that involved a noun preceded by either a concrete modifier or an abstract modifier ("green book" vs. "engaging book") in a visual half-field presentation, Huang et al. (2010) found distinct hemispheric responses. Thus, ERP studies employing diverse tasks support the notion that different cognitive systems are associated with the semantic processing of high and low imageable words.

Neuroimaging data supports the notion that neural activity is modulated by imageability. A number of studies find that abstract or low imageable words elicit greater activation than concrete or high imageable words in superior regions of the left temporal lobe (Mellet et al., 1998; Giesbrecht et al., 2004; Noppeney and Price, 2004; Binder et al., 2005; Sabsevitz et al., 2005) and inferior regions of the left prefrontal cortex (Giesbrecht et al., 2004; Noppeney and Price, 2004; Binder et al., 2005; Sabsevitz et al., 2005; Goldberg et al., 2006). This evidence fits with imaging studies that implicate the left inferior frontal gyrus or IFG in language processing (Bookheimer, 2002). When researchers make the comparison in the reverse direction, the pattern is less clear. Whereas some studies find no areas of increased activation (Kiehl et al., 1999; Perani et al., 1999; Tyler et al., 2001; Grossman et al., 2002; Noppeney and Price, 2004), others find increased activation in right hemisphere areas (Mellet et al., 1998; Jessen et al., 2000;

Binder et al., 2005; Sabsevitz et al., 2005). This divergence with respect to activation patterns fits with the neuropsychological observation that patients are more likely to have a selective deficit for abstract or low imageable words than for concrete or high imageable words.

Sabsevitz et al. (2005) carried out a particularly careful fMRI study. Their study incorporated a larger sample (28 adults) than previous studies and a task (judgment of semantic similarity) that is more likely to elicit deep semantic processing than a more superficial task, such as lexical decision. Participants were visually presented with three words (e.g., *cheetah*, *wolf*, and *tiger*) in the form of a triangle. The task was to decide which of the two bottom words was most semantically similar to the top word. In this task, abstract nouns elicited greater activation in the left superior temporal and left inferior frontal cortex than concrete nouns, while concrete nouns elicited greater activation in a bilateral network of association areas than abstract nouns.

The upshot of this survey is that imageability effects have been found in multiple disciplines by investigators in a number of labs using different research methodologies and measures. These effects provide support for the notion that abstract or low imageability concepts are processed somewhat differently than concrete or high imageability concepts. Areas associated with language processing appear to be more active during semantic tasks associated with abstract or low imageability concepts. This pattern of activation fits with both the hypothesis that language is a dis-embodied form of cognition and the hypothesis that it is an amodal form of cognition. The decision between these two hypotheses turns on the role played by the observed activity in language related areas. Is it part of linguistic sensorimotor simulations or is it part of amodal linguistic processes? This question awaits further research.

CONCLUSION

In this essay, I have attempted to assess the generality of embodied cognition. The current evidence for conceptual embodiment is compelling but, unfortunately, circumscribed. Part of the problem is that there has not been enough research on abstract concepts. Beyond this evidential lacuna, though, abstract concepts represent an important theoretical challenge to embodied cognition. The most promising attempts to deal with this problem appeal to what I have called dis-embodied representations. I have argued that there are good reasons to think that natural language itself is a form of dis-embodied cognition. The acquisition of competence with respect to a natural language provides access to syntactically recombinable symbol system that extends our cognitive reach.

The speculation that natural language extends the cognitive capacities of embodied minds points the way to new research opportunities. One question that needs to be answered more fully is just how the two types of conceptual symbol systems interact. The potential for interaction is implicit in the dual functionality of linguistic symbols. On the account developed here linguistic representations can serve as elicitors of non-linguistic perceptual symbols and as semantic symbols in their own right. Presumably, we have the ability to employ these systems in a context-sensitive and flexible way. However, the nature of this flexibility remains to be seen. Another question that arises is the extent to which language might explain other significant features of cognition. For example, both Dennett (1996) and Carruthers (2002) suggest that language may be the medium for conscious deliberation. Although this is not implied by the position outlined in this essay, the possibility that conscious deliberation involves language-based perceptual symbols seems worthy of investigation. In the end, the hypothesis that language is a dis-embodied form of cognition has both empirical support and theoretical promise.

REFERENCES

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., and Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J. Neurosci.* 20, 2683–2690.

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439.

Altmann, G. T. M., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264.

Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., and Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease-dementia-aphasia syndrome. *Brain* 124, 103–120.

Barr, R. A., and Caplan, L. J. (1987). Category representations and their implications for category structure. *Mem. Cogn.* 10, 82–93.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 1177–1187.

Barsalou, L. W., and Prinz, J. J. (1997). "Mundane creativity in perceptual symbol systems," in *Conceptual Structures and Processes: Emergence, Discovery, and Change*, eds T. B. Ward, S. M. Smith, and J. Vaid (Washington, DC: American Psychological Association), 267–307.

Barsalou, L. W., Santos, A, Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford, UK: Oxford University Press), 245–284.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thought*, eds D. Pecher and R. Zwaan (Cambridge, UK: Cambridge University Press), 129–163.

Bates, E., and Wulfeck, B. (1989). Comparative aphasiology: a cross-linguistic approach to language breakdown. *Aphasiology* 3, 111–142.

Bates, E., Wulfeck, B., and MacWhinney, B. (1991). Cross-linguistic studies of aphasia: an overview. *Brain Lang.* 41, 123–148.

Berndt, R. S., Haendiges, A. N., Burton, M. W., and Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: their effects are independent. *J. Neurolinguistics* 15, 353–371.

Binder, J., Westbury, C., McKiernan, K., Possing, E., and Medler, D. (2005). Distinct brain systems for processing concrete and abstract concepts. *J. Cogn. Neurosci.* 17, 905–917.

Bird, H., Franklin, S., and Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behav. Res. Methods* 33, 73–79.

Bird, H., Howard, D., and Franklin, S. (2003). Verbs and nouns: the importance of being imageable. *J. Neurolinguistics* 16, 113–149.

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Stud. Philos.* 10, 615–78.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188.

Borghi, A. M. (2005). "Object concepts and action," in *Grounding Cognition: The Role of Perception and Action in Memory*, eds D. Pecher and R. A. Zwaan (Cambridge, UK: Cambridge University Press), 8–34.

Borghi, A. M., and Cimatti, F. (2009). "Words as tools and the problem of abstract words meanings," in *Proceedings of the 31st Annual Conference of the Cognitive Science*

Society, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 2304–2309.

Borghi, A. M., Glenberg, A. M., and Kaschak (2004). Putting words in perspective. *Mem. Cogn.* 32, 863–873.

Boroditsky, L., and Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychol. Sci.* 13, 185–189.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res.* 24, 355–363.

Camp, E. (2009). Putting thoughts to work: concepts, systematicity, and stimulus-independence. *Philos. Phenomenol. Res.* 78, 275–311.

Caplan, D., Hildebrandt, N., and Makris N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain* 119, 933–949.

Caramazza, A., and Mahon, B. Z. (2006). The organization of conceptual knowledge in the brain: the future's past and some future directions. *Cogn. Neuropsychol.* 23, 13–38.

Carey, S., and Sarnecka, B. W. (2006). "The development of human conceptual representations: a case study," in *Processes of Change in Brain and Cognitive Development*, eds Y. Munakata and M. Johnson (Oxford, UK: Oxford University Press), 473–496.

Carruthers, P. (2002). The cognitive functions of language. *Behav. Brain Sci.* 25, 657–674.

Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593.

Chao, L. L., and Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12, 478–484.

Chatterjee, A. (2010). Disembodying cognition. *Lang. Cogn.* 2, 79–116.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Chomsky, N. (1966). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper and Row.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. (1998). "Embodied, situated, and distributed cognition," in *A Companion to Cognitive Science*, eds W. Bechtel and G. Graham (Malden, MA: Blackwell Publishers), 506–517.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford, UK: Oxford University Press.

Coltheart, M., Patterson, K., and Marshall, J. C. (eds). (1980). *Deep Dyslexia*. London: Routledge and Kegan Paul.

Crepaldi, D., Aggujaro, S., Arduino, L. S., Zonca, G., Ghirardi, G., Inzaghi, M. G., Colombo, M., Chierchia, G., and Luzatti, C. (2006). Noun–verb dissociation in aphasia: The role of imageability and functional location of the lesion. *Neuropsychologia* 44, 73–89.

Crutch, S. J., and Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain* 128, 615–627.

Damasio, A. R., and Damasio, H. (1994). "Cortical systems for retrieval of concrete knowledge: the convergence zone framework," in *Large-Scale Neuronal Theories of the Brain. Computational neuroscience*, eds C. Koch and J. L. Davis (Cambridge, MA: MIT Press), 61–74.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tviskin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974.

Deñabeitia, J. A., Aviles, A., Afonso, O., Scheepers, C., and Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: evidence from the visual-word paradigm. *Cognition* 110, 284–292.

Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Harper Collins.

Dick, F., Bates, E. Wulfeck, B., Utman, J. A., Dronkers, N., and Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychol. Rev.* 108, 759–788.

Dove, G. O. (2009). Beyond perceptual symbols: a call for representational pluralism. *Cognition* 110, 412–431.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.

Finger, S. (1994). *Origins of Neuroscience*. Oxford: Oxford University Press.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71.

Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: evidence from choice and simple response time tasks. *J. Mem. Lang.* 4, 396–410.

Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.

Giesbrecht, B., Gamblin, C., and Swaab, T. (2004). Separable effects of semantic priming and imageability on word processing in human cortex. *Cereb. Cortex* 14, 521–529.

Glenberg, A. M., de Vega, M., and Graesser, A. C. (2008). "Framing the debate," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford, UK: Oxford University Press), 1–10.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Glenberg, A. M., and Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Process.* 28, 1–26.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61, 905–919.

Goldberg, R. F., Perfetti, C. A., and Schneider, W. (2006). Distinct and common cortical activations for multimodal semantic categories. *Cogn. Affect. Behav. Neurosci.* 6, 214–222.

Goldin-Meadow, S. (ed.). (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Belknap Press.

Grèzes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19.

Grodzinsky, Y. (2000). The neurology of syntax: language use without Broca's area. *Behav. Brain Sci.* 23, 1–71.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., and Gee, J. (2002). The neural basis for category-specific knowledge: An fMRI study. *Neuroimage* 15, 936–948.

Harman, G. (1982). Conceptual role semantics. *Notre Dame J. Formal Logic* 28, 252–256.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Hauk, O., Johnsrude, I. and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context- availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742.

Huang, H., Lee, C., and Federmeier (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *Neuroimage* 49, 1116–1123.

Jackendoff, R. (1992). *Languages of the Mind: Essays on Mental Representation*. Cambridge, MA: MIT Press.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia* 33, 1419–1432.

Jessen, F., Heun, R., Erb, M., Granath, D.-O., Klose, U., Papassotiropoulos, A., and Grodd, W. (2000). The concrétense effect: evidence for dual coding and context availability. *Brain Lang.* 74, 103–112.

Kiehl, K. A., Liddle, P. F., Smith, A. M., Mendrek, A., Forster, B. B., and Hare, R. D. (1999). Neural Pathways involved in the processing of concrete and abstract words. *Hum. Brain Mapp.* 7, 225–233.

Kellenbach, M. L., Wijers, A. A., Hovis, M., Mulder, J., and Mulder, G. (2002). Neural differentiation of lexico-syntactic categories or semantic features? Event related potential evidence for both. *J. Cogn. Neurosci.* 14, 561–577.

Kemmerer, D. (2010). "How words capture visual experience: the perspective from cognitive neuroscience," in *Words and the World: How Words Capture Human Experience,* eds B. Malt and P. Wolff (Oxford: Oxford University Press), 287–327.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M., and Shwartz, S. P. (1977). A simulation of visual imagery. *Cogn. Sci.* 1, 265–295.

Kounios, J., and Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 804–823.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Process* 25, 259–284.

Liberman, A. M., and Whalen, D. H. (2000). On the relation of speech to language. *Trenes Cogn. Sci.* 4, 187–196.

Louwerse, M., and Jeuniaux, P. (2008). "Language comprehension is both embodied and symbolic," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A.

C. Graesser (Oxford, UK: Oxford University Press), 309–326.
Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., and Pinna, G. (2002). Verb–noun double dissociation in aphasic lexical impairments: the role of word frequency and imageability. *Brain Lang.* 81, 432–444.
Machery, E. (2007). Concept empiricism: a methodological critique. *Cognition* 104, 19–46.
Machery, E. (2010). Précis of "doing without concepts." *Behav. Brain Sci.* 33, 195–244.
Maguire, M. J., and Dove, G. O. (2008). "Speaking of events: event word learning and event representation," in *Understanding Events: How Humans See, Represent, and Act on Events*, eds T. F. Shipley and J. Zacks (Oxford, UK: Oxford University Press), 193–218.
Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.
Markman, A., and Dietrich, E. (2000). In defense of representation. *Cogn. Psychol.* 40, 138–171.
Marshall, J., Pring, T., Chiat, S., and Robson, J. (1996). Calling a salad a federation: an investigation of semantic jargon. Part 1 – nouns. *J. Neurolinguistics* 9, 237–250.
Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature* 379, 649–652.
McGurk, H., and MacDonald, T. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.
Mellet, E., Tzourio, N., Denis, M., and Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport* 9, 803–808.
Murphy, G. L. (1996). On metaphoric representation. *Cognition* 60, 173–204.
Murphy, G. L. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition* 62, 99–108.
Nishitani, N., Schürmann, M., Amunts, K., and Hari, R. (2005). Broca's region: from action to language. *Physiology* 20, 60–69.
Noppeney, U., and Price, C. J. (2004). Retrieval of abstract semantics. *Neuroimage* 22, 164–170.
Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston.
Paivio, A. (1986). *Mental Representations: A Dual-Coding Approach.* New York: Oxford University Press.
Pecher, D., Van Dantzig, S., and Schifferstein, H. N. J. (2009). Concepts are not represented by conscious imagery. *Psychon. Bull. Rev.* 16, 914–919.
Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychol. Sci.* 14, 119–124.
Perani, D., Cappa, S., Schnur, T., Tettamanti, M., Collina, S., Rosa, M., and Fazio, F. (1999). The neural correlates of verb and noun processing: a PET study. *Brain* 122, 2337–2344.
Petito, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., and Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proc. Natl. Acad. Sci. U.S.A.* 97, 13961–13966.
Pezzulo, G., and Castelfranchi, C. (2007). The symbol detachment problem. *Cogn. Process* 8, 115–131.
Pinker, S. (1994). *Language Instinct.* New York: Harper Collins.
Posner, M. I., and Raichle, M. E. (1994). *Images of Mind*, 2nd Edn. New York: W. H. Freeman and Co.
Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis.* Cambridge, MA: MIT Press.
Pulvermüller, F (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.
Pulvermüller, F., Hauk, O., Nikulin, V. V., Ilmonlemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797.
Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychol. Bull.* 80, 1–24.
Pylyshyn, Z. W. (1981). The imagery debate: analogue media versus tacit knowledge. *Psychol. Rev.* 88, 16–45.
Richardson, D. C., Spivey, M. J., Barsalou, L. W., and McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cogn. Sci.* 27, 767–780.
Sabsevitz, D., Medler, D., Seidenberg, M., and Binder, J. (2005). Modulation of the semantic system by word imageability. *Neuroimage* 27, 188–200.
Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M., Dold, G., and Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* 380, 526–528.
Sadoski, M., and Paivio, A. (2004). "A dual coding theoretical model of reading," in *Theoretical Models and Processes of Reading*, 5th Edn., eds R. B. Ruddell and N. J. Unrau (Newark, DE: International Reading Association), 1329–1362.
Saffran, E. M. (2000). Aphasias and the relationship of language and brain. *Semin. Neurol.* 20, 409–418.
Schwanenflugel, P. J., and Shoben, E. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 82–102.
Scorolli, C. (2009). *Language and Embodiment: Sensory-Motor and Linguistic-Social Experience. Evidence on Sentence Comprehension* (Doctoral dissertation). Available at: http://en.scientificcommons.org/55511932
Scorolli, C., and Borghi, A. M. (2007). Sentence comprehension and action: effector specific modulation of the motor system. *Brain Res.* 1130, 119–124.
Scorolli, C., Borghi, A. M., and Glenberg, A. M. (2009). Language-induced motor activity in bimanual object lifting. *Exp. Brain Res.* 193, 43–53.
Simmons, W. K., and Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol.* 20, 451–486.
Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., and Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45, 2802–2810.
Spivey, M., and Richardson, D. (2009). "Language processing embodied and embedded," in *The Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (Cambridge, UK: Cambridge University Press), 382–400.
Spivey, M., Tanenhaus, M., Eberhard, K., and Sedivy, J. (2002). Eye-movements and spoken language-comprehension: effects of visual context on syntactic ambiguity resolution. *Cogn. Psychol.* 45, 447–481.
Stanfield, R. A., and Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychol. Sci.* 12, 153–156.
Swaab, T., Baynes, K., and Knight, R. (2002). Separable effects of priming and imageability on word processing: an ERP study. *Cogn. Brain Res.* 15, 99–103.
Taddeo, M., and Floridi, L. (2005). The symbol grounding problem: a critical review of fifteen years of research. *J. Exp. Artif. Intell.* 17, 419–445.
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634.
Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V. Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–281.
Thoenissen, D., Zilles, K., and Toni, I. (2002). Differential involvement of parietal and precentral regions in movement preparation and motor intention. *J. Neurosci.* 22, 9024–9034.
Toglia, M. P., and Battig, W. F. (1978). *Handbook of Semantic Word Norms.* New York: Erlbaum.
Toni, I., de Lange, F. P., Noordzij, M., and Hagoort, P. (2008). Language beyond action. *J. Physiol. Paris* 102, 71–79.
Tyler, L. K., Russell, R., Fadili, J., and Moss, H. (2001). The neural representation of nouns and verbs: PET studies. *Brain,* 124, 1619–1634.
van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590.
Wattenmaker, W., and Shoben, E. (1987). Context and the recallability of concrete and abstract sentences. *J. Exp. Psychol.* 13, 140–150.
Weiskopf, D. (2007). Concept empiricism and the vehicles of thought. *J. Conscious. Stud.* 14, 156–183.
West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037.
Wiemer-Hastings, K., and Xu, X. (2005). Content differences for abstract and concrete concepts. *Cogn. Sci.* 29, 719–736.
Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.
Winner, E., Rosenstiel, A. K., and Gardner, H. (1976). The development of metaphoric understanding. *Dev. Psychol.* 12, 289–297.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2010; accepted: 23 December 2010; published online: 25 January 2011.
Citation: Dove G (2011) On the need for embodied and dis-embodied cognition. Front. Psychology ***1****:242. doi: 10.3389/fpsyg.2010.00242*
This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.
Copyright © 2011 Dove. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 22 November 2010
doi: 10.3389/fpsyg.2010.00197

A connectionist approach to embodied conceptual metaphor

Stephen J. Flusberg*, Paul H. Thibodeau, Daniel A. Sternberg and Jeremy J. Glick

Department of Psychology, Stanford University, Stanford, CA, USA

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Elisa Magosso, University of Bologna, Italy
Giovanni Pezzulo, Institute of Cognitive Sciences and Technologies, National Research Council, Italy

****Correspondence:***
Stephen J. Flusberg, Department of Psychology, Stanford University, Jordan Hall, 450 Serra Mall, Building 420, Stanford, CA 94305, USA.
e-mail: sflus@stanford.edu

A growing body of data has been gathered in support of the view that the mind is embodied and that cognition is grounded in sensory-motor processes. Some researchers have gone so far as to claim that this paradigm poses a serious challenge to central tenets of cognitive science, including the widely held view that the mind can be analyzed in terms of abstract computational principles. On the other hand, computational approaches to the study of mind have led to the development of specific models that help researchers understand complex cognitive processes at a level of detail that theories of embodied cognition (EC) have sometimes lacked. Here we make the case that connectionist architectures in particular can illuminate many surprising results from the EC literature. These models can learn the statistical structure in their environments, providing an ideal framework for understanding how simple sensory-motor mechanisms could give rise to higher-level cognitive behavior over the course of learning. Crucially, they form overlapping, distributed representations, which have exactly the properties required by many embodied accounts of cognition. We illustrate this idea by extending an existing connectionist model of semantic cognition in order to simulate findings from the embodied conceptual metaphor literature. Specifically, we explore how the abstract domain of time may be structured by concrete experience with space (including experience with culturally specific spatial and linguistic cues). We suggest that both EC researchers and connectionist modelers can benefit from an integrated approach to understanding these models and the empirical findings they seek to explain.

Keywords: connectionism, models, embodiment, conceptual metaphor, time, space

INTRODUCTION

In recent years, a growing body of data has been gathered in support of the idea that the mind is situated and embodied and that cognition is grounded in sensory-motor interactions with the world (Varela et al., 1991; Clark, 1998; Barsalou, 1999, 2008; Lakoff and Johnson, 1999; Gibbs, 2006; Spivey, 2007; Chemero, 2009). The guiding tenet of the embodied cognition (EC) movement holds that cognitive processes are shaped and structured by the fact that an agent has a particular kind of body and is embedded in a particular kind of environment. Crucially, the effects of embodiment can and should be observed at all levels of cognitive processing, from vision and memory (Glenberg, 1997; Noë, 2004; Proffitt, 2006), to emotion and action perception (Rizzolatti and Craighero, 2004; Niedenthal et al., 2005), to language and abstract thought (Barsalou, 1999; Lakoff and Johnson, 1999; Feldman, 2006; Gibbs, 2006; Barsalou, 2008). It has been argued that this "body-up" approach to cognition poses a serious challenge to more traditional "mind-down" approaches in cognitive science (Lakoff and Johnson, 1999; Spivey, 2007; Barsalou, 2008; Chemero, 2009), which have attempted to define cognition in terms of discrete, amodal, symbolic information-processing mechanisms divorced of any particular physical instantiation (Fodor, 1975; Marr, 1982; Kemp and Tenenbaum, 2008).

This debate has been particularly contentious in discussions of high-level cognition, where the amodal symbolic view has typically dominated. As a result, the embodiment of metaphor and abstract thought has become one of the most hotly researched, discussed, and debated issues within cognitive science (Lakoff and Johnson, 1980, 1999; Gibbs, 1994, 1996, 2006; Murphy, 1996, 1997; Boroditsky, 2000; Boroditsky and Ramscar, 2002; Feldman, 2006; Pinker, 2007). Lakoff and Johnson (1980, 1999) famously observed that natural language is exceedingly figurative. When we talk about complex or abstract topics, we rely heavily on systems of metaphors, borrowing words and phrases from other, more concrete domains. For example, to talk about *theories*, people often rely on *building* metaphors. Indeed, theories must have a *solid foundation* and be *well-supported* by the data or they might *fall apart*, and you can *build them up, tear them down*, or even *explode them* in light of new findings.

While traditional theories of language treat metaphor as mere ornamental flourish (e.g., Grice, 1975; Searle, 1979; Pinker, 2007), Lakoff and Johnson (1980, 1999) argue that metaphor is not simply the way we *talk* about abstract things, but how we *think* about them as well. On this view, we understand and reason about abstract domains like theories, time, and love through our concrete, embodied experiences (e.g., of interacting with physical buildings). Thus, our perceptual and motor experiences actually structure our ability to engage in abstract thinking. Empirical demonstrations of embodied metaphor have taken the form of experiments showing that activating a concrete source domain (e.g., space) influences responses and inferences in the abstract target domain (e.g., time; Boroditsky and Ramscar, 2002; Casasanto and Boroditsky, 2008; Jostmann et al., 2009; Ackerman et al., 2010).

One important challenge facing researchers is to account for this view of metaphorical thought at a more precise, mechanistic level of description (Murphy, 1996, 1997; Barsalou, 2008). This may

be particularly problematic because EC is not a singular, unified framework, but rather a collection of heterogeneous viewpoints that may be only loosely related to one another in terms of theoretical commitments and empirical investigation (Wilson, 2002; Ziemke, 2003; Gibbs, 2006; Barsalou, 2008; Chemero, 2009). In addition, because these competing perspectives are commonly described only verbally, it can be difficult to use them to generate the precise predictions that might allow us to directly compare them (but see Lakoff and Johnson, 1999; Feldman, 2006).

Taking a computational modeling approach may provide a potential remedy to these issues. The development of specific, simplified models can help researchers understand complex cognitive processes at a level of detail that theories of EC have sometimes lacked (see, e.g., Broadbent, 1987; Smolensky, 1988; Hintzman, 1991; Seidenberg, 1993; Barsalou, 1999, 2008; Spivey, 2007; McClelland, 2009). The process of constructing a model differs from a verbally described theory in that it forces the researcher to commit at least temporarily to a particular *internally consistent* instantiation of the environment and the agent that acts within it. As a result, computational models can make precise predictions that can be tested empirically. Grounding empirical findings in terms of a model and making principled modifications to that model in order to accommodate these findings can help researchers explore and clarify ideas (McClelland, 2009). In addition, because models can often reveal principles that underlie a given set of phenomena, modeling frameworks can sometimes help unify various areas of empirical inquiry (Estes, 1955; Rescorla, 1988; McClelland et al., 1995; Ramscar et al., 2010). This special topic of Frontiers in Cognition is evidence that more researchers are starting to take computational modeling seriously as a method for exploring the principles and mechanisms that support EC (see Spivey, 2007 for a call to arms on this issue).

At the same time, the findings from EC outlined above provide computational modelers the opportunity to look to for evidence of the ways in which cognition naturally unfolds in a real, embodied agent (for a recent review, see Barsalou, 2008). This will strongly influence not only the details of the model environment, but also the choice of the learning problem to be solved by the model. Modelers focused on understanding learning processes should attend to the fact that the information reaching the cognitive system is always structured by the relationship between the organism and its environment, which may lead to surprising new ways of thinking about everything from visual perception (Noë, 2004) to semantics (Barsalou, 1999).

The present paper has both a narrow and a broad goal. The narrow goal is to capture the effects of embodied conceptual metaphor using a connectionist model. In lieu of instantiating a particular EC theory of metaphor, we repurpose an existing connectionist model of semantic cognition (Rogers and McClelland, 2004) to explore how our experience of space can structure how we think and reason about time. This approach may be especially fruitful because it promises to bring together more established modeling principles with the novel findings from EC.

Our network receives direct experience with both space and time in its simplified environment, including experience that is analogous to the use of linguistic or cultural cues. However, the network's experience in the spatial domain is more richly structured than its experience in the temporal domain, in much the same way that we can freely move around and interact with our spatial, but not temporal, environment. Because the model is sensitive to the ways in which the structure of time is similar to the structure of space, it develops representations of time that are partially constituted by its knowledge of space. Therefore, even in the absence of direct co-occurrence of space and time during learning, the network is able to exploit this structural similarity to draw inferences about temporal events by using what it knows about space. This demonstrates a novel learning mechanism that operates over the course of development and gives rise to deeply metaphorical semantic representations, which may serve as a tractable implementation of existing theories of metaphorically structured thought (e.g., Lakoff and Johnson, 1980; Boroditsky, 2000).

The broad goal of this paper is to serve as an example of how computational models and EC can reciprocally inform one another. In particular, we make the case that connectionist architectures can help explain many surprising results from the EC literature (for related views, see Bechtel, 1997; Clark, 1998; Spivey, 2007). Crucially, our model focuses on the learning process and forms overlapping, distributed representations, which have exactly the properties required by many embodied accounts of cognition. In particular, these representations, together with the learning process, support the integration of experience from multiple modalities, including perceptual-motor, linguistic, and cultural information. At the same time, extending the scope of the model to incorporate insights from EC transforms our interpretation of the modeling approach more generally. This can lead to new ways of thinking about how to set up and investigate particular ideas within this modeling framework. Ultimately, we suggest that this integrative approach can serve as a unifying framework that may help drive future progress within cognitive science.

MATERIALS AND METHODS

GENERAL MODELING FRAMEWORK

The network can be thought of as an agent experiencing its world. Over the course of "training" the agent repeatedly experiences events in the world, predicts their outcomes, and learns something about how the actual events differ from its predictions. The environment and the agent are simplified so as to render the learning process tractable, while still retaining those aspects of environmental structure which are crucial for producing the phenomena the model is supposed to explain, and to make it possible to analyze what the agent has learned (for a discussion of this issue, see McClelland, 2009).

In this model, the environment consists of the various items in the world that the agent experiences in their various relational contexts (collectively forming the input patterns), together with the subsequent states of the world that the network attempts to predict (the target output patterns). The network that comprises the agent is wired up in a strictly feed-forward fashion, as shown in **Figure 1**. While we assume that in reality agents interact with the world in a dynamic fashion, for simplicity we consider only one portion of this dynamic interaction. On each trial, the agent experiences some portion of the world (e.g., that it is standing in a particular section of space and moving in a particular direction), makes a prediction about what it will experience next (e.g., that

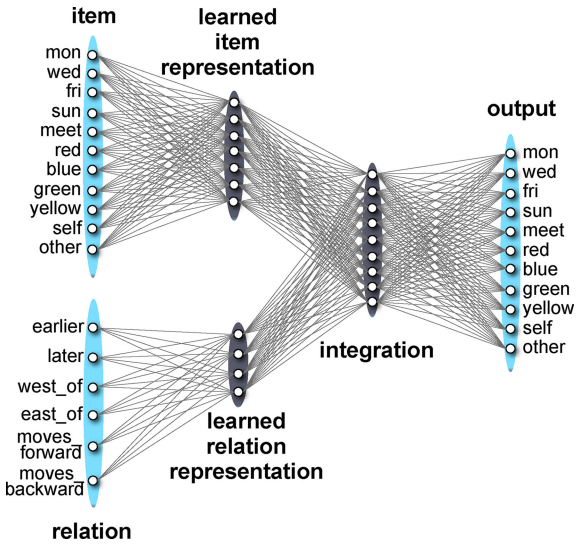


FIGURE 1 | Diagram of network architecture.

it will encounter another particular section of space), and learns about the ways in which it was incorrect, thereby improving future predictions.

The network's knowledge is stored in the weights between the layers. When a pattern of activation occurs across one of the layers, that activation propagates forward through the weights to the next layer. The patterns of activation at the input layers are thought of as multimodal sensory-motor input from the environment. In the *Item* layer, these inputs stand for the experience of physical locations in space and temporally extended events such as the days of the week or a meeting. In the *Relation* layer, the inputs stand for different kinds of relationships that these items can have to each other; for example, we might ask the network what day is *earlier than Wednesday*, or what section of space is *West of* the *blue* section.

While these layers consist of labeled units, they are best thought of as standing for distributed representations that were learned from other, lower-level (possibly modality-specific) patterns of perceptual-motor experience. This simplification does not strongly affect how the model works because the network is forced to create its own distributed representations of these perceptual inputs in the layers that immediately follow (see, e.g., Rogers and McClelland, 2004). In particular, the *Learned Item Representation* is a re-representation of the *Item* inputs, integrating all of the information it has learned across all relations to create a densely overlapping set of patterns that encode the structural regularities that hold between the items. The *Learned Relation Representation* serves the complementary function for the *Relation* inputs. Activation in these layers then propagates forward to the *Integration* layer. Here, information about the two input pathways is combined in a way that we presume is similar to how modality-specific information is integrated at earlier layers. This integrated representation is used to make a prediction about the target pattern, which is represented by activations of the *Output* layer. In the current model, the target pattern consists of another item (or set of items) that bears the appropriate relation to the input.

Initially, the network is instantiated with small random weights connecting each of the layers. As a result, its internal representations of all items and all relations will be similar, and therefore its predictions about the world will be the same for all inputs. Whenever the network's output fails to match the target pattern, however, it receives an error signal in proportion to the squared output error. This error signal informs the network both when it has predicted events that do not occur and when it has failed to predict an event that did occur. In practice, this error signal serves to adjust the weights from the inputs to the outputs in proportion to the error that they caused, using the standard backpropagation learning algorithm.

Since different input patterns predict different events "in the world," the network will gradually learn to differentiate the items from each other, and the relations from each other. This process of differentiation is driven by differences in what the various items predict about what else may happen in the world, not directly by, for example, the degree of overlap in the perceptual inputs (for related views, see Gibson and Gibson, 1955; Rogers and McClelland, 2008; Ramscar et al., 2010). However, wherever there is similarity between different items, these similarities will be encoded in the learned, distributed representations. The "similarity," as we will show, can be similarity either in the explicit overlap between their predictions or in the systematic structural relationships among the various items within a domain. These internal representations therefore capture, in a graded and sub-symbolic fashion, both the similarities and the differences between the items. In the simulations that follow, we examine whether this framework can account for some of the empirical findings from the conceptual metaphor literature. In order to motivate the simulations, we begin with a discussion of a specific example of conceptual metaphor.

A CASE STUDY OF EMBODIED CONCEPTUAL METAPHOR: TIME AS SPACE

One the best documented cases of how abstract thinking can be metaphorically structured by concrete experience comes from the domain of time (Clark, 1973; Lakoff and Johnson, 1980; McGlone and Harding, 1998; Boroditsky, 2000, 2001; Boroditsky and Ramscar, 2002; Gentner et al., 2002; Evans, 2004; Matlock et al., 2005; Casasanto and Boroditsky, 2008). The language we use to talk about time is heavily infused with the language we use to talk about space, as when we talk about a *long* meeting or two birthdays being *close* together (Clark, 1973; Lakoff and Johnson, 1980). Consistent with the EC perspective, our actual perception of space can influence how we experience and reason about time (Casasanto and Boroditsky, 2008). For example, Casasanto and Boroditsky (2008) found that the length (in spatial extension) of a line on a computer screen affected how long (in temporal duration) it was judged to remain on the screen: the longer the line, the longer the time.

Like many other abstract, complex domains, there is more than one system of metaphor for talking and thinking about time (Clark, 1973; Lakoff and Johnson, 1980; Gentner et al., 2002). For instance, we can imagine ourselves moving forward through time, like when we talk about *coming up on* the holidays (ego-moving perspective), but we can also imagine remaining stationary as time moves toward us, like when we talk about the holidays *fast approaching* (time-moving perspective). Some spatial words that we use to talk about temporal events are ambiguous because they can be interpreted

differently depending on which metaphorical perspective is adopted. For example, if you are told that Wednesday's meeting has been moved *forward* 2 days and you had adopted the ego-moving perspective, you would conclude that the meeting is now on Friday. However, if you had adopted the time-moving perspective you would conclude that the meeting is now on Monday (McGlone and Harding, 1998; Boroditsky, 2000; Boroditsky and Ramscar, 2002). Several experiments have demonstrated that the way people are currently thinking about space directly affects which of these perspectives they select and therefore how they reason about time (Boroditsky, 2000; Boroditsky and Ramscar, 2002). For example, people who are asked the Wednesday's meeting question at an airport are more likely to take the ego-moving perspective (i.e., give the Friday response) because they are about to take a flight (i.e., move through space) than when they are waiting to pick someone up (i.e., someone is approaching them in space; Boroditsky and Ramscar, 2002).

These findings suggest that we automatically use our online representations of space to structure our thinking about time. Why might this be the case and what mechanisms support this process? Researchers have highlighted at least two rich sources of information in our experience that could give rise to the metaphorical mapping between time and space. First, time and space co-occur in meaningful ways in our experience moving and acting in our environment (Lakoff and Johnson, 1980; Boroditsky, 2000). For instance, walking a longer distance typically takes a longer amount of time. Second, the structure of our linguistic experience, including the specific spatial metaphors we use as well as features of the language such as writing direction, might also influence how the concept of time is structured in terms of space (Boroditsky, 2000, 2001; Boroditsky and Gaby, 2010; Fuhrman and Boroditsky, 2010). For example, in both explicit event sequencing tasks and implicit temporal judgment tasks people represent time as progressing from the past to the future in a manner consistent with the writing direction of their language (Fuhrman and Boroditsky, 2010). In the following simulations we use the general modeling framework described above to explore how these metaphorical mappings may emerge gradually over the course of learning. Unlike previous proposals about the emergence of metaphor over developmental time, the mappings are not driven primarily by direct co-occurrence, but by the similarities in the structural regularities across domains.

MODEL SIMULATION 1

In the first simulation, the network learns about space and time through experience trying to predict how space and time are structured in the model environment (see **Table 1** for detailed model specifications). In the simplified environment of the model, space is laid out along a single dimension running from *West* to *East* (unlike our own environment, in which space is three-dimensional and also includes north and south, up and down! See **Figure 2**). To make the simulation easier to talk about and understand, space is divided into sections of different colors, going from *red* to *blue* to *green* to *yellow* as you move toward the *East*. Throughout the course of training the network will attempt to learn that two relations – *East of* and *West of* – structure the spatial arrangement of the colored sections in the environment. Training proceeds by asking the model to predict what color section of space it will "see" if it looks toward the *East* or *West* of its current position (and adjusting the weights in proportion to the error of this prediction, as described above). In practice, this works by presenting the network with one or more items along with a relation in the input layers and asking it to generate all appropriate outputs. For instance, if the network were presented with *blue* and *West of* it would have to output *green* and *yellow* (i.e., what the network would see if it looked toward the *East* while standing on the *blue* section: the sections of space that the *blue* section lies to the *West of*).

Time is also laid out in a single dimension from *earlier* to *later* events. Time is divided up into distinct moments, the days of the week, which follow a specific temporal sequence (going from

Table 1 | Detailed simulation parameters.

	Sim 1	Sim 2
LAYERS (# UNITS)		
Item	11	10
Relation	6	6
Learned item representation	7	7
Learned relation representation	4	4
Integration	9	9
Output	11	10
OTHER PARAMETERS		
Initial weight range (–/+)		–0.05/0.05
Activation function		Sigmoid
Error measure		SSE
Learning rate		0.1
Momentum		0

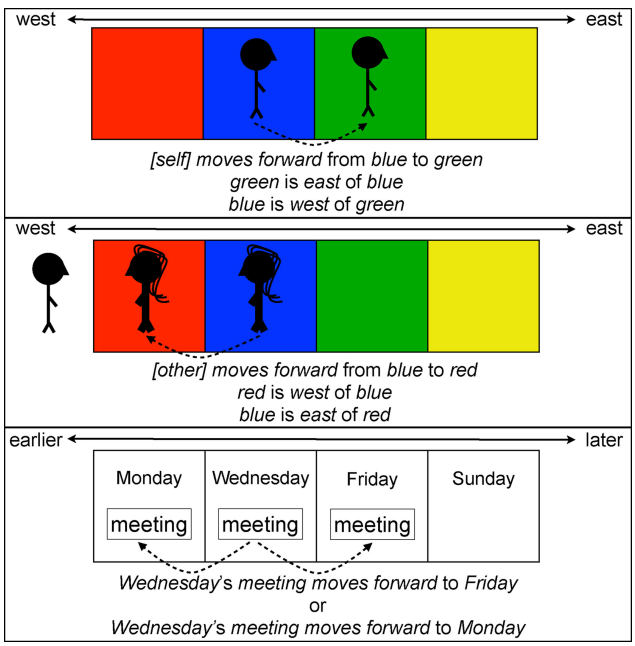


FIGURE 2 | A diagram illustrating the structure of spatial and temporal relations in the model environment. The network learns about the consequences of both itself and other agents moving in the environment, though movement in the temporal domain is ambiguous.

Monday to *Wednesday* to *Friday* to *Sunday* as you progress *later* in time). During training the network will attempt to learn that two relations – *earlier than* and *later than* – structure the temporal sequence of the days of the week. The network learns about time in the same way that it learns about space. Thus, if the network were presented with *Monday* and *earlier than* as inputs it would have to generate *Wednesday*, *Friday*, and *Sunday* as outputs (i.e., the days of the week that Monday is *earlier than*).

Crucially, the network enjoys a richer, more structured set of experiences in the spatial domain because it can observe the consequences of its own movements in space (as well as the consequences of the movements of other agents in the environment). We can imagine that, like most mobile organisms, the network has both a front and back and can move forward and backward in space. To keep things simple, let us imagine that the network is standing on the *blue* section of space facing toward the *East*. If the network *moves forward*, it will move toward the *East* end up on the *green* section of space, while if it *moves backward* it will move toward the *West* and end up on the *red* section of space. However, *forward* and *backward* movements in space are not simply the same as moving toward the *East* or *West*. Now imagine that the network is observing another agent in the environment that is standing on the *blue* section of space and facing *West*. If this other agent *moves forward*, it will move toward the *West* end up on the *red* section of space, while if it *moves backward* it will move toward the *East* and end up on the *green* section of space. Thus the network has to learn that the consequences of moving *backward* and *forward* in space depends on whether it is attending to its own movements or to the movements of another agent. In practice, this works by including *self* and *other* items in the input layer to let the network know whose movements it is observing (see **Figure 2**). To keep things simple, we assume that the model is always facing toward the *East* and the *other* agent in the environment is always facing toward the *West*.

While the effects of movement in the spatial environment are unambiguous in the presence of either the *self* or *other* context, the model's experience of "movement" in the temporal domain is ambiguous in that there is no consistent mapping between *forward*/*backward* and *earlier*/*later*. The model learns that when a *Wednesday* meeting *moves forward*, it sometimes is moved to *Monday*, and other times it is moved to *Friday*. The same can occur when a meeting *moves backward*. Structuring the temporal domain in this way allows us to study to the ambiguity explored in Boroditsky (2000) and Boroditsky and Ramscar (2002). In particular, while the network has no experience with the *self*/*other* distinction in the temporal domain, we can examine whether it can use its experience with the effects of these contexts in the spatial domain to resolve the ambiguity of "movement" in the temporal domain. That is, we can test whether activating a particular spatial frame of reference (i.e., the *self* or *other* perspective) in the context of reasoning about a temporal event (i.e., moving the *Wednesday* meeting *forward*) will influence the network's expectations about the effects of "movement" in the temporal domain.

MODEL SIMULATION 2

The first simulation investigated whether the network would learn to metaphorically map its relatively rich experience with space onto the parallel but experientially impoverished domain of time in order to resolve an ambiguous temporal reasoning task. In Simulation 2, we explore whether the network can learn to map the directionality of time (from *earlier* to *later*) onto other spatial cues in the environment (e.g., the directionality of space, from *West* to *East*). Several studies have demonstrated that culturally specific spatial cues, such as writing direction (Fuhrman and Boroditsky, 2010) and absolute spatial coordinate systems (Boroditsky and Gaby, 2010), can influence and structure how people think about the directional "flow" of time.

The model was set up in a very similar manner as in Simulation 1. However, where Simulation 1 included an ambiguity in the temporal domain, Simulation 2 removes that ambiguity in order to closely align the meanings of the temporal and spatial relations. In particular, the *moves forward* relation was made unambiguous in both the spatial and temporal domains, by removing the *other* item. In the temporal domain, *moves forward* always predicted that the *Wednesday meeting* should occur on *Friday*, never *Monday,* and *moves backward* always predicted that *Wednesday* meetings should occur on *Monday*, never *Friday*. This might be interpreted as a culturally specific bias, analogous to the experience of reading temporally sequenced material like calendars and comics from left to right (or even writing direction itself, see Fuhrman and Boroditsky, 2010). In the spatial domain, we removed the patterns in which the *other* agent *moves forward* from *blue* to *red* and *moves backward* from *blue* to *green*, again rendering the situation unambiguous. Removing these four patterns, two from the temporal domain and two from the spatial domain, leaves the *moves forward* relation consistent with the *earlier than* relation in the temporal domain and with the *West of* relation in the spatial domain. This can be seen in the predicted outcomes of the events: *moves forward* from *Wednesday* predicts *Friday*, and *Wednesday* is *earlier than Friday*, and so on for the spatial domain.

If the model is sensitive to the structural similarities present in this environment, it should learn that the *West of* and *earlier than* relations make similar predictions in their respective domains, as do the *East of* and *later than* relations. As a result, the learned distributed representations for these pairs of relations should become similar as a function of experience – allowing, for instance, spatial words like *East of* and *West of* to be sensibly interpreted in the temporal domain (e.g., *Wednesday* is *East of Monday* or *Wednesday* is *West of Friday*). The model only ever observes *West of* and *East of* in the spatial domain, and *earlier than* and *later than* in the temporal domain, so an account based on direct co-occurrence would not generate the same prediction. This would provide a demonstration of how culturally specific features of the environment such as writing direction or dominant spatial coordinate systems could come to organize our representations of abstract domains such as time.

RESULTS

SIMULATION 1

In the first simulation we explored whether an ambiguity in the temporal domain (i.e., that a *Wednesday* meeting sometimes *moves forward* to *Monday* and sometimes *moves forward* to *Friday*) can be resolved by activating a particular spatial frame of reference. That is, even though the model has no experience with the *self*/*other* distinction in the temporal domain, we can nevertheless activate one of these spatial frames of reference in the temporal domain when asking the model whether

it thinks *Wednesday* meetings *move forward* to *Monday* or *Friday*. If these reference frames influence the model's interpretation of *moves forward* in a way that is consistent with empirical results (Boroditsky, 2000; Boroditsky and Ramscar, 2002), we would expect that including *self* as an input (along with *Wednesday*, *meeting*, and *moves forward*) would yield relatively more activation in the *Friday* output unit than the *Monday* output unit. Alternatively, we would expect that including *other* as an input instead would result in relatively more activation in the *Monday* output unit than the *Friday* output unit.

To investigate this, we exposed the network to 10,000 epochs of training in the simplified environment, at which point we froze the weights to prevent further learning and began the testing phase. The statistics reported for both simulations include activation values that have been averaged across 10 instances of the model to ensure that any effects are not the result of a random bias in a particular instance. First, we tested whether the network had learned the unambiguous spatial and temporal structure of its environment by presenting it with the same input–output pairings that it was trained on. Indeed, the network performed quite well on this test (mean tss = 2.31, SD = 0.32)[1], demonstrating that it had correctly learned the features of its environment that it had been directly exposed to during training. Next we tested whether including a spatial frame of reference (i.e., *self* vs. *other*) influenced the network's predictions for the effect of *moves forward* in the temporal domain. We measured this effect by comparing three test patterns: (1) the ambiguous pattern that the network was trained on in which the *Wednesday* and *meeting* items were paired with the *moves forward* relation, (2) this same pattern with the *self* item included as an input, and (3) this same pattern with the *other* item included as an input (instead of *self*).

The model learned that *moves forward* was ambiguous in the temporal domain when the *self* and *other* items were not included as inputs. Specifically, when tested on the ambiguous pattern, the model fully predicted *meeting* on the output (mean = 0.981, SD = 0.0043) and partially predicted both *Monday* (mean = 0.500, SD = 0.0075) and *Friday* (mean = 0.503, SD = 0.0069). No other units had average activations greater than 0.02. A regression model that predicted output activation of the two target units (*Monday* and *Friday*) with contrast-coded predictors for Day (*Monday*, *Friday*) and Perspective (*self*, *other*) as well as a Day × Perspective interaction term, was fit to the two test patterns. Both main effects were significant (Day: $\beta = 0.062$, $p < 0.05$; Perspective: $\beta = -0.100$, $p < 0.01$) as was the interaction term ($\beta = 0.081$, $p < 0.01$), indicating that including the perspective units shifted the degree to which the model predicted that *Wednesday's meeting* would *move forward* to *Monday* or *Friday* (see **Figure 3**).

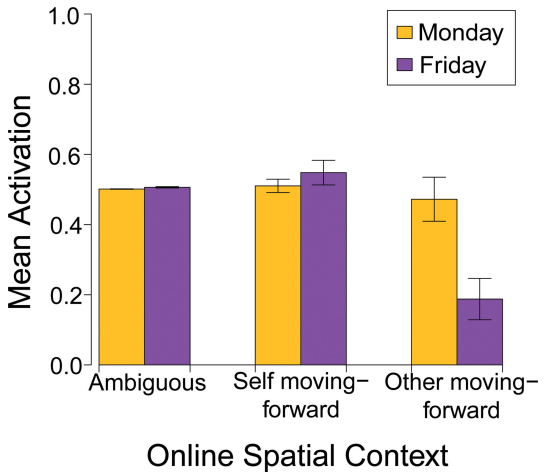


FIGURE 3 |The results of Simulation 1 showing that activating a particular spatial frame of reference biases the network's predictions about movement in the temporal domain.

SIMULATION 2

In the second simulation, we explored whether the network could in principle learn to map the directionality of time (from *earlier* to *later*) onto the directionality of space (from *West* to *East*). In order to clearly explore this possibility, we modified the model's environment slightly from that of Simulation 1 so that the *moves forward* relation was consistent with *later than* in the temporal domain and *East of* in the spatial domain (and so that *moves backward* was consistent with *earlier than* and *West of*). If the network is able to take advantage of this similarity in a way that is consistent with empirical findings (e.g., Fuhrman and Boroditsky, 2010; Boroditsky and Gaby,2010), then it should be able to interpret, for example, the relations *East of* and *West of* in the domain of time (i.e., *Wednesday* is *East of Monday* and *Wednesday* is *West of Friday*) even though these relations were never explicitly paired with temporal inputs or outputs in training.

We investigated this issue by exposing the network to 10,000 epochs of training and then freezing the weights. At this point the network had learned to make correct predictions for each of the training patterns (mean tss = 0.0898, SD = 0.0375). We then presented the network with two novel test patterns: one pairing *Wednesday* with *East of* on the input, the other pairing *Wednesday* with *West of* on the input.

When *Wednesday* was paired with the *East of* relation, the model predicted *Monday* (mean = 0.525, SD = 0.421) more than *Friday* (mean = 0.022, SD = 0.046), whereas when *Wednesday* was paired with the *West of* relation, the model predicted *Friday* (mean = 0.324, SD = 0.387) more than *Monday* (mean = 0.098, SD = 0.306). A regression model that predicted output activation with contrast-coded predictors for Day (*Monday*, *Friday*) and Relation (*East of*, *West of*) as well as an interaction term was fit to the two test patterns. Neither main effect was significant (Day: $\beta = 0.032$, $p = 0.54$; Relation: $\beta = -0.069$, $p = 0.19$); however, the Day × Perspective interaction was significant ($\beta = -0.182$, $p < 0.01$), indicating that these relation units held meaning in the domain of time even though this was the first time the model had encountered them in the temporal context (**Figure 4**).

In order to determine why this effect occurred, we submitted the representations for each of the relations of interest (i.e., *West of*, *East of*, *earlier*, *later*, *moves forward*, and *moves backward*) in the *Learned Representation Layer* to a hierarchical clustering analysis (shown in **Figure 5**). This analysis shows that the representation of *West of* is very similar to the representation of *earlier*, and the representation of *East of* is similar to the representation of *later*. It also shows that *moves forward* is more similar to *West of* and *earlier*, and *moves backward* is more similar to *East of* and *later*.

[1]The imposed ambiguity in the temporal domain (i.e., that *Wednesday's* meeting can move both *forward* and *backward* to *Monday* or *Friday*) made it impossible for the network's tss to improve beyond 2.

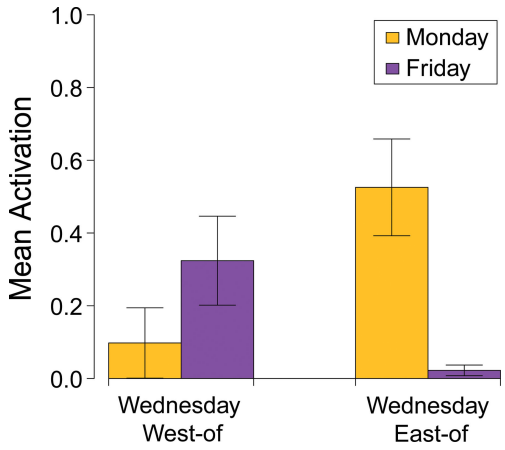


FIGURE 4 | The results of Simulation 2 showing that spatial relations that were never directly experienced in the temporal domain can still structure temporal reasoning.

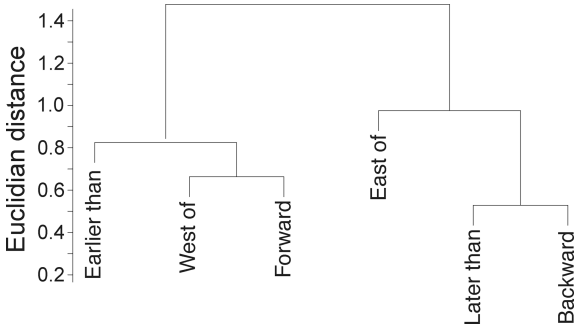


FIGURE 5 | The structure of the *learned relation representation* layer in Simulation 2.

SUMMARY OF FINDINGS

Both model simulations successfully learned a representation of the temporal structure of the world based partially on their experience with space. In Simulation 1, this allowed the network to resolve an ambiguity in the temporal domain by relying on additional structure only present in the spatial domain: a true application of conceptual metaphor to aid cognition. In Simulation 2, the network's representations of the spatial and temporal domains were shaped by a structural homology between the domains: in this case, a "culturally driven" bias to scan from *West* to *East* through time. In both cases, the network's metaphoric concepts were not driven by direct co-occurrence between concepts within the domains (e.g., distance with duration). Rather, the available information for learning the metaphor was the second-order relations between items within each domain (e.g., things move around in space in a similar way to how events can be sequenced in time).

DISCUSSION

WHY THE MODEL WORKS

Several theorists have proposed that the grounding of abstract thought in concrete knowledge may be due in part to the direct co-occurrence of certain domains in experience, for example, of time with space or of love with physical warmth (Lakoff and Johnson, 1999; see also the afterward to the 2003 edition of Lakoff and Johnson, 1980). The model presented here demonstrates another, more general, yet equally grounded pathway to metaphor: structural similarity between the target and the source domain (see also Boroditsky, 2000). However, this use of structural similarity is not a distinct online, rule-based algorithm operating over symbolic representations, as in other theories of structural alignment in metaphor comprehension and analogical reasoning (e.g., Falkenhainer et al., 1989); rather, it is a result of the gradual process of differentiation that takes place over the entire course of learning. It is fair to say that the network's knowledge of time is partially constituted by the learned structural relations in the spatial domain. This is demonstrated by the metaphoric remapping between time and space, which, in this model, share almost no input (only the *moves forward* and *moves backward* units) and no output units at all. Merely having distributed representations, as most connectionist models do, is not sufficient for this kind of behavior to emerge; the process by which those representations were acquired through experience with the environment is also critical.

To understand why the remapping occurs, recall that the model initially treats all items and relations as equivalent (due to its small and random initial weights) and only discriminates objects as it is forced to do so by the flow of information from the world. Over the course of this differentiation process, the model constructs several high-dimensional and highly overlapping representations for the items, the relations, and the item–relation conjunctions, all passing through the same sets of weights and patterns of activation over the same sets of units. If the network can reuse certain dimensions of its representations because of similarity in the structural relationships between and among items and contexts, it will tend to do so. Since the spatial and temporal domains share most, though not all, of their respective relational structure, the network learns a partially unified representation for the two domains. These overlapping representations, which are a direct result of the differentiation process, give rise to the influence of the concrete spatial experience on the abstract temporal reasoning task.

This model demonstrates that the structural homology between domains of experience is one aspect of the environment that can drive generalization (or, more properly, partial lack of differentiation) for metaphorical inference. But it is not the only way that metaphors can be learned. As mentioned above, co-occurrence of more abstract with more concrete domains of experience may also cause the learner to build metaphorical semantic representations. This is because the experience with the abstract domain will often predict properties that are also predicted during experience with the concrete domain, which may drive the representations to become more similar than they would otherwise be. In Simulation 2, it is indeed co-occurrence that drives learning, but it is indirect, not direct, co-occurrence. The *moves forward* unit in this simulation is unambiguously similar to *West of* when it occurs in the spatial domain, and to *earlier than* when it occurs in the temporal domain. Notice that *West of* and *earlier than* never predict similar outputs in a way that would cause them to become similar, so this is not a matter of raw co-occurrence. Still, the model is encouraged to draw its representations of *moves forward*, *West of*, and *earlier than* into a similar semantic structure because these relations must be

re-represented in the *Learned Representation Layer*. In a parallel fashion, *moves backward* draws together the cross-domain relations *East of* and *later than*. The bridging between structures can occur because of similarity in the structural relationships among the items within each domain, or because of some direct or indirect co-occurrence (or co-prediction) in the environment, or (as we believe is probably the case in most natural settings) both.

Another mechanism that drives metaphorical structuring is language use. There are at least two routes through which language can bring about metaphoric alignment, one slow and one fast. In the current framework, linguistic experience is considered to be another aspect of the environment (this point is discussed in more detail below). The influence of language here would be across the (slow) course of development, serving as additional scaffolding for similar high-order structures (as in Simulation 1) or as an indirect co-occurrence cue or outcome (as in Simulation 2). On the other hand, language might be used online to point out a novel metaphorical structural mapping, such as "an atom is like the solar system." The agent's task is then to take two existing semantic structures and figure out what aspects of those structures the conversational partner intends to highlight. Our model deals with a very slow process of learning and differentiation, but does not have a way of rapidly integrating new information, so this kind of novel metaphorical language is a problem in this model. However, we are not claiming that our model describes the whole story, and any model of learning like ours will eventually need to take into account fast-learning processes as well (McClelland et al., 1995). Our model is nevertheless a novel contribution to the literature, as existing models of metaphor (and analogy) that do deal with online structural alignment (e.g., Falkenhainer et al., 1989) do not attempt to slowly integrate structural information over the course of development.

The possibility that speakers of different languages might categorize or even perceive the world in different ways has been a focus of scrutiny in recent work (e.g., Boroditsky et al., 2003; Majid et al., 2004; Winawer et al., 2007). One might expect that if embodiment holds, then the environment itself would fully determine the semantic representations possessed by the agents within that environment (and therefore language use would not really have any effect on conceptual representation; for a related position, see Gleitman and Papafragou, 2005). While this viewpoint recognizes the importance of the statistics of the environment in semantic learning, it fails to appreciate that linguistic information is itself another rich source of environmental statistics. The modeling approach described here provides a principled way of integrating the effects of language on cognition with EC (see also Dilkina et al., 2007; Boroditsky and Prinz, 2008; Andrews et al., 2009).

In our modeling framework, the key to this integration is to allow the network to experience a linguistic relational context alongside contexts conveying other kinds of perceptual and motor information. The network integrates information for each item across many different relational contexts, though these are limited to a fairly small set of physical and temporal relations. However, in other related models the contexts can take on a much broader meaning (e.g., Rogers and McClelland, 2004; Thibodeau et al., 2009). For example, in the semantic cognition model of Rogers and McClelland (2004, 2008), contexts include the Quillian-like *is-a*, *is, can*, and *has* relations. These relations may be thought of as different kinds of world contexts within which the items might be encountered. For example, when first observing a robin, one might notice that it is red; when attempting to catch a robin, one might observe that it can fly away; and when discussing robins with other people, one might be informed that a robin is a bird and an animal. If language is used to describe things in the world, we would expect that the relational context within which linguistic information is acquired should bear some structural resemblance to relational contexts grounded in perceptual-motor experience. In this case, as described above, the structural information may shape the representations even across different relations, leading to just the effects of language on thought that have been shown in experimental work.

IMPLICATIONS OF THE MODEL FOR EMBODIMENT

We have presented this model as an exploration of effects in the embodied conceptual metaphor literature, and as having implications for theories of EC as a whole. Thus it is important to address the possible criticism that this model neither is embodied nor speaks to issues in EC.

For one, because the inputs and outputs of this network are labeled units, of which only a few are active at a given time, the model may appear to support a more classically symbolic approach to cognition than EC would endorse. Indeed, these representations are highly simplified and abstracted from realistic sensorimotor information. We make the claim that this is an acceptable simplification because falling back to a relatively more localist representation does not fundamentally change the way the model works. Simulations by Rogers and McClelland (2004) using a similar model demonstrated that replacing the localist input units with a distributed input representation (e.g., of the visual features of animals, rather than their names) did not affect the model's performance in any significant way. There is reason to expect this result, since the model is not allowed to manipulate these localist inputs in any direct fashion; rather, as we noted earlier, it is forced to create its own distributed representations of these inputs in the layers that immediately follow. Our inputs may look like "linguistic" rather than "sensorimotor" representations precisely because they are localist, and many in the field think of linguistic units as localist symbols. While we do not exclude linguistic information as part of the experience relevant to the time/space effects (and neither do most researchers in the field), we do intend our model to stand for the entire space of experiences available to the agent.

Of course, this localist/distributed argument is somewhat distinct from the question of whether our training patterns accurately reflect the sensorimotor inputs to an agent, which in this case they do not and cannot. Even distributed representations would have to be greatly simplified and abstracted relative to the enormous flow of information that continuously impinges on the sensory receptors of any biological agent. Better input and output representations are surely possible. One promising approach would be to use unsupervised learning mechanisms such as the deep belief networks of Hinton and Salakhutdinov (2006) in order to extract distributed representations from more ecologically valid datasets. As this is not the focus of the current research, we used minimally distributed item representations, which nevertheless allowed us to capture the learning processes of interest.

For all that we believe that our simplifications are both justified and necessary, it might still be argued that the cumulative effect is to render the model "disembodied." However, to the extent that this model has consequences for EC, we would characterize it as a model of embodiment. We take our model as a kind of metaphor (as all models ultimately are), which points to a certain kind of statistical learning process that could help explain many of the results in the EC literature. We show that the statistical support available for learning environmental regularities is much stronger than the raw co-occurrence-based mechanisms previously proposed. Importantly, this helps explain how simple learning mechanisms (of the sort that may be plausibly instantiated by perceptual and motor brain regions) can give rise to "higher-level" cognitive processes such as conceptual metaphor. This is an example of the sort of back-and-forth engagement between connectionist and embodied approaches that we hope to foster in this paper. By situating our model in the EC perspective, we provide stronger support for the validity of the EC approach in general, and in particular, for the generality of the learning mechanisms that underlie many embodied theories.

FUTURE DIRECTIONS

The current model could be improved upon by including a more ecologically valid environment structure and set of training data. At present, we have made several simplifying assumptions that do not realistically map onto the ways in which humans experience space and time. For example, the network receives the same amount of experience moving forward and backward in space, only ever faces in one direction, and does not actually experience moving into every location in the surrounding space. While these simplifications allowed us to more easily explore the mechanisms underlying a small number of relevant findings, future versions of the model could incorporate a more realistic environment structure based on empirical findings in order to generate more precise and accurate predictions and explanations.

In the process of further developing this model, it will become increasingly important to explore the relationship between the model and the way in which conceptual metaphor is realized in the brain (see also Lakoff and Johnson, 1999; Feldman, 2006). One way to approach this question would be to use the model to make predictions about how neurological damage should affect metaphorical knowledge or the ability to reason metaphorically. We believe this approach could be fruitful given that Rogers and McClelland (2004) used a variant of the model that we adapted for our simulations to understand the degradation of semantic knowledge in patients with particular patterns of neurological damage. This work has led to more biologically plausible models of semantic dementia that highlight the specific role multimodal integration layers in the anterior temporal lobes play in semantic representation (Rogers et al., 2004).

Research on conceptual metaphor suggests that the effects of damage to particular brain regions will depend on the metaphorical domain in question. As described above, the reason that the conceptual metaphor approach fits naturally within the wider scope of EC theories is that abstract knowledge is thought to be grounded in lower-level sensory and motor mechanisms. This view suggests, for example, that to understand metaphors that rely on mappings to the motor domain (e.g., "The reader *grasped* the ideas in the paper"), we would draw on neural mechanisms that support actual motor planning or execution. Recently, neuroimaging evidence has been gathered in support of this claim (Boulenger et al., 2009). Other work has shown that processing metaphorical language about movement in space (e.g., "Stock prices *soared*") is sufficient to adapt direction-selective perceptual circuits and lead to a visual motion after-effect (Toskos Dils and Boroditsky, 2010).

In light of these findings, we would predict that brain areas responsible for representing spatial experience would also be important for certain aspects of temporal reasoning. In fact, recent research has implicated parietal cortex in representing space, time, number, and other domains that involve magnitudes (for a review, see Hubbard et al., 2005; Bueti and Walsh, 2009). Other researchers have suggested that the cerebellum, which is important for coordinating fine motor movements and balance in space, might also play a role in representing the temporal aspects of linguistic processing (Oliveri et al., 2009). The results of our current simulations suggest that any brain networks that represent the structure of space or time in experience might play a role in these metaphorical processes. Future research will explore the relationship between the model and the brain more directly.

CONCLUSION

In the introduction we outlined both a narrow and a broad goal for the modeling approach described in this paper. The narrow goal, capturing embodied effects in conceptual metaphor using a connectionist model, has been described in some detail above. We would now like to return to the broader goal of showing how connectionist models and theories of EC can mutually inform one another, and how marrying these approaches can benefit cognitive science as a whole.

For one, we have demonstrated that it is both possible and useful for proponents of EC to engage with the rich literature on domain-general learning mechanisms for insight into how to construct models of their findings and generate testable, mechanistic theories. This approach promises to provide an implementation of many ideas that EC theorists have proposed. Lakoff and Johnson (1980, 1999) saw that the conceptual system is deeply metaphorical, and we can now understand why this might be the case for a particular kind of learner embedded in a particular kind of environment. Our model provides an illustration of how conceptual metaphor naturally emerges within a system that learns the statistical structure of the environment through progressive differentiation and stores its representations as distributed and overlapping patterns of activation.

Connectionists, in turn, can gain a new understanding of their own models by examining the empirical findings from EC. The model of semantic cognition we extended here was originally tested on a simple Quillian semantic hierarchy (Collins and Quillian, 1969; Rogers and McClelland, 2004), and showed the right patterns of learning to account for traditional ideas of conceptual development. However, as the EC critique has ably pointed out, the physical abilities and limitations of the agent provide an extremely powerful source of statistics that pervades the agent's interactions with all features of its environment (Noë, 2004; Gibbs, 2006). This observation transforms the implications of the semantic cognition model, allowing us to

think about it not as an observer gradually learning to construct a mirror image of the world inside its head, but as an active and embodied agent learning to predict the world around it. Similarly, embodiment effects, together with metaphorical overlap between learned contexts, provide new ways of thinking about traditional controversies. For example, language learning might be thought of as a task that occurs not in symbolic isolation, but within the broader context of learning to discriminate sounds in general, to produce sounds with the mouth in general, to predict the behavior of other agents in general, and so on. Thus the EC way of thinking seems to fit naturally into the domain-general approach to cognition that has been championed by connectionist researchers for decades.

Finally, we would like to suggest that this integrative approach to thinking about both EC and computational models might itself serve as a model for future research within cognitive science. We believe that we are now approaching a point in the evolution of cognitive science where various and diverse theoretical and experimental approaches can be usefully synthesized into a greater whole. Movement in this direction is ongoing (see, for example, the recent volume edited by Spencer et al., 2009) and we hope that this paper may serve as an additional nudge to the field.

ACKNOWLEDGMENTS

The authors would like to thank Lera Boroditsky and Jay McClelland for inspiration and helpful, critical discussion of the content of this paper and the issues it addresses. We would also like to thank Alexia Toskos Dils, Teenie Matlock, Dharshan Kumaran and Adriana Weisleder for additional help in thinking through the logic of our simulations and for comments on an earlier draft. This material is based on work supported under a National Science Foundation Graduate Research Fellowship and a Stanford Graduate Fellowship.

REFERENCES

Ackerman, J., Nocera, C., and Bargh, J. (2010). Incidental haptic sensations influence social judgments and decisions. *Science* 328, 1712.

Andrews, M., Vigliocco, G., and Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychol. Rev.* 116, 463–498.

Barsalou, L. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bechtel, W. (1997). "Embodied connectionism," in *The Future of the Cognitive Revolution*, eds D. M. Johnson and C. E. Erneling (Oxford, UK: Oxford University Press), 187–208.

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition* 75, 1–28.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cogn. Psychol.* 43, 1–22.

Boroditsky, L., and Gaby, A. (2010). Absolute spatial representations of time in an aboriginal Australian community. *Psychol. Sci.* doi:10.1177/0956797610386621

Boroditsky, L., and Prinz, J. (2008). "What thoughts are made of," in *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches*, eds G. Semin and E. Smith (New York, NY: Cambridge University Press), 98.

Boroditsky, L., and Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychol. Sci.* 13, 185–189.

Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). "Sex, syntax, and semantics," in *Language in Mind: Advances in the Study of Language and Thought*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 61–79.

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Broadbent, D. (1987). "Simple models for experimentable situations," in *Modelling Cognition*, ed. P. Morris (Chichester: John Wiley & Sons), 169–185.

Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1831–1840.

Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593.

Chemero, A. (2009). *Radical Embodied Cognitive Science.* Cambridge, MA: MIT Press.

Clark, A. (1998). *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: MIT Press.

Clark, H. (1973). "Space, time, semantics, and the child," in *Cognitive Development and the Acquisition of Language*, ed. T. E. Moore (New York, NY: Academic Press), 27–63.

Collins, A. M., and Quillian, M. R. (1969). Retrieval time from semantic memory. *J. Verbal Learn. Verbal Behav.* 8, 240–247.

Dilkina, K., McClelland, J. L., and Boroditsky, L. (2007). "How language affects thought in a connectionist model," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, eds D. S. McNamara and J. G. Trafton (Austin, TX: Cognitive Science Society), 215–220.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychol. Rev.* 62, 145–154.

Evans, V. (2004). *The Structure of Time: Language, Meaning and Temporal Cognition.* Amsterdam: John Benjamins.

Falkenhainer, B., Forbus, K., and Gentner, D. (1989). The structure-mapping engine: algorithm and examples. *Artif. Intell.* 41, 1–63.

Feldman, J. (2006). *From Molecule to Metaphor: A Neural Theory of Language.* Cambridge, MA: MIT Press.

Fodor, J. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fuhrman, O., and Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cogn. Sci.* 34, 1430–1451.

Gentner, D., Imai, M., and Boroditsky, L. (2002). As time goes by: evidence for two systems in processing space → time metaphors. *Lang. Cogn. Process.* 17, 537–565.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding.* Cambridge, UK: Cambridge University Press.

Gibbs, R. (1996). Why many concepts are metaphorical. *Cognition* 61, 309–319.

Gibbs, R. (2006). *Embodiment and Cognitive Science.* Cambridge, UK: Cambridge University Press.

Gibson, J., and Gibson, E. (1955). Perceptual learning: differentiation or enrichment? *Psychol. Rev.* 62, 32–41.

Gleitman, L., and Papafragou, A. (2005). "Language and thought," in *Cambridge Handbook of Thinking and Reasoning*, eds K. Holyoak and B. Morrison (Cambridge, UK: Cambridge University Press), 633–661.

Glenberg, A. (1997). What memory is for: creating meaning in the service of action. *Behav. Brain Sci.* 20, 41–50.

Grice, H. (1975). "Logic and conversation," in *Readings in Language and Mind*, eds H. Geirsson and M. Losonsky (Cambridge, MA: Blackwell Publishers), 41–58.

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507.

Hintzman, D. (1991). "Why are formal models useful in psychology," in *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock*, eds W. E. Hockley and S. Lewandowsky (Hillsdale, NJ: Lawrence Erlbaum), 39–56.

Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448.

Jostmann, N. B., Lakens, D., and Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychol. Sci.* 20, 1169–1174.

Kemp, C., and Tenenbaum, J. B. (2008). The discovery of structural form. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10687–10692.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By.* Chicago, IL: University of Chicago Press.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York, NY: Basic Books.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., and Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends Cogn. Sci.* 8, 108–114.

Marr, D. (1982). *Vision.* New York, NY: W.H. Freeman.

Matlock, T., Ramscar, M., and Boroditsky, L. (2005). On the experiential link between spatial and temporal language. *Cogn. Sci.* 29, 655–664.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Top. Cogn. Sci.* 1, 11–38.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–437.

McGlone, M., and Harding, J. (1998). Back (or forward?) to the future: the role of perspective in temporal language comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1211–1223.

Murphy, G. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition* 62, 99–108.

Murphy, G. L. (1996). On metaphoric representation. *Cognition* 60, 173–204.

Niedenthal, P., Barsalou, L., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Pers. Soc. Psychol. Rev.* 9, 184.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Oliveri, M., Bonnì, S., Turriziani, P., Koch, G., Lo Gerfo, E., Torriero, S., Vicario, C. M., Petrosini, L., and Caltagirone, C. (2009). Motor and linguistic linking of space and time in the cerebellum. *PLoS ONE* 4, e7933. doi: 10.1371/journal.pone.0007933.

Pinker, S. (2007). *The Stuff of Thought: Language as a Window into Human Nature*. New York, NY: Viking Press.

Proffitt, D. (2006). Embodied perception and the economy of action. *Perspect. Psychol. Sci.* 1, 110.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cogn. Sci.* 34, 909–957.

Rescorla, R. A. (1988). Pavlovian conditioning: it's not what you think it is. *Am. Psychol.* 43, 151–160.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rogers, T., and McClelland, J. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.

Rogers, T. T., Lambon Ralph, M. A., Garrad, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol. Rev.* 111, 205–235.

Rogers, T. T., and McClelland, J. L. (2008). Précis of *Semantic Cognition: A Parallel Distributed Processing Approach*. *Behav. Brain Sci.* 31, 689–749.

Searle, J. (1979). "Metaphor," in *Metaphor and Thought*, ed. A. Ortony (Cambridge, England: Cambridge University Press), 83–111.

Seidenberg, M. (1993). Connectionist models and cognitive theory. *Psychol. Sci.* 4, 228–235.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behav. Brain Sci.* 11, 1–74.

Spencer, J., Thomas, M., and McClelland, J. (2009). *Toward a Unified Theory of Development*. Oxford, UK: Oxford University Press.

Spivey, J. (2007). *The Continuity of Mind*. Oxford, UK: Oxford University Press.

Thibodeau, P. H., McClelland, J. L., and Boroditsky, L. (2009). "When a bad metaphor may not be a victimless crime: the role of metaphor in social policy," *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 809–814.

Toskos Dils, A., and Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16396–16400.

Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: MIT Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7780–7785.

Ziemke, T. (2003). "What's that thing called embodiment," in *Proceedings of the Annual Conference of the Cognitive Science Society*, eds R. Alterman and D. Kirsch (Mahwah, NJ: Cognitive Science Society), 1305–1310.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 July 2010; accepted: 22 October 2010; published online: 22 November 2010.

Citation: Flusberg SJ, Thibodeau PH, Sternberg DA and Glick JJ (2010) A connectionist approach to embodied conceptual metaphor. Front. Psychology ***1****:197. doi: 10.3389/fpsyg.2010.00197*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Flusberg, Thibodeau, Sternberg and Glick. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 31 December 2010
doi: 10.3389/fpsyg.2010.00240

Are temporal concepts embodied? A challenge for cognitive neuroscience

Alexander Kranjec* and Anjan Chatterjee

Department of Neurology, Center for Cognitive Neuroscience, University of Pennsylvania, Philadelphia, PA, USA

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Max Louwerse, University of Memphis, USA
Seana Coulson, University of California at San Diego, USA

****Correspondence:***
Alexander Kranjec, Department of Neurology, University of Pennsylvania, 3 West Gates Building, 3400 Spruce Street, Philadelphia, PA 19104, USA.
e-mail: akranjec@mail.med.upenn.edu

Is time an embodied concept? People often talk and think about temporal concepts in terms of space. This observation, along with linguistic and experimental behavioral data documenting a close conceptual relation between space and time, is often interpreted as evidence that temporal concepts are embodied. However, there is little neural data supporting the idea that our temporal concepts are grounded in sensorimotor representations. This lack of evidence may be because it is still unclear how an embodied concept of time *should* be expressed in the brain. The present paper sets out to characterize the kinds of evidence that would support or challenge embodied accounts of time. Of main interest are theoretical issues concerning (1) whether space, as a mediating concept for time, is itself best understood as embodied and (2) whether embodied theories should attempt to bypass space by investigating temporal conceptual grounding in neural systems that instantiate time perception.

Keywords: embodiment, abstract concepts, time and space, schemas

INTRODUCTION

Time is frequently talked about in terms of *space* (Clark, 1973; Hasplemath, 1997; Evans, 2004; Tenbrink, 2007). For example, languages refer to the related temporal concepts of past, present, and future in spatial terms. Languages commonly conceptualize the past as *behind*, the future as *in front of*, and the present as *here* or co-locational with the space around the body. Behavioral data suggest that such conventions in language are not arbitrary; conceptual relations between space and time seem to reflect a psychological reality that is more than "language deep." That is, time appears to be thought about as well as talked about in terms of space (Boroditsky, 2000, 2001; Gentner and Boroditsky, 2001; Boroditsky and Ramscar, 2002; Matlock et al., 2005; Kranjec, 2006; Nunez and Sweetser, 2006; Nunez et al., 2006; Torralbo et al., 2006; Santiago et al., 2007; Casasanto and Boroditsky, 2008; Kranjec et al., 2010; Kranjec and McDonough, 2011).

Along these lines, some accounts of the data make the further claim that the tight coupling observed between these two domains supports an *embodied* view of cognition. Embodied cognition theories vary in their details, but most evoke a simulation of the concept under consideration (Wilson, 2002). They typically maintain that abstract concepts are given structure in the mind by the constraints and experiential couplings imposed by human physiology, as reflected in perception and action. Thus, for example, we may talk and think about past times as being "behind us" because, when we walk, what has already been experienced in time tends also to be located behind us in space.

Generally speaking, neuroscience research that investigates relations between temporal concepts and spatial semantics is limited (Kemmerer, 2005; Teuscher et al., 2008). Despite the wealth of linguistic and behavioral evidence, there is surprisingly little neural data supporting the idea that our temporal concepts are grounded in space, or more generally embodied in sensorimotor or perceptual representations. This lack of data is partly because it is unclear what such supportive evidence should look like in the first place.

Although the details of individual models differ, *simulation* accounts of embodied cognition (Barsalou, 1999; Glenberg and Kaschak, 2002; Zwaan, 2004) suggest that concepts activate the same perceptual and sensorimotor neural networks that represent actual experience with their real-world referents. Such explanations are intuitively plausible when accounting for the conceptualization of relatively concrete categories of objects and actions. For example, according to simulation theories, comprehending the word *punch* might be expected to activate sensorimotor representations associated with arms more than legs, whereas processing a word like *kick* might be expected to do the opposite. Results generally compatible with such predictions have been found (Buccino et al., 2005; Pulvermuller et al., 2005; Kemmerer et al., 2008) although questions about whether the precise details of the findings confirm the embodied hypothesis have been raised, and knowing how to best interpret the results from such studies is often unclear (Mahon and Caramazza, 2008; Chatterjee, 2010; Arévalo et al., 2010).

A more substantial problem for embodied theories concerns the representation of abstract concepts that are less obviously accessible to perception or direct experience. What does it mean to make claims about important, or even critical underlying neuroanatomical sensorimotor structures for a concept with no obvious perceptual or action-based referent? *Conceptual metaphor theory* (Lakoff and Johnson, 1999) is often invoked to address this issue: central is the idea that we commonly talk, and importantly also think about relatively abstract domains (like time) in terms of more concrete domains (like space).

In order for embodied accounts of temporal conceptual representation to be either confirmed or disconfirmed first we must ask: What hypotheses about the neural instantiation of temporal concepts do metaphorical and simulation models generate? Although directed neural hypotheses are intuitive for the embodiment of concrete concepts – such as that of "apple" activating certain shapes, colors, and tastes – similar mappings onto sensory or

action experiences are not transparent for more abstract concepts. So before clear-cut neural hypotheses can be formed, some foundational issues will need to be clarified. The present paper attempts to address issues related to the embodiment of temporal concepts.

We begin by reviewing the kinds of temporal concepts that tend to get mapped onto spatial relations. However, it is not clear to us that space, as relevant to the kinds of relations onto which temporal concepts are mapped, is itself embodied. A subsequent section therefore addresses the question of whether representations coding spatial concepts are best understood as embodied in the first place. Finally, we address the question of whether temporal concepts could be embodied more directly. Rather than making the claim that temporal concepts are embodied because they are mapped onto spatial concepts, one could ask if temporal concepts are embodied because they map onto sensory and motor representations of time itself. Perhaps embodied temporal concepts can bypass space. We briefly review temporal perception from a cognitive neuroscience perspective as relevant to embodied theories.

SPACE, TIME, LANGUAGE, AND THOUGHT

Conceptual metaphor theory (Lakoff and Johnson, 1999) provides a powerful framework for investigating the embodied representation of abstract concepts. At the core of conceptual metaphor theory is the idea that we commonly talk and think about relatively abstract domains (like time) in terms of more concrete domains (like space). Indeed, it seems as though *spatial relations* do provide structure for many abstract concepts. So, for example, we tend to conceptualize emotional states along a vertical axis (*happy* as *up*) and similarity in terms of proximity (*difference* as *far*; Meier and Robinson, 2004; Casasanto, 2008; Boot and Pecher, 2009). In this manner, "mental metaphors" (Casasanto, 2009) presumably help us to organize abstract concepts by mentally mapping a concept that we cannot easily perceive onto a concept more directly associated with perceptual or motor representations. These cross-domain mappings are not only argued to be *conceptual* in nature – i.e., more than mere linguistic artifacts of interest to etymologists – but they are argued to be *asymmetrical* – i.e., concrete concepts are thought to structure more abstract ones, but not *vice versa*. Experimental data collected using non-linguistic tasks bolster these claims. For example, Casasanto and Boroditsky (2008) found that the remembered size of a line in space concordantly modulates recall for its duration, but not *vice versa*. That is, (spatially) longer lines are remembered as being presented for longer times, but lines of greater durations are not remembered as having greater spatial extent. This, and other related findings (Boot and Pecher, 2009), are used to support the claim that patterns observed in language (i.e., the systematicity and asymmetry of space-time mappings) reflect deeper relations that influence other kinds of thinking. The spatial conceptualization of time has, in particular, been studied by both experimentalists and cognitive linguists in such great detail that time is postulated to be "the model system of choice for linguistic and psychological tests of relationships between metaphorical source and target domains" (Casasanto, 2009).

The model status of space–time relations owes something to the fact that conceptual divisions within the domain of space map intuitively onto complementary temporal concepts. For example, we talk about temporal extent or duration in terms of paths or distances in space (e.g., *The Friday meeting marks the end of a long week*); temporal order judgments, or sequences, in terms of dynamic object relations in space (e.g., *The meeting comes before lunch*); and the past and future in egocentric spatial terms (e.g., *The meeting is behind us*). Such a division of the time concept (duration; sequential order; past/future) seems to map onto traditional tripartite reference frame models that distinguish between extrinsic, intrinsic, and deictic frames of reference and also lend themselves to distinct spatial schemas (Kranjec, 2006; Zinken, 2009; Kranjec and McDonough, 2011; see **Figures 1A–C**).

Why are these relations between space and time so intuitive? Spatial and temporal relations share many similarities by being linked in experience. It does generally take more time to visually scan larger objects; objects that arrive earlier than other objects, in time, are generally in front of later objects, in space; and, as ambulatory, front-facing organisms, events in our past tend to involve locations behind us. To some extent, this accounts for why temporal meanings can be easily expressed using spatial schemas. As we discuss in detail later, by spatial schemas we mean pared down, analog representations of spatial relations that can be depicted by points, lines, and vectors

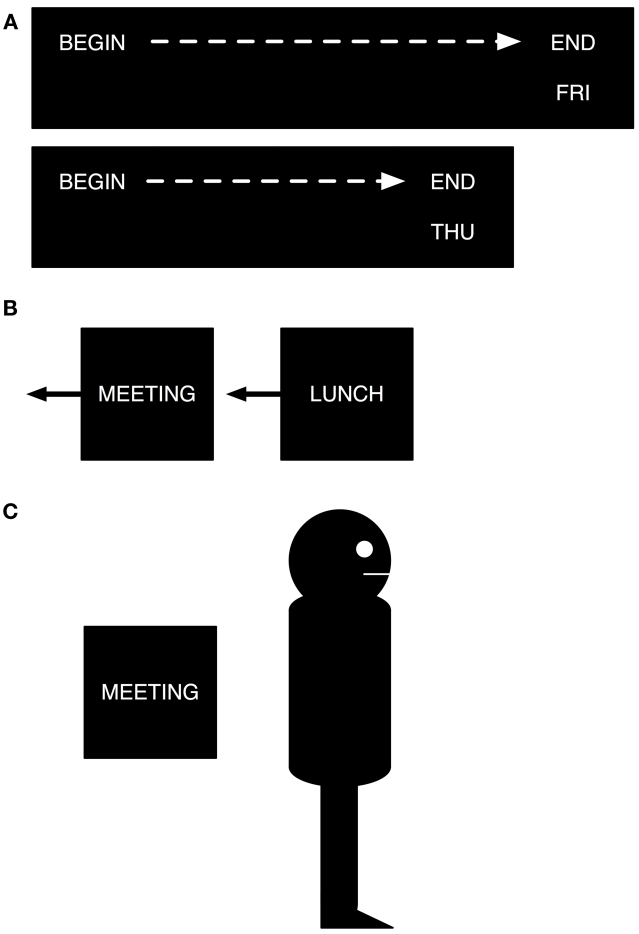


FIGURE 1 | Schematic depictions of three temporal concepts. **(A)** Duration; *A long work week ends Friday, a short one ends Thursday,* **(B)** Sequence; *The meeting comes before lunch,* **(C)** Past/Future; *The meeting is behind us.*

(Mandler, 1996; Talmy, 2000). The analog spatial characteristics of schemas would appear to map directly onto temporal phenomena. However, do these experiential–semantic correlations necessarily tell us that temporal concepts are embodied at the neural level?

ABSTRACTING SPACE

The idea that time is embodied because it is mapped onto space raises two fundamental questions; first, (1) how are spatial relations that are potentially important for structuring other concepts themselves represented in the brain and, second, (2) whether spatial relations represented as such are best understood as embodied. These two issues concern the representational continuity of spatial perception, thought, and language (Chatterjee, 2001, 2008). Roots of these ideas can be traced back to Pavio's (1986) dual coding hypothesis that suggested that information can be coded in analog as well as symbolic formats. Here, we are concerned with their interactions with respect to time and space and the ramifications for defining the limits of what can be reasonably considered an embodied abstract concept in the first place (Chatterjee, 2010). As things currently stand, the extent to which spatial relational information is in fact grounded in perceptual or sensorimotor content – i.e., spatial representation at the level described by embodied theories of cognition – is far from clear.

In conceptual metaphor theory, the idea of spatial relational *schemas* (Johnson, 1987; Mandler, 1992; Lakoff and Johnson, 1999; Talmy, 2000) plays a critical role in embodying abstract concepts. Schemas are generally described as "boiled down" abstractions of frequently observed spatial and motion patterns. Although they are often portrayed as static, iconic figures, or diagrams, *mental schemas* are better understood as multi or supramodal abstractions of frequently occurring patterns of movement, spatial relations, and force dynamics. Conceptual metaphor theory makes it clear that schemas are not the same as percepts ("We do not see spatial relations the way we see physical objects"; Lakoff and Johnson, 1999). Although their semantics often map onto the meanings of prepositions, they are thought to develop prior to word-like representations (Mandler, 1992). Schemas are also distinct from the kind of propositional or predicate logic structures that can operate on spatial relations in a mental model approach (Miller and Johnson-Laird, 1976) or the conceptual structures proposed by Jackendoff (1990). In many ways, schemas often seem to be defined in terms of what they are not, rather than what they are. Very broadly though, and the sense in which they are used in this paper, schemas can be understood as representations that code for abstract spatiotemporal relations among objects – like paths, containment, contact, and support relations – that provide a conceptual base onto which language can be mapped. They are not visual despite having analog properties; schemas represent the relations among objects, not the objects. They represent a pared down product of percepts, but not the percepts themselves (Mandler, 2000). Schemas therefore are generally treated as intermediate formats in multiple system approaches to mental representation (see **Figure 2**).

It is our view that the same basic spatial relation can be encoded in several formats, instantiated by distinct brain regions. Such a view may be compatible with other multiple system approaches. For example, Barsalou and colleagues propose that both words and perceptually grounded simulations play a role in representing concepts (Simmons et al., 2008); that supramodal representations of space and time serve an overarching role in structuring "perceptual symbols" (Barsalou, 1999); and that "relation simulators" serve as the mechanism for extracting abstract spatial and temporal relations from more perceptually rich imagistic representations (Barsalou, 2003). Kemmerer and Tranel (2000) found evidence for a double dissociation between linguistic and perceptual representations of spatial relations. A patient with left fronto-parietal damage did poorly on verbal tasks relying on categorical representations with relatively intact performance on coordinate visuospatial tasks, while a patient with damage to right frontal, parietal, and temporal areas displayed the opposite pattern. This suggests that abstract spatial relations may be stored separately in verbal and non-verbal formats.

Spatial schemas are, however, presently a theoretical construct. Although much has been written about schemas from a philosophical, developmental, and cognitive linguistic perspective, very little about their neural organization is understood. For present purposes, we can only review in more general terms how such *abstract but discrete spatial relations*, might be represented in the brain based on available neural data. In conducting research in this broader area, cognitive neuroscience has traditionally focused on (1) the left-hemisphere representation of spatial prepositions and categorical relations and (2) deficits in spatial representation associated with damage to the right-hemisphere (i.e., neglect).

For the current discussion, a short review of spatial prepositions and categorical relations in the context of cognitive neuroscience is a reasonable place to start. The spatial representation of time is

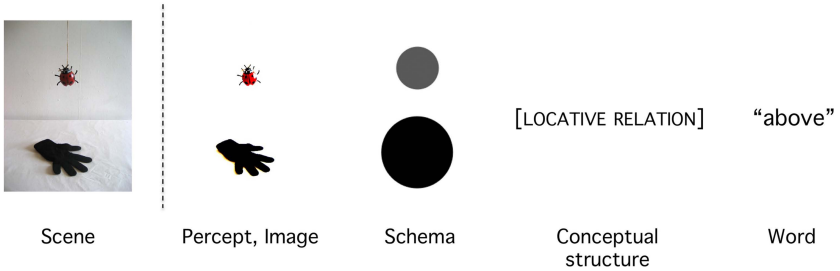


FIGURE 2 | The mental representation of spatial relations. The perceptual or imagistic representation captures the sensory-rich features of the actual scene. The schema abstracts the relative positions of the objects while retaining some analog structure. Conceptual structure is language-like, algebraic, and propositional. A verbal representation encodes the label referring to a discrete, categorical spatial relation (Adapted from Chatterjee, 2001).

reliant on prepositions for expressing different kinds of temporal concepts. When communicating information about durations (e.g., We examined policy *across* the decade), sequences (The meeting was *before* lunch), and the past or future (The illness is far *behind* me) prepositions are used to invoke different kinds of spatial relations. Time concepts map onto spatial prepositions (Kranjec et al., 2010) and spatial prepositions, practically speaking, operate as relational schemas. For example, the preposition *on* can be used to represent the semantic relations between a pen on a table, and a boat on a lake. Similarly, when prepositions like *on* are used in the temporal domain as with, *The meeting is on Wednesday*, what is being invoked is a relational schema, emphasizing collocation, and contact.

Frederici (1981) demonstrated that aphasics with left posterior temporal–parietal lesions show semantic impairments that dissociate from syntactic ones in processing locative prepositions. Subsequently, Landau and Jackendoff (1993) proposed that the parietal cortex might process prepositions because it serves as the terminus of the dorsal "where" pathway. Damasio and colleagues (Damasio et al., 2001; Emmorey et al., 2002) corroborated this idea by finding a role for the left supramarginal gyrus and inferior frontal gyrus in the comprehension of locative prepositions. Noordzij et al. (2008) found that understanding the meanings of locative prepositions embedded in sentences, and the spatial relations expressed in picture formats were both associated with activation in the left supramarginal gyrus. And Wu et al. (2007) found that damage to the left inferior frontal–parietal cortices impaired knowledge of locative relations lexicalized by prepositions in simple sentences. In studies investigating categorical spatial relations using non-verbal tasks, additional work from our own laboratory and others also implicate left-hemisphere areas in the representation of the type of categorical spatial relations encoded by locative prepositions. Specifically, these studies find the inferior parietal lobe and frontal operculum to be involved (Kosslyn, 1987; Kosslyn et al., 1989, 1998; Amorapanth et al., 2010). The general view that emerges from both literatures – on spatial prepositions and on categorical spatial relations – suggests that the left-hemisphere, more than the right, processes these kinds of relations across verbal and non-verbal tasks. Furthermore, the inferior parietal cortex and possibly parts of dorsolateral prefrontal cortex involving inferior and posterior middle frontal gyri are critical in mediating this kind of spatial relational knowledge.

However, despite this left-hemisphere bias for processing prepositions and categorical spatial relations, spatial representation and spatial attention are more generally regarded as right-lateralized functions of the inferior parietal cortex, and the supramarginal gyrus in particular (McFie et al., 1950; Vallar and Perani, 1986). In contrast to left-hemisphere word-like representations, right-hemisphere spatial representations are often described as being analog in format (Bisiach, 1993). Furthermore, representations of space in the right parietal lobe are thought to be supramodal, as there is evidence spatial deficits associated with neglect affect sensory modalities other than vision (Farah et al., 1989) and that right-hemisphere spatial representations are dissociable from visual imagery (Farah et al., 1988). Thus, spatial representation in the right-hemisphere can also be conceived as separable from any particular sensory modality. The picture that emerges from the cognitive neuroscience literature is one in which word-like spatial relations are represented in the left supramarginal gyrus and analog spatial relations are represented in the right supramarginal gyrus; where both kinds of representations may be more abstract than any particular sensory modality (see **Figure 3**).

The potential for abstract spatial relations to be instantiated in distinct formats of (1) amodal, left-hemisphere representations closely linked to language and/or (2) supramodal, right-hemisphere spatial representations linked more closely to perception, makes investigating image schemas with respect to sensorimotor grounding complicated. If representations of spatial relations provide organizational structure across sensorimotor modalities, how can one operationally distinguish between representations best characterized as either disembodied-and-amodal (in the left-hemisphere) or embodied-and-supramodal (in the right)? In fact, some philosophers think that making such an operational distinction between amodal and supramodal representations may be impossible (Dove, 2009; Machery, 2009). Barsalou (1999) states that supramodal spatial representations "constitute fundamental parts of perception that integrate…specific modalities (p. 638)." But if language itself is modeled as "part of perception" and also plays a role in "integrating modalities," the model has the risk of becoming circular; especially if spatial language is critical for providing structure across a range of abstract concepts. For example, some neo-Whorfian accounts suggest that verbal category labels exert an influence over perception and conceptualization (Lupyan, 2008; Regier and Kay, 2009; Lupyan et al., 2010). And it has been observed that patterns of spatial language exert an influence in both directions as well: over both time perception and abstract conceptualization (Boroditsky and Ramscar, 2002; Casasanto and Boroditsky, 2008). Although such work establishes that links exist between spatial language, perception, and the conceptualization of other domains – links that can persist even when non-linguistic tasks are employed – this work tells us less about the amodal, modal, or supramodal nature of the participating representations underlying a given behavioral effect.

While the research outlined above suggests that perceptual and conceptual systems are more integrated than previously thought, neural studies may permit researchers investigating space-time relations to make finer-grained comparisons of representations that serve as the basis for this integration. That is, by delineating the properties of potentially related, but distinct, mental representations in a multiple systems model, we can begin to ask if more word-like spatial labels or more percept-like image schemas ground temporal concepts. One step may involve determining if the neural instantiation of image schema-like spatial representations have more in common with the kinds of spatial relations thought to be instantiated in the left- or right-hemispheres. The more general challenge for cognitive neuroscience is to develop specific tasks and methods, such that, if there are discrete anatomical areas dedicated to processing the verbal, perceptual, or schematic representations of spatial relations, they can be identified. Once identified, the question would be: Do temporal concepts also engage the same neural structures?

Although empirical evidence for the existence of schematic representation is lacking, as a thought experiment, let us imagine that there was good functional and anatomical evidence available for both (1) the neural instantiation of schematic representations *and* specifically (2) the mapping of temporal relations onto structurally

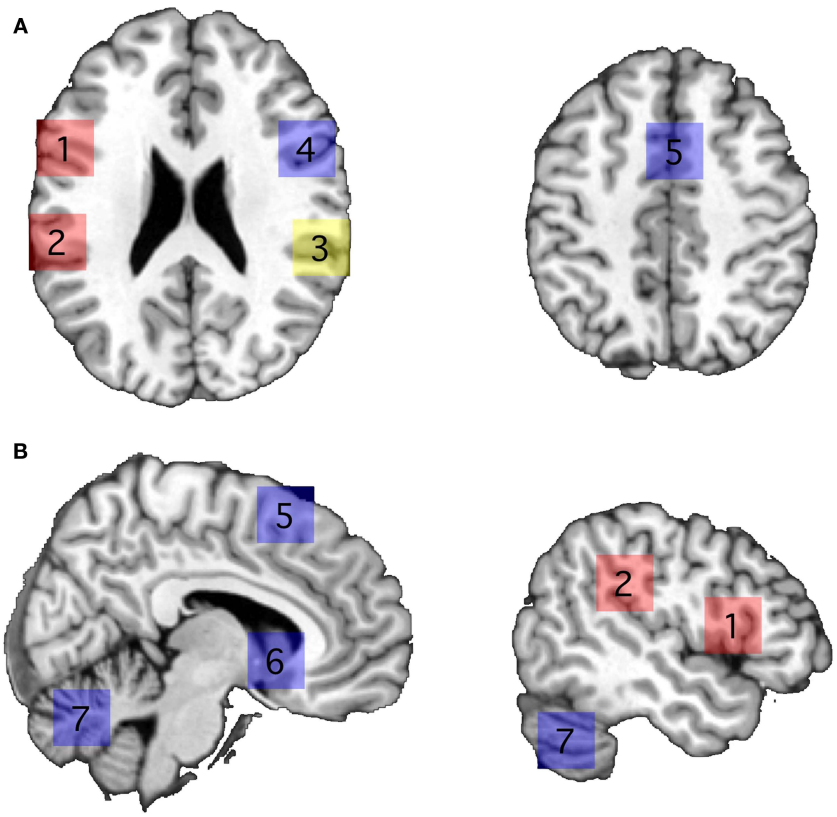


FIGURE 3 | (A) Transverse and **(B)** sagittal slices selected to illustrate anatomical areas discussed in Sections "Abstracting Space" and "Simulating Time" hypothesized to be important for locative and categorical spatial information (in red), supramodal spatial representation and attention (in yellow), and timing (in blue): Colored squares highlight areas including: (1) left inferior frontal gyrus, (2) left supramarginal gyrus, (3) right supramarginal gyrus, (4) right inferior frontal gyrus, (5) supplementary motor area, (6) basal ganglia (location of caudate shown here), and (7) cerebellum.

concordant spatial schemas. The question would still remain: are schemas themselves best understood as embodied? In conceptual metaphor theory, schemas play a critical role in structuring more abstract concepts (like emotions, similarity, and time). However, since the idea of a schema remains a theoretical construct, it is particularly important to consider what finding schemas instantiated in the brain would actually tell us about embodiment.

Not much about schematic representation is agreed upon, but schemas, by most definitions, involve a process of moving away from the concrete perceptual attributes of objects in order to represent the more abstract relations between them. Thus, there are good reasons to hypothesize that representations resulting from such a process would in fact *not* involve perceptual or sensorimotor neural networks, or if they did, they would be greatly attenuated. In fact, the main proponents of embodied theories of mind acknowledge this limit. Lakoff (1987) writes that "image schemas are kinesthetic in nature, that is, they have to do with the sense of spatial locations, movement, shape, etc., independent of any particular sensory modality." (See Hampe, 2005 for more recent discussions, especially chapters by Dean and Grady). And Barsalou (2003) describes how abstract representations of spatial relations filter out the irrelevant details of rich simulations. Thus, even according to conceptual metaphor and simulation theories, when devising neural studies for tasks designed to invoke schemas, structures *outside* (but perhaps adjacent to) those sensorimotor areas thought to represent the most concrete aspects of direct sensory experience might be doing the most important work in representing abstract spatial relations. Using similar logic, stroke patients with lesions directly in perceptual or sensorimotor areas might be expected to retain the ability to represent the relations contained within perceptual arrays because those areas involved in fully representing the rich, perceptual features of visual displays may not be necessary for extracting the abstract relations between objects (in adult brains at least).

Considered from such a perspective, the role of schemas in a continuous, graded model of conceptual representation could actually be useful for marking where in this system embodiment *trails off* and abstraction begins. Before we can conclude one way or the other, many more details pertaining to how (and if) schematic relations are represented in the brain need to be known. Although empirical data relating to schematic representation is in short supply, we entertain the theoretical position that abstract, schema-like representations mediate between perception and language but that these representations are not embodied in the strict sense; i.e., there are no convincing reasons to think that schemas necessarily involve simulating previous experience with instances of a concept. According to our view, if schemas are instantiated in the brain they will be found to code the same kind of abstract structural roles

among objects in space (and time) that are evoked by prepositions, but in a format and with a neural implementation that is distinct from both perceptual and verbal representations.

Conceived as such, schemas are disembodied at least as much as they are embodied. According to such a view one can, in principle, hold that schemas may be (a) derived from perceptual and motor experiences, (b) have some analog, and (c) some computational properties of abstract representations (as Barsalou, 1999 claims) and still maintain that such features are as characteristic of a "disembodied" representation as much as they are an "embodied" one. The notion that an abstract, analog representation need not be embodied in the strict simulation sense is the fulcrum of this position. It is an important point to consider because when one makes strong claims about the embodiment of image schemas one necessarily underplays the role of higher-order processes like relational thought, abstraction, and analogy that may play an especially important role during early cognitive development (Mandler, 2000; Gentner, 2003) and continue to mediate relations between spatial concepts and the more abstract concepts that find structure in space. Although not opposed to embodied cognition accounts, the view put forth here is compatible with a graded account of mental representation that resists making strong claims about the necessity of sensorimotor simulation in grounding language and thought (Chatterjee, 2010). It is also consistent with Talmy's (2000) view and Mandler's (2004) developmental framework which suggest that image schemas provide the foundation for *explicitly* accessible concepts. It is our position that conceptual meanings must be grounded on representations with content accessible to conscious analysis, thus it is unclear how sensorimotor representations impenetrable to conscious analysis could serve this purpose[1]. Schemas conceived in such a manner resemble the kind of *meaningful but not perceptual* right-hemisphere spatial representations described by early researchers of neglect:

> Spatial analogs may be claimed to be unfit to convey the full meaning a representation is supposed to be endowed with. In the absence of further comparative analysis of what is meant by "meaning" and "representation", my reply is that, on the one hand, meaning may be inherent in the kinetic features of spatial analogs (Bisiach and Berti, 1990), and, on the other, analogue representation may be conceived as being "enthymematic"– that is, leaving unexpressed a great deal of the antecedents and entailments of which meaning consists (Bisiach, 1992). Unlike pictures, as Sterelny (1990) would say, analogue representations are "preinterpreted" (Bisiach, 1993).

This view is consistent with the developmental mechanism that Mandler (2000) terms *perceptual analysis* and defines as "a process in which perceptual input is attentively analyzed and recoded into a new format." For schemas to be independently meaningful, they must come "preinterpreted."

SIMULATING TIME

But does an embodied temporal concept necessarily need a spatial schema for grounding? Consider the fact that we do not only talk about time in terms of space. We frequently talk about time in its own terms; both grammatically and semantically. The focus on spatial metaphoric accounts of abstract conceptualization by embodied theories has had the ironic consequence of creating a blind spot for temporal language. By most accounts, the use of "purely" temporal, time-specific language precedes the use of metaphoric, spatial–temporal language in development (Nelson, 1996). Simulation approaches might closely examine how the brain codes obligatory grammatical categories like (viewer-centered) tense and (event-related) aspect, examining the ways in which such grammaticalization patterns could relate back to non-linguistic processes associated with the representation of distinct temporal reference frames. Similarly, the lexicalization of pure temporal concepts associated with particular temporal perspectives (e.g., egocentric: *now, yesterday*; duration: *while, yet*; sequence: *later*; frequency: *always, never*) also occurs sooner in development compared to spatial metaphoric language for time. Although little is known about the development of embodied concepts, the early acquisition of time-pure lexical concepts and grammatical rule use implies, at least, that spatial grounding is not necessary for all temporal conceptualization. If space is not necessary for grounding temporal concepts, it suggests that support for simulation accounts might be found by looking for semantic grounding in sensorimotor networks associated with temporal, not spatial, processing.

The major claims made by simulation accounts of embodied cognition center around the basic idea that representing a particular concept at least partially entails the same sensorimotor neural networks activated during direct experience with the thing to which the concept refers. Recall that a general approach like this makes straightforward predictions. For example, representing concepts associated with actions (e.g., *kick*) should involve networks in primary motor cortex or visual motion areas, while representing concepts associated with sounds (e.g., *thunder*) should involve networks in primary auditory cortex. Predictions can be made with finer granularity within a domain such that an embodied hypothesis might predict that comprehending the meaning of *punch* would activate homuncular motor representations associated with arms more than legs, and *kick*, legs more than arms. If simulation accounts view concrete concepts as co-activating distributed networks of sensorimotor and domain-specific linguistic representations, how would one expect such a perspective to be applied to time? Although research on conceptual metaphor suggests that we look for the neural grounding of temporal concepts in representations of spatial schemas, as discussed above, neural structures supporting spatial schemas might not be sufficient to support simulation theories. Fortunately, the domain of time may permit a relatively direct approach for testing hypotheses about the embodiment of temporal concepts.

The brain has distinct neural mechanisms for representing different timescales and types of temporal information. Roughly, sub-second intervals are processed by subcortical areas, supra-second intervals by more diffuse cortical circuits, and the clock that underlies large-scale circadian rhythms is located in the suprachiasmatic nucleus of the hypothalamus (Buhusi and Meck, 2005). Recent meta-analyses further corroborate this view indicating that sub-cortical networks including the cerebellum and basal ganglia are strongly associated with processing the motor and perceptual components of sub-second timing tasks, while cortical areas like the supplementary motor area (SMA) and prefrontal cortex are more

[1]See Kranjec and McDonough (2011) for a thorough theoretical account for why image schema *content* should be expected to be explicit and therefore represented outside sensorimotor areas. We also report experimental evidence in support of this idea.

involved in supra-second timing (Wiener et al., 2010). However, a related lesion study suggests that the basal ganglia are not critical for timing *per se*, but rather with the production of timed movements (Coslett et al., 2010). Other recent work from our laboratory using fMRI and a duration discrimination task, also implicates the SMA in supra-second timing, in addition to left inferior frontal, and superior temporal cortical structures (Wencil et al., 2010). However, the perception of time is not simply a matter of representing duration. The neural instantiation of duration information dissociates with that for sequence information. There is evidence that making ordinal sequence judgments involves distinct premotor cortical areas as compared to making duration judgments (Schubotz and von Cramon, 2001) and that learning the sequence of a motor response involves right parietal structures while learning the duration of the same response, the cerebellum (Sakai et al., 2002). And although less accurately described as perceptual in nature, the neural architecture underlying thinking about the future and the past has also been investigated (Hassabis et al., 2007; Abraham et al., 2008; Arzy et al., 2009a,b). Other work finds the parietal cortex to be important for representing time in a more abstract sense of magnitude (Bueti and Walsh, 2009). Unlike other very abstract concepts (like *hope* for example), something is already known about the neural bases for several distinct cognitive processes associated with temporal cognition.

Considering what we know about timing in the brain, it seems that simulation approaches would benefit from using tasks and methods designed to probe *within* the domain of time to look for specific relations between neural areas specialized for a particular kind of time-specific cognitive processing (e.g., duration, sequence, or past/future representation) and an associated linguistic representation (e.g., distinctions between lexicalized concepts or grammatical class). If a simulation account of temporal conceptualization does not predict the grounding of time concepts in neural areas dedicated to processing temporal information it should explicitly state why this should be the case.

CONCLUSIONS

Time is the most frequently used noun in the English language (Soanes and Stevenson, 2007). But where do we look for temporal concepts in the brain? Current cognitive science approaches tend to investigate the spatial organization of temporal concepts and draw conclusions about embodied cognition. Although linguistic and experimental behavioral data are abundant, little neural data is available. This is not an ideal state for the field because providing direct evidence for the embodiment of abstract concepts is potentially possible within cognitive neuroscience (e.g., What are the neural bases for spatial schemas? For time perception? To what extent do time concepts map onto such structures?).

A study by Kemmerer (2005) is notable. In one of the few neuroscientific studies investigating the relations between space, time, and semantics, Kemmerer found evidence for a double dissociation for impaired comprehension of either the spatial or temporal meanings of prepositions in patients with focal brain lesions. The left supramarginal gyrus was determined to be a critical structure for representing the spatial meaning of prepositions, whereas areas important for representing temporal meanings were less clearly defined. However, the left perisylvian cortex was implicated. The results at least suggest that the spatial and temporal meanings of prepositions are represented separately, but are equivocal with respect to the current discussion. However, they suggest that the language of time is stored separately from the language of space. In general, embodied accounts would benefit by probing further for domain-specific dissociations between representations of temporal and spatial concepts in areas associated with spatial and temporal perception using cognitive neuroscience methods. Lesion studies, especially larger group studies, are particularly well suited for testing hypotheses about embodied cognition. They allow for making strong inferences regarding the necessity of a particular anatomical structure (e.g., perceptual or sensorimotor structures) for performing a given cognitive function. Well-designed lesion studies could provide especially valuable information concerning both (1) the roles that perceptual or schematic formats play in representing spatial relational information and (2) the part that such spatial representations vs. separate structures dedicated to temporal perception play in grounding our concepts of time (see **Figure 3**). Clearly, tasks that compare neural representations across verbal and perceptual tasks are required.

The ways in which we have described the conceptual grounding of time in either spatial abstractions or more directly in sensorimotor processes are not necessarily mutually exclusive. By probing domain-specific sensorimotor areas associated with temporal perception and perspective, cognitive neuroscientists place themselves in a position to potentially confirm aspects of simulation accounts while bypassing conceptual metaphor theories reliant on spatial schematic grounding. Alternatively, if temporal concepts find structure primarily in spatial abstractions, time could become the model domain for demonstrating how a particular abstract concept can become disembodied, or disconnected from the neural basis of its concrete representation rather than the other way around. The notion that conceptual processing involves both verbal and nonverbal representations is not new (Pavio, 1986) and has been refined in the context of more recent debates concerning embodiment and simulation theories (Louwerse, 2010; Riordan and Jones, 2010). In the case of space, distinguishing between formats presents a particular challenge, as it may be that the same basic spatial meanings encoded by amodal left-hemisphere verbal representations are also inherent to supramodal right-hemisphere representations.

Answers to questions related to these issues will depend partly on the neural evidence as it becomes available, but also on how theoretical discussions concerning the nature of spatial abstraction and conceptual scaffolding unfold. The present paper has tried to frame core issues concerning the embodied structure of abstract concepts with respect to some of what is currently known about the functional neuroanatomy of spatial and temporal representations. The depiction of embodied theories presented here may strike some as superficial. However, we are hopeful that the way common issues have been broadly framed will make addressing them more amenable to direct hypothesis testing using empirical methods. If timing areas play no role in grounding time concepts, those taking a strong embodied approach should be able to provide a rigorous explanation for why this is the case. If spatial representations ground time concepts, a more fine-grained neuroanatomical account capable of distinguishing between word-like representations in the left-hemisphere and more perceptual ones in the right-hemisphere may be possible. And finally, if the kinds of spatial schemas that are thought to ground temporal concepts cannot be easily identified,

described, or contrasted with other kinds of spatial representations (i.e., symbolic or perceptual), then the validity of this proposed intermediate representational construct may need to be reconsidered by those taking an empirical approach.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health [RO1 DC004817, RO1 DC008779] and the National Science Foundation [subcontract under SBE0541957].

REFERENCES

Abraham, A., Schubotz, R. I., and von Cramon, D. Y. (2008). Thinking about the future versus the past in personal and non-personal contexts. *Brain Res.* 1233, 106–119.

Amorapanth, P., Widick, P., and Chatterjee, A. (2010). The neural basis for spatial relations. *J. Cogn. Neurosci.* 8, 1739–1753.

Arévalo, A., Baldo, J., and Dronkers, N. (2010). What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system? *Cortex.* doi: 10.1016/j.cortex.2010.06.001. [Epub ahead of print].

Arzy, S., Bick, A., and Blanke, O. (2009a). Mental time in amnesia: evidence from bilateral medial temporal damage before and after recovery. *Cogn. Neuropsychol.* 26, 503–510.

Arzy, S., Collette, S., Ionta, S., Fornari, E., and Blanke, O. (2009b). Subjective mental time: the functional architecture of projecting the self to past and future. *Eur. J. Neurosci.* 30, 2009–2017.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Phil. Trans. Roy. Soc. Lond. B Biol. Sci.* 358, 1177–1187.

Bisiach, E. (1992). "Understanding consciousness," in *The Neuropsychology of Consciousness*, eds A. D. Milner and M. Rugg (London: Academic Press), 113–137.

Bisiach, E. (1993). Mental representation in unilateral neglect and related disorders: the twentieth Bartlett Memorial lecture. *Q. J. Exp. Psychol.* 46A, 435–461.

Bisiach, E., and Berti, A. (1990). "Waking images and neural activity," in *The Psychophysiology of Mental Imagery*, eds R. G. Kunzendorf and A. A. Sheikh (Amityville, NY: Baywood), 67–88.

Boot, I., and Pecher, D. (2009). Similarity is closeness: metaphorical mapping in a conceptual task. *Q. J. Exp. Psychol. (Colchester)* 63, 942–954.

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition* 75, 1–28.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cogn. Psychol.* 43, 1–22.

Boroditsky, L., and Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychol. Sci.* 13, 185–189.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., and Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Cogn. Brain Res.* 24, 355–363.

Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Phil. Trans. Roy. Soc. B Biol. Sci.* 364, 1831–1840.

Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* 6, 755–765.

Casasanto, D. (2008). Similarity and proximity: when does close in space mean close in mind? *Mem. Cognit.* 36, 1047–1056.

Casasanto, D. (2009). "When is a linguistic metaphor a conceptual metaphor?" in *New Directions in Cognitive Linguistics*, eds V. Evans and S. Pourcel (Amsterdam: John Benjamins), 127–145.

Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593.

Chatterjee, A. (2001). Language and space: some interactions. *Trends Cogn. Sci.* 5, 55–61.

Chatterjee, A. (2008). The neural organization of spatial thought and language. *Semin. Speech Lang.* 29, 226–238.

Chatterjee, A. (2010). Disembodying cognition. *Lang. Cogn.* 2, 79–116.

Clark, H. (1973). "Space, time, semantics and the child," in *Cognitive Development and the Acquisition of Language*, ed. T. Moore (New York: Academic Press).

Coslett, H. B., Wiener, M., and Chatterjee, A. (2010). Dissociable neural systems for timing: evidence from subjects with basal ganglia lesions. *PLoS ONE* 5, e10324. doi: 10.1371/journal.pone.0010324

Damasio, H., Grabowski, T. J., Tranel, D., Ponto, L. L., Hichwa, R. D., and Damasio, A. R. (2001). Neural correlates of naming actions and of naming spatial relations. *Neuroimage* 13(Pt 1), 1053–1064.

Dove, G. (2009). Beyond perceptual symbols: a call for representational pluralism. *Cognition* 110, 412–431.

Emmorey, K., Damasio, H., McCullough, S., Grabowski, T., Ponto, L. L., Hichwa, R. D., and Bellugi, U. (2002). Neural systems underlying spatial language in American sign language. *Neuroimage* 17, 812–824.

Evans, V. (2004). *The Structure of Time: Language, Meaning and Temporal Cognition*. Amsterdam: John Benjamins.

Farah, M. J., Hammond, K. M., Levine, D. N., and Calvanio, R. (1988). Visual and spatial mental-imagery – dissociable systems of representation. *Cogn. Psychol.* 20, 439–462.

Farah, M. J., Wong, A. B., Monheit, M. A., and Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention – modality-specific or supramodal. *Neuropsychologia* 27, 461–470.

Frederici, A. (1981). Production and comprehension of prepositions in aphasia. *Neuropsychologia* 19, 191–199.

Gentner, D. (2003). "Why we're so smart," in *Language in Mind*, eds D. Gentner and S. Goldin-Meadows (Cambridge, MA: The MIT Press), 195–235.

Gentner, D., and Boroditsky, L. (2001). "Individuation, relativity and early word learning," in *Language Acquisition and Conceptual Development*, eds M. Bowerman and S. C. Levinson (England: Cambridge University Press), 215–256.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Hampe, B. (2005). *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, Vol. 29. Berlin: Mouton de Gruyter.

Hasplemath, M. (1997). *From Space to Time: Temporal Adverbials in the World's Languages*. Munich: Lincom Europa.

Hassabis, D., Kumaran, D., Vann, S. D., and Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1726–1731.

Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA: The MIT Press.

Johnson, M. (1987). *The Body in Mind*. Chicago: Chicago University Press.

Kemmerer, D. (2005). The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologia* 44, 1607–1621.

Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., and Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI. *Brain Lang.* 107, 16–43.

Kemmerer, D., and Tranel, D. (2000). A double dissociation between linguistic and perceptual representations of spatial relationships. *Cogn. Neuropsychol.* 17, 393–414.

Kosslyn, S. (1987). Seeing and imagining in the cerebral hemispheres: a computational analysis. *Psychol. Rev.* 94, 148–175.

Kosslyn, S., Koenig, O., Barrett, C., Cave, C., Tang, J., and Gabrielli, J. (1989). Evidence for two types of spatial representations: hemispheric specialization for categorical and coordinate relations. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 723–735.

Kosslyn, S., Thompson, W., Gitelman, D., and Alpert, N. (1998). Neural systems that encode categorical versus coordinate spatial relations: PET investigations. *Psychobiology* 26, 333–347.

Kranjec, A. (2006). Extending spatial frames of reference to temporal concepts. *Paper Presented at the Twenty-Eighth Annual Conference of the Cognitive Science Society*, Vancouver.

Kranjec, A., Cardillo, E. R., Schmidt, G. L., and Chatterjee, A. (2010). Prescribed spatial prepositions influence how we think about time. *Cognition* 114, 111–116.

Kranjec, A., and McDonough, L. (2011). The implicit and explicit embodiment of time. *J. Pragmat.* 43, 735–748. (Special Issue: The Language of Space and Time).

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh*. New York, NY: Basic Books.

Landau, B., and Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behav. Brain Sci.* 16, 217–265.

Louwerse, M. (2010). Symbol interdependency in symbolic and embodied cognition. *Top. Cogn. Sci.* doi: 10.1111/j.1756-8765.2010.01106.x. [Epub ahead of print].

Lupyan, G. (2008). The conceptual grouping effect: categories matter (and named categories matter more). *Cognition* 108, 566–577.

Lupyan, G., Thompson-Schill, S. L., and Swingley, D. (2010). Conceptual penetration of visual processing. *Psychol. Sci.* 21, 682–691.

Machery, E. (2009). *Doing without Concepts*. New York: Oxford University Press.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychol. Rev.* 99, 587–604.

Mandler, J. M. (1996). "Preverbal representation and language," in *Language and Space*, eds P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett (Cambridge, MA: The MIT Press), 365–384.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *J. Cogn. Dev.* 1, 3–36.

Mandler, J. M. (2004). *The Foundations of Mind: Origins of Conceptual Thought*. New York, NY: Oxford University Press.

Matlock, T., Ramscar, M., and Boroditsky, L. (2005). On the experiential link between spatial and temporal language. *Cogn. Sci.* 29, 655–664.

McFie, J., Piercy, M. F., and Zangwill, O. L. (1950). Visual-spatial agnosia associated with lesions of the right cerebral hemisphere. *Brain* 73, 167–190.

Meier, B., and Robinson, M. (2004). Why the sunny side is up: associations between affect and vertical position. *Psychol. Sci.* 15, 243–247.

Miller, G., and Johnson-Laird, P. (1976). *Language and Perception*. Cambridge, MA: The Belknap Press of Harvard University Press.

Nelson, K. (1996). *Language in Cognitive Development: Emergence of the Mediated Mind*. Cambridge: Cambridge University Press.

Noordzij, M. L., Neggers, S. F. W., Ramsey, N. F., and Postma, A. (2008). Neural correlates of locative prepositions. *Neuropsychologia* 46, 1576–1580.

Nunez, R., Motz, B., and Teuschner, U. (2006). Time after time: the psychological reality of the ego-and time-reference-point distinction in metaphorical construals of time. *Metaphor Symbol* 21, 133–146.

Nunez, R., and Sweetser, E. (2006). With the future behind them: convergent evidence from Aymara language and gesture in the cross-linguistic comparison of spatial construals of time. *Cogn. Sci.* 30, 133–146.

Pavio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford: Oxford University Press.

Pulvermuller, F., Hauk, O., Nikulin, V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797.

Regier, T., and Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends Cogn. Sci.* 13, 439–446.

Riordan, B., and Jones, M. N. (2010). Redundancy in perceptual and linguistic experience: comparing feature-based and distributional models of semantic representation. *Top. Cogn. Sci.* doi:10.1111/j.1756-8765.2010.01111.x. [Epub ahead of print].

Sakai, K., Ramnani, N., and Passingham, R. E. (2002). Learning of sequences of finger movements and timing: frontal lobe and action-oriented representation. *J. Neurophysiol.* 88, 2035–2046.

Santiago, J., Lupianez, J., Perez, E., and Funes, M. J. (2007). Time (also) flies from left to right. *Psychon. Bull. Rev.* 14, 512–516.

Schubotz, R. I., and von Cramon, D. Y. (2001). Interval and ordinal properties of sequences are associated with distinct premotor areas. *Cereb. Cortex* 11, 210–222.

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., and Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *J. Physiol. Paris* 102, 106–119.

Soanes, C., and Stevenson, A. (eds). (2007). *Concise Oxford English Dictionary*. Oxford, UK: Oxford University Press, 1728.

Sterelny, K. (1990). "*The imagery debate,*" *in Mind and Cognition*, ed. W. G. Lycan (Cambridge, MA: Blackwell), 607–626.

Talmy, L. (2000). *Towards a Cognitive Semantics: Concept Structuring Systems*. Cambridge, MA: The MIT Press.

Tenbrink, T. (2007). *Space, Time, and the use of Language: An Investigation of Relationships*. Berlin: Mouton de Gruyter.

Teuscher, U., McQuire, M., Collins, J., and Coulson, S. (2008). Congruity effects in time and space: behavioral and ERP measures. *Cogn. Sci.* 32, 563–578.

Torralbo, A., Santiago, J., and Lupianez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cogn. Sci.* 30, 745–757.

Vallar, G., and Perani, D. (1986). The anatomy of unilateral neglect after right-hemisphere stroke lesions – a clinical CT-scan correlation study in man. *Neuropsychologia* 24, 609–622.

Wencil, E. B., Coslett, H. B., Aguirre, J. K., and Chatterjee, A. (2010). Carving the clock at its component joints: neural bases for interval timing. *J. Neurophysiol.* 104, 160–168.

Wiener, M., Turkeltaub, P., and Coslett, H. B. (2010). The image of time: a voxel-wise meta-analysis. *Neuroimage* 49, 1728–1740.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Wu, D. H., Waller, S., and Chatterjee, A. (2007). The functional neuroanatomy of thematic role and locative relational knowledge. *J. Cogn. Neurosci.* 19, 1542–1555.

Zinken, J. (2009). "Temporal frames of reference," in *Language, Cognition, and Space*, eds P. Chilton and V. Evans (London: Equinox).

Zwaan, R. A. (2004). "The immersed experiencer: toward an embodied theory of language comprehension," in *The Psychology of Learning and Motivation*, Vol. 44, ed. B. H. Ross (New York: Academic Press), 35–62.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 June 2010; accepted: 20 December 2010; published online: 31 December 2010.

Citation: Kranjec A and Chatterjee A (2010) Are temporal concepts embodied? A challenge for cognitive neuroscience. Front. Psychology ***1****:240. doi: 10.3389/fpsyg.2010.00240*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Kranjec and Chatterjee. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 08 June 2011
doi: 10.3389/fpsyg.2011.00121

Numbers in space: differences between concrete and abstract situations

***Diane Pecher*[1]* and *Inge Boot*[1,2]**

[1] Brain and Cognition, Department of Psychology, Erasmus University Rotterdam, Rotterdam, Netherlands
[2] Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

Edited by:
Anna M. Borghi, University of Bologna and Institute of Cognitive Sciences and Technologies, Italy

Reviewed by:
Michael Kaschak, Florida State University, USA
Mariagrazia Ranzini, University of Bologna, Italy

***Correspondence:**
Diane Pecher, Department of Psychology, Erasmus University Rotterdam, T12-33, Post Box 1738, 3000 DR Rotterdam, Netherlands.
e-mail: pecher@fsw.eur.nl

Numbers might be understood by grounding in spatial orientation, where small numbers are represented as low or to the left and large numbers are represented as high or to the right. We presented numbers in concrete (*seven shoes in a shoe shop*) or abstract (29 – 7) contexts and asked participants to make relative magnitude judgments. Following the judgment a target letter was presented at the top or bottom (Experiments 1–3) or left or right (Experiment 4) of the visual field. Participants were better at identifying letters at congruent than incongruent locations, but this effect was obtained only when numbers were presented in concrete contexts. We conclude that spatial grounding might have a smaller role for numbers in abstract than in concrete context.

Keywords: numbers, quantity, context, grounded cognition, abstract concept

INTRODUCTION

If you would ask someone whether seven pairs of shoes is a lot or a little, the person would probably respond that this depends on the context. For example, it would be a little to have in a shoe shop but a lot to bring on a weekend to Paris (at least for sensible people). In order to evaluate magnitude one needs to mentally compare the number to a reference quantity such as an approximate number of shoes in a regular shoe shop or in an overnight bag. In the present study we investigated the mental representation of number magnitude. In particular, we compared to what extent numbers in concrete contexts (e.g., *seven pairs of shoes*) and numbers in abstract contexts (e.g., 7) have spatial representations.

Because number meaning can vary between very concrete and very abstract, depending on the context, the representation of numbers provides an interesting case for grounded theories of cognition (Glenberg, 1997; Barsalou, 1999; Lakoff and Núñez, 2000). On one end of the continuum numbers refer to perceptible quantities of concrete stuff such as shoes and oranges, and on the other end numbers are used to refer to things that have no perceptible referents in the world such as negative numbers and square roots. Much of the empirical support for the grounded cognition framework comes from studies that showed sensory–motor grounding for representations of concrete objects and actions, but so far not many studies have shown sensory–motor grounding for abstract concepts (Pecher et al., 2011). Since abstract concepts are an essential part of cognition, explaining how abstract concepts are represented by sensory–motor simulations is a critical challenge for the grounded framework (Machery, 2007; Mahon and Caramazza, 2008; Borghi et al., 2011; Dove, 2011).

Several researchers have proposed that abstract concepts are grounded in sensory–motor simulations by metaphorical mappings between abstract concept and concrete, spatial domains (Lakoff and Johnson, 1980, 1999; Lakoff, 1987; Gibbs, 1994, 2005). On this account, mental representations of abstract concepts take their structure from concrete image schemas (but see Kranjec and Chatterjee, 2010). For example, in natural language people often use spatial terms when they talk about time (e.g., *the meeting has been moved forward*), which suggests that representations for space and time are partially overlapping (see also Walsh, 2003). Initially, evidence for such mappings came from linguistic expressions, but more recently results from behavioral experiments have supported the cognitive metaphor theory (Boroditsky and Ramscar, 2002; Richardson et al., 2003; Meier and Robinson, 2004; Spivey et al., 2005; Casasanto, 2008; Van Dantzig, 2009; Boot and Pecher, 2010, 2011; Zanolie et al., 2011; Sell and Kaschak, 2011).

A second proposal explaining how abstract concepts are grounded in sensory–motor simulations posits an important role for situations. Barsalou and Wiemer-Hastings (2005) argued that the specific situations in which abstract concepts occur might be mentally simulated. To investigate this idea they asked participants to list properties for concepts at varying levels of abstraction. Their results confirmed that situational properties were important for abstract concepts. On this account, numbers might get their meaning from situational information such as when the number refers to a quantity of something (*seven pairs of shoes*) in a relevant context (*the shoe shop*). Therefore, numbers should have richer representations in concrete contexts than without such a context.

The metaphor explanation for number representations is supported by the spatial numerical association of response codes (SNARC) effect. Many researchers have suggested that numbers are represented visuo-spatially on a mental number line. The best known effect is the horizontal SNARC effect (e.g., Dehaene et al., 1993) although there are a few studies that also showed a vertical SNARC effect for hand responses (Ito and Hatta, 2004; Gevers et al., 2006; Ristic et al., 2006) and eye movement responses (Schwartz and Keus, 2004). Participants respond faster with the left hand to low numbers than to high number, and faster with the right hand to high numbers than to low numbers (see Gevers et al., 2010

for similar effects with verbal responses) even when magnitude is irrelevant to the task, as in parity judgment. Such findings could be interpreted as showing that participants represent numbers along a horizontal mental number line with small numbers on the left and large numbers on the right of the continuum. On an alternative account, however, the SNARC effect could be attributed to processes that occur during response selection (Keus and Schwarz, 2005; Keus et al., 2005; Müller and Schwarz, 2007) or to polarity alignment (Proctor and Cho, 2006; Landy et al., 2008; Santens and Gevers, 2008; Bae et al., 2009). If the effect can be attributed to response selection, no underlying mental representation needs to be assumed to explain the SNARC effect. Thus, at present there is no agreement yet as to whether the SNARC effect indicates a spatial representation of numbers.

Stronger evidence for a spatial representation of numbers is provided by studies that show effects of number magnitude on spatial attention (Fischer et al., 2003; Nicholls et al., 2008; Salillas et al., 2008). Fischer et al. (2003) found that participants were faster to detect a target at the left when it was preceded by a low digit (one or two) and faster to detect a target at the right when preceded by a high digit (eight or nine). These attention effects indicate automatic activation of a spatial representation of numbers on a horizontal dimension. Unfortunately, the effect on spatial attention is not obtained very consistently (Galfano et al., 2006; Ristic et al., 2006; Lindemann et al., 2008; Zanolie and Pecher, 2011). In the attentional paradigm spatial information might be less salient than in the traditional SNARC studies because spatial information is not relevant for the response. Zanolie and Pecher (2011) showed that the effect of number magnitude on spatial attention critically depends on the relevance of magnitude in the task context. Thus, there is some evidence that number representation affects spatial attention although the effect is probably due to active rather than automatic processing of number magnitude.

The SNARC effect has been investigated mostly in the horizontal direction. Linguistic sources, however, provide very little evidence for a horizontal mental number line but much evidence for a vertical mental number line. In natural language vertical words are often used to talk about magnitude. For example, people might say *prices are high, mortgage rates dropped, incomes can rise* or *heat is turned up or down* (Johnson, 1987; Lakoff and Núñez, 2000; Bergen et al., 2007). Why would they not say that *incomes are moving to the right* and *heat is turned left*? The systematic way in which vertical but not horizontal terms are used for magnitude suggests that magnitude is represented in terms of verticality (Lakoff and Johnson, 1980). Therefore, it may be the case that in representations of magnitude vertical orientation is more important than horizontal orientation. In the present study we investigated whether thinking about magnitude directed spatial attention vertically or horizontally.

A second goal of the present study was to investigate whether spatial attention is affected differently by magnitude in concrete (e.g., *seven pairs of shoes in a shop*) and abstract (e.g., seven) contexts. So far, most evidence for spatial number representations has been obtained with numbers in abstract contexts. In the present study we presented numbers either in a concrete, natural context or in a more abstract context of just other numbers. On the one hand we might expect that numbers in concrete contexts are more closely associated to spatial directions, because they refer to concrete magnitudes such as a row of shoes or a stack of books that actually have spatial dimensions. A study by Bergen et al. (2007) suggests that vertical spatial representations might be restricted to situations and events that refer to concrete vertical space. They investigated the effect of described motion on spatial attention along the vertical midline. They found that sentences with concrete movements (e.g., *the mule climbed*) interfered with processing of an unrelated visual stimulus, but no such effect was found for sentences with abstract movements (e.g., *the cost climbed*). Bergen et al. (2007) argued that representations of concrete situations interfered with visual processing because they use the same resources. Abstract representations do not involve visual simulation, however, and thus do not interfere with visual processing.

On the other hand, numbers in abstract contexts may need grounding in the spatial domain precisely because they lack perceptual properties. Several studies have obtained effects of abstract representations on spatial attention, suggesting that these representations are grounded in spatial image schemas (Richardson et al., 2003; Meier and Robinson, 2004; Van Dantzig, 2009). These findings are predicted by cognitive metaphor theorists, who claim that abstract concepts are understood by metaphorical mappings between abstract and concrete domains. Whereas mental simulation of visual details interferes with perceptual processing, effects of image schemas on spatial attention should facilitate processing of visual stimuli (Estes et al., 2008).

In the present study we investigated how processing of visual stimuli at different locations is affected by numbers in concrete and abstract contexts. Because the spatial position of a number is relative to other numbers (Reynvoet and Brysbaert, 1999; Nathan et al., 2009), we asked participants to make relative magnitude judgments to numbers in either a concrete or an abstract context. To assess the activation of spatial image schemas the magnitude judgment was followed by an unrelated stimulus (the letter *p* or *q*) in a spatial location that was either congruent or incongruent with the image schematic location of the number. If spatial attention is affected by number magnitude, identification of the letter should be facilitated in congruent positions compared to incongruent positions.

The letter identification task has been used in numerous studies that investigated effects on spatial attention (e.g., Meier and Robinson, 2004; Meier et al., 2007; Estes et al., 2008; Van Dantzig, 2009). The advantage of using a secondary task is that in the experimental set-up magnitude judgment (the first task) is not related to spatial position, and letter identification (the second task) is not related to either spatial position or magnitude judgment. Proctor and Cho (2006) have suggested that congruency effects in binary tasks can be explained by polarity alignment. This effect occurs if stimulus and response dimensions have polarity, that is, can be coded as + (plus) and – (minus). Proctor and Cho (2006), but see Pecher et al., 2010 assume that in spatial dimensions, above and right are coded as +. They also assume that "yes" responses are coded as +. Therefore, the polarity principle predicts that "yes" responses will be faster when they are aligned with stimuli or response sides that are also + (i.e., above or right). In the letter identification task such alignment effects will not play a role because the "*p*" or "*q*" response does not have an obvious polarity. Even if participants would code the responses as + and –, however, this would not affect the overall results, because "*p*" and "*q*" responses are collapsed.

EXPERIMENT 1

In Experiment 1 we presented sentences in which a number was presented in a natural context. Participants decided whether the magnitude was relatively *few* or *many* (the Dutch words *weinig* and *veel* were used, which do not refer to height and can be used for both mass and count nouns). Whether the magnitude was relatively few or many depended on the situation, because the same absolute numbers appeared in both conditions (e.g., *The man had two books in his bookcase* vs. *The man read two books a day*). Thus, participants had to use the situation in order to make a magnitude judgment. Magnitude judgments were followed by identification of a letter *p* or *q* which appeared at the top or bottom of the screen (as in the study by Meier and Robinson, 2004).

If magnitude is represented by vertical position we expected that attention would shift to the top for relatively high magnitudes and to the bottom for relatively low magnitudes. This attention shift should facilitate identification of targets in the congruent position compared to targets in the incongruent position. It should be noted that we did not use quantities that explicitly referred to vertical positions (e.g., we did not use items like *the tree was 6 meters tall*). Neither did we use words that referred to vertical movement or direction (e.g., *dropped*, *up*). Thus, any effect on vertical attention was due to the representation of the concept magnitude rather than to the literal orientation of actions or objects mentioned in the sentence.

METHODS

Participants

Thirty-two undergraduate psychology students at the Erasmus University Rotterdam received course credit for participating. All participants were fluent speakers of Dutch.

Materials

All instructions and materials were in Dutch. We created 40 sentences in the few condition containing a relatively small quantity and 40 sentences in the many condition containing a relatively large quantity (examples are shown in the Appendix). The same absolute quantities were used in the few and many conditions, so whether it was *few* or *many* depended on the sentence context and not on the absolute number. Different kinds of quantity indications were used (e.g., *100 meters*, *10 minutes*, *10 euro*). No words were used that referred to vertical position or movement. Across participants, sentences were counterbalanced over the target letter identity, target letter position, and block order. Each combination of letter identity and letter position was used equally often in each condition and block. Twenty-one additional sentences were created for practice and instruction.

Procedure

Participants were tested individually on PCs separated by walls. Participants responded by pressing a button on a response box that had five aligned buttons. Four different response mappings were used, such that participants used two fingers of each hand to respond, and used one hand for each task so that the two fingers of the same hand were used to make the two responses in one task. Response mapping was varied between subjects. Buttons were labeled with the letter *v* (*veel* – Dutch for *many*), the letter *w* (*weinig* – Dutch for *few*), the letter *p*, and the letter *q*. Participants were instructed to make magnitude judgments by choosing *few* or *many*. A sentence was presented in the center of the screen until a response was made. Then, 200 ms after the response, the letter *p* or *q* (28 pt Garamond) was presented at the top or bottom of the screen which had to be identified as quickly and accurately as possible by pressing the corresponding button. The next trial started 500 ms after the response to the target was made or after the feedback. Feedback (*fout* – Dutch for *error*) was provided for 1500 ms after incorrect responses to the letter targets. Twenty practice trials preceded the 80 experimental trials. The experimental sentences were presented in random order in two blocks, with the restriction that sentences that contained the same quantity were presented in different blocks. Between the two blocks participants could take a break and feedback on their accuracy in the first block was provided.

RESULTS

Four items were removed from the analysis because fewer than 60% of the participants gave the intended magnitude response (Two in the many and two in the few condition). We calculated mean reaction times and accuracy for the letter identification responses. Incorrect responses (to prime or target, 8.1%) and responses more than 2 SD from the subject's mean (4.2%) were excluded from the reaction time analyses. Mean reaction times and error rates are presented in **Figure 1**. In the reaction times we obtained a significant interaction between magnitude and vertical position, $F(1, 31) = 4.75$, MSE = 732.6, $\eta^2 = 0.13$, $p < 0.05$. Participants were faster to identify a letter presented at the top when they first read a sentence in which the magnitude was many compared to a few, $t(31) = 5.08$, SEM = 5.08, $\eta^2 = 0.24$, $p < 0.0001$. Responses to letters at the bottom of the screen were not different between conditions, $t(31) = 1.49$, SEM = 5.1, $\eta^2 = 0.03$, $p > 0.10$. Responses to letters were faster in the many than in the few condition, $F(1, 31) = 4.58$, MSE = 1478.7, $\eta^2 = 0.13$, $p < 0.05$, and responses were faster to letters at the top than the bottom of the screen,, $F(1, 31) = 5.38$, MSE = 3421, $\eta^2 = 0.15$, $p < 0.05$.

In the error rates we also obtained an interaction effect between position and magnitude, $F(1, 31) = 5.06$, MSE = 0.000, $\eta^2 = 0.14$, $p < 0.05$. None of the other effects reached significance.

EXPERIMENT 2

Next we investigated if the magnitude-position congruency effect could be replicated with quantities in an abstract context of just other numbers. As in Experiment 1, participants made relative magnitude judgments. In Experiment 2 the numbers 0–100 were used and participants judged magnitude by comparing each number to 50.

METHOD

Participants

Participants were 32 students from the same pool who had not participated in Experiment 1.

Materials

The numbers 0–49 and 51–100 were used.

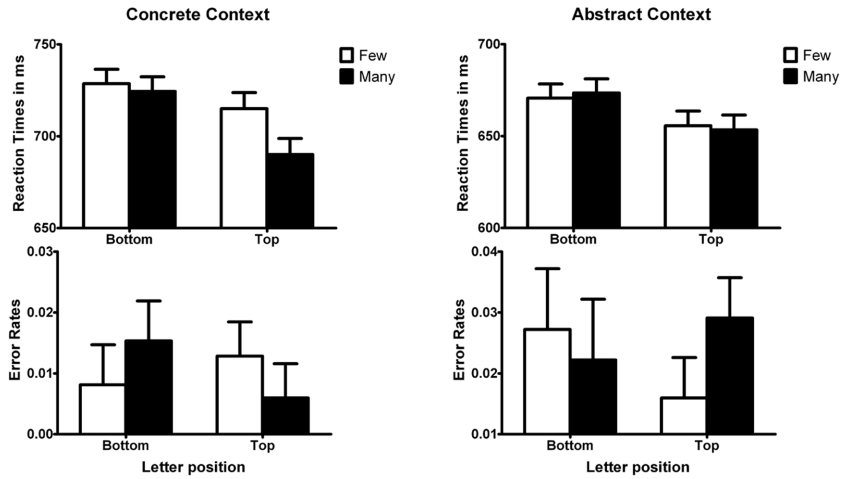


FIGURE 1 | Reaction times and error rates for the letter identification task in Experiments 1 (Concrete Context) and 2 (Abstract Context). Error bars represent standard error of the mean difference for adjacent bars.

Procedure

The procedure was the same as in Experiment 1, except that prior to the letter identification task participants judged whether the number was a few or many compared to 50. Twenty numbers were used for practice; the remaining 80 numbers were used as experimental trials. Counterbalancing was the same as in Experiment 1, and in addition odd and even numbers were also counterbalanced across letter position.

RESULTS

The same data analysis procedure was used as for Experiment 1. The results are displayed in **Figure 1**. Incorrect responses (5.3%) and remaining outliers (4.6%) were removed. In the reaction times we obtained no interaction effect between magnitude and vertical position, $F < 1$ nor in the error rates, $F(1, 31) = 1.13$, MSE = 0.001, $\eta^2 = 0.04$, $p > 0.25$.

EXPERIMENT 3

In Experiment 1 we found that number magnitude directed spatial attention, whereas no such effect was observed in Experiment 2. The main difference between the experiments was that the numbers in Experiment 1 referred to concrete situations and in Experiment 2 to abstract situations. This suggests that the vertical representation of magnitude is activated more strongly when people represent numbers that refer to concrete situations than numbers that refer to abstract situations. Besides abstractness, however, there were a few other procedural differences between the two experiments. In order to directly compare the two types of number representation we ran an additional experiment in which we made the two tasks more similar. In particular, we controlled for two differences between Experiments 1 and 2. First, in Experiment 1 the reference point (the "normal" magnitude in that context) was implicit, whereas it was explicitly provided in Experiment 2. Second, the reference point was variable in Experiment 1 because it depended on the context. In Experiment 2, however, it was fixed (i.e., it was always 50). In Experiment 3 we presented numbers in both concrete and abstract contexts, and an explicit reference point was given for both types of stimuli. In addition, the value of the reference point varied between trials for both types of stimuli.

METHOD

Participants

Thirty-two students from the same pool who had not participated in Experiments 1 and 2 participated for course credits.

Materials

The same materials as in Experiments 1 and 2 were used. In addition, all prime sentences and numbers (on which the magnitude judgment was made) were now preceded by a reference sentence or reference number that provided an explicit reference point for the prime. Thus, the order of events for each trial was: reference – magnitude stimulus – letter stimulus. For each magnitude sentence from Experiment 1 we created a referent sentences which contained a different quantity in the same context (e.g., *60,000 people can be seated in the stadium*) as the magnitude sentence (e.g., *80 people were seated in the stadium*).

To prevent participants simply comparing the two quantities from referent and magnitude sentence without considering the situation half of the referent sentences were replaced by fillers that had the same quantity in a different context (e.g., *80 people attended the business meeting*). Sentences used as experimental trials and fillers were counterbalanced between subjects. For each magnitude number from Experiment 2 we selected two different referent numbers. Referent numbers could be at large or small numerical distances from the magnitude numbers, and were chosen such that the absolute size of the referent was not predictive of the response.

In both blocks the target task was identification of the letters *p* and *q* at the top or bottom of the screen as in Experiments 1 and 2.

Procedure

The same procedures as in Experiment 1 and 2 were used. One block had the same procedure as Experiment 1 except that the magnitude sentence was preceded by a referent sentence. The other block had

the same procedure as in Experiment 2 except that each magnitude number was preceded by a referent number. Participants were instructed to press the middle button (labeled *sentence1/number1*) after they processed the referent (sentence or number) and to use it as a comparison for the next stimulus. After making the magnitude judgment the target letter *p* or *q* was presented at the top or bottom of the screen. Assignment of sentences to experimental and filler trials was counterbalanced across participants, as was the order of sentence and number blocks. At the end of the experiment we asked participants whether they had any hypotheses about the purpose of the experiment.

RESULTS

None of the participants guessed the purpose of the experiment correctly. The mean reaction times and error rates are displayed in **Figure 2**. The same data analysis procedure was used as for Experiments 1 and 2. Incorrect responses (4.4% in the sentence condition and 5.0% in the number condition) and remaining outliers (3.8% in the sentence condition and 7.0% in the number condition) were removed.

The results of Experiment 3 are consistent with those of Experiments 1 and 2. The three-way interaction effect (magnitude × vertical position × type of quantity) approached significance in the reaction times, $F(1, 63) = 2.86$, MSE = 1630.3, $\eta^2 = 0.043$ $p = 0.096$ but not in the error rates, $F < 1$. This marginal interaction indicated that the interaction effect between magnitude and vertical position was different between numbers in the two types of contexts.

For trials that were preceded by numbers in concrete contexts, the interaction effect between magnitude and vertical position approached significance, $F(1, 63) = 3.60$, MSE = 2349.9, $\eta^2 = 0.054$, $p = 0.062$. Participants responded faster to a letter at the top when they first read a sentence in the many condition compared to the few condition, $t(63) = 2.22$, SEM = 6.06, $\eta^2 = 0.042$, $p < 0.05$, whereas participants responded faster to a letter at the bottom when they first read a sentence in the few compared to the many condition, although this difference was not statistically significant, $t(63) = 1.57$, SEM = 6.06, $\eta^2 = 0.012$, $p > 0.05$. In the error rates we obtained no significant interaction effect, $F < 1$.

For trials that were preceded by numbers in abstract contexts, there was no significant interaction between magnitude and vertical position in the reaction times, $F < 1$, or in the error rates, $F(1, 63) = 1.65$, MSE = 0.001, $\eta^2 = 0.025$, $p > 0.20$. A main effect of vertical position showed that participants responded faster to letters at the bottom than at the top, $F(1, 63) = 8.24$, MSE = 1672.8 $\eta^2 = 0.12$, $p < 0.01$.

EXPERIMENT 4

The previous results suggest that numbers in concrete contexts direct vertical spatial attention but numbers in abstract contexts do not. A possible explanation is that numbers in concrete situations have a mainly vertical orientation, whereas numbers in abstract situations have a mainly horizontal orientation. This explanation would be consistent with the SNARC effect and with effects of number magnitude on spatial attention, which are usually obtained with numbers in abstract contexts. The effects of number on spatial attention are not very robust, however, and appear to depend on how strongly numerical magnitude information is activated by the number (Galfano et al., 2006; Ristic et al., 2006; Lindemann et al., 2008; Zanolie and Pecher, 2011). An alternative explanation is that numbers in concrete contexts activate spatial representations more strongly than numbers in abstract contexts, irrespective of the spatial orientation. To distinguish between these two explanations we used the same design as in Experiment 3 except that the letters in the letter identification task were presented at the left and right rather than top and bottom of the screen.

METHOD

Participants

Thirty-one students from the same pool who had not participated in any of the previous experiments participated for course credits.

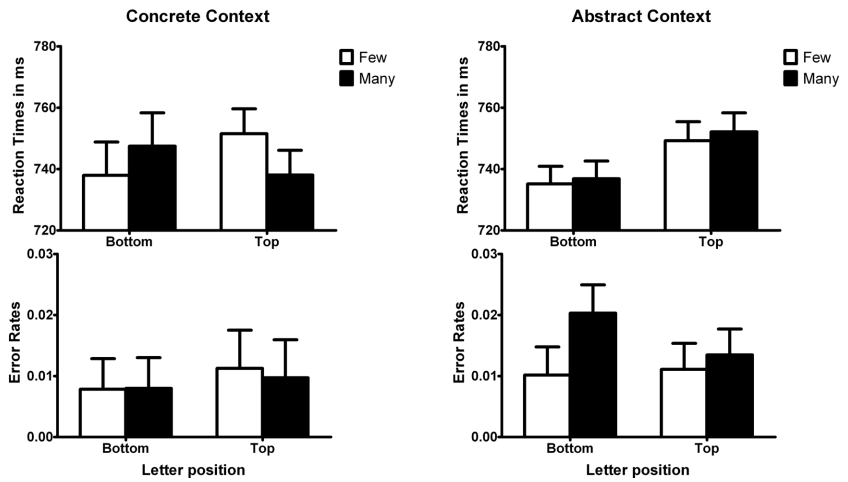


FIGURE 2 | Reaction times and error rates for the letter identification task in Experiment 3. Error bars represent standard error of the mean difference for adjacent bars.

Materials and procedure

The materials and procedure were the same as in Experiment 3 with two exceptions. Most important, in the target task the letters *p* and *q* were presented at the left and right on the computer screen (in the vertical center position). In addition, for the magnitude sentences we used only the same context referent sentence. Twenty additional fillers were created in which the magnitude sentence was preceded by a same number, different context referent sentence.

RESULTS

None of the participants guessed the purpose of the experiment correctly. The mean reaction times and error rates are displayed in **Figure 3**. The same data analysis procedure was used as for the previous Experiments. Incorrect responses (3.1% in both sentence and number condition) and remaining outliers (9.9% in the sentence condition and 8.3% in the number condition) were removed.

The results of Experiment 4 basically replicated those of Experiment 3 and showed similar effects for horizontal and vertical spatial orientation. The three-way interaction effect (magnitude × horizontal position × type of quantity) approached significance, $F(1, 30) = 3.07$, MSE = 5135.2, $\eta^2 = 0.093$ $p = 0.090$ for reaction times and $F(1, 30) = 2.77$, MSE = 0.0001, $\eta^2 = 0.084$ $p = 0.011$ for error rates. This interaction indicated that the interaction effect between magnitude and horizontal position was different between the two types of contexts.

For trials that were preceded by numbers in concrete contexts, the interaction effect between magnitude and horizontal position was significant, $F(1, 30) = 9.02$, MSE = 1699.1, $\eta^2 = 0.23$, $p < 0.01$. Participants responded faster to a letter at the left when they first read a sentence in the few condition compared to the many condition, $t(30) = 2.41$, SEM = 12.00, $p < 0.05$, whereas participants responded faster to a letter at the right when they first read a sentence in the many compared to the few condition, $t(30) = 1.80$, SEM = 8.63, $p = 0.08$. Overall, participants responded faster to letters at the left than at the right, $F(1, 30) = 10.62$, MSE = 5406.9, $\eta^2 = 0.26$, $p < 0.01$. In the error rates we obtained no significant interaction effect, $F < 1$.

For trials that were preceded by numbers in abstract contexts, there was no significant interaction between magnitude and horizontal position, $F < 1$ for reaction times and $F(1, 30) = 1.17$, $p > 0.20$ for error rates. A main effect of magnitude showed that participants responded faster to letters following a "few" decision than following a "many" decision, $F(1, 30) = 11.58$, MSE = 3755.3 $\eta^2 = 0.28$, $p < 0.01$.

GENERAL DISCUSSION

In four experiments we investigated the effect of magnitude representations on spatial attention. Participants made magnitude decisions to numbers presented in a concrete or abstract context. Subsequently, a target letter was presented at the top or bottom (Experiments 1–3) or the left or right (Experiment 4) of the computer screen. We found that identification of target letters was influenced by the magnitude decision. Lower magnitudes directed attention to the bottom or left of the screen and higher magnitudes directed attention to the top or right of the screen. This was only the case, however, when magnitude decisions were made to numbers in concrete contexts. When numbers were presented in abstract contexts no effect of magnitude on spatial attention was observed.

The interaction between magnitude and spatial position was not due to the response itself, as has been suggested for the SNARC effect (Keus and Schwarz, 2005; Keus et al., 2005; Müller and Schwarz, 2007; Landy et al., 2008; Santens and Gevers, 2008; Bae et al., 2009). In our experiments, the target response (*p* or *q*) was unrelated to letter position, magnitude decision, and the congruency between position and quantity. Even if participants had used a response mapping between the magnitude response (*few* or *many*) and response on the target (*p* or *q*) this cannot explain the results because the data in each condition are based on equal numbers of *p* and *q* responses. Thus, any positive and negative effects of such mapping would have been collapsed into one average for each condition. Therefore, the interaction between magnitude and spatial position on target identification is better explained by differences in spatial attention for low and high magnitudes.

Our finding that the effect of magnitude on spatial attention was larger in concrete than abstract contexts was surprising for two reasons. First, some researchers have argued that abstract concepts

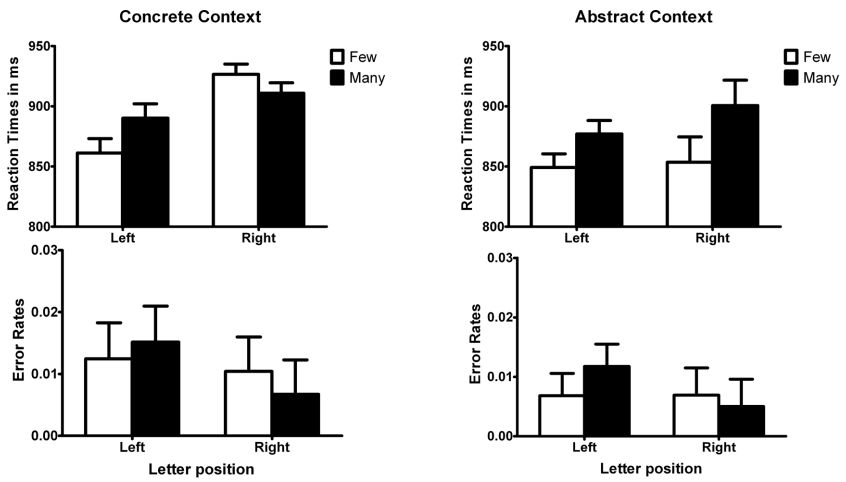


FIGURE 3 | Reaction times and error rates for the letter identification task in Experiment 4. Error bars represent standard error of the mean difference for adjacent bars.

are grounded in sensory–motor processing by metaphorical mappings (Lakoff and Johnson, 1980, 1999; Lakoff, 1987; Gibbs, 1994, 2005). On this account, abstract concepts are understood by metaphorical mappings between the abstract concept and concrete, spatial domains. Such mappings should be more essential to understand numbers in abstract contexts than in concrete contexts. Abstract numbers have no grounding in sensory–motor processing whereas numbers in concrete contexts can be simulated by sensory–motor processing. Thus, abstract numbers should activate image schemas to a larger extent than concrete numbers. The present results do not support this account because we showed that numbers in concrete contexts have a larger effect on both vertical and horizontal spatial attention than numbers in abstract contexts. A possible metaphorical explanation might be that alternative mappings were used for abstract numbers. For example, magnitude might be represented as motion along a path (Lakoff and Núñez, 2000) or as values in a non-spatial sensory domain (Núñez et al., in press). Although we cannot exclude these alternatives based on the present data, the spatial number line appears to be the most "natural" grounding, at least for educated Europeans (Núñez et al., in press). Thus, it seems reasonable to assume that *if* participants in our study used metaphorical mappings these should most likely have been vertical or horizontal.

Second, even if the effect on spatial attention was the same for concrete and abstract magnitudes, the positive effect of concrete magnitudes should have been opposed by a negative interference effect. Spatial congruency effects might be explained by two opposing mechanisms (Estes et al., 2008). First, a representation might direct visual attention to the congruent spatial location just as arrows or verbal commands (*left*, *up*) do. As a result, target processing at the congruent location is facilitated compared to target processing at incongruent locations. Second, simulation of perceptual experiences with objects might occupy the same resources as those needed to identify the visual stimulus. This will cause interference similar to the Perky effect (Craver-Lemley and Reeves, 1992; Bergen et al., 2007) at the congruent location. For example, mental simulation of *cowboy hat* will interfere more with perception of stimuli in a high position than mental simulation of *cowboy boot*. Interference should only be expected for concrete and not for abstract concepts, because abstract concepts do not have the perceptual details that might compete for resources with visual perception. Facilitation, on the other hand, might be expected for both concrete and abstract concepts. On this account, more facilitation is expected for abstract than concrete magnitudes, because for the concrete magnitudes the effect is counteracted by perceptual interference.

Our finding of larger facilitation for concrete than abstract magnitudes seems at odds with this idea that concrete concepts cause more visual interference than abstract concepts. The contexts described by our stimuli, however, did not have a specific spatial direction. For example, *seven pairs of shoes in a shoe shop* might have any spatial layout. Therefore, any visual interference caused by perceptual details of the concrete contexts would not be systematically related to magnitude and thus would not affect the interaction between magnitude and letter position. Although our results thus do not necessarily contradict the interference account, it does not explain why concrete magnitude had a larger effect on spatial attention than abstract magnitude.

It is possible that magnitude judgments in concrete contexts required deeper or richer processing than those in abstract contexts. In that case, our results would be consistent with findings that spatial attention induced by number magnitude is weak and depends on the relevance and depth of number processing in the task (Galfano et al., 2006; Ristic et al., 2006; Lindemann et al., 2008; Zanolie and Pecher, 2011). Judging the magnitude of seven pairs of shoes in a shop might result in a richer representation of the relative magnitude than judging the magnitude of 7 compared to 23. With this assumption differences in depth of processing might explain the differences between abstract and concrete context. Because our paradigm differed in several ways from those used in prior studies one may ask whether previous findings have any bearing on our results. First, the response categories, few and many, are quantifiers. One could argue that quantifiers are processed differently than absolute numbers, for example because they are less exact. However, the stimuli still contained exact numbers, so it seems unlikely that participants did not process exact numerical information. Second, attention was measured in a letter identification task rather than a target detection task (i.e., present/absent decision). Although both tasks should be affected by manipulations of spatial attention, they may still differ in how they are affected by unrelated tasks. Third, in previous studies the interval between presentation of the number and that of the target was short. In our study, the target stimulus was time-locked to the magnitude decision and thus varied with decision speed. Bearing in mind these differences, we tentatively argue that the effect of magnitude on spatial attention appears to depend on the depth and relevance of number meaning activation.

The assumption that grounding in spatial representations depends on the depth of number representations leads to the question how more superficial representations can still result in good performance. We propose that experienced number users have developed shortcuts that allow them to perform certain tasks without activating a rich and grounded meaning of numbers. For example, after learning the multiplication tables students can answer multiplication questions without even fully understanding the concept of multiplication. Such associative knowledge might be sufficient to perform simple tasks (such as magnitude judgment) in abstract contexts. Barsalou et al. (2008; see also Simmons et al., 2008) have proposed a similar mechanism for language processing. They argue that in some linguistic tasks, participants do not need to fully simulate a concept but instead can base their responses on simple word associations. It is possible that experienced number users can rely on a similar set of associative links between numbers in simple number tasks. This idea is supported by findings that participants sometimes treat numbers represented by fingers as symbolic (Di Luca et al., 2010). Thus, even though the exact number is presented in an analog form processing benefits from a symbolic representation. Such findings indicate that number processing using associative shortcuts is possible and in some cases perhaps even more efficient, for example in simple or over-learned tasks. When more meaningful processing is required, however, as with numbers in a more concrete context, representations involve sensory–motor simulations, including metaphorical mappings.

ACKNOWLEDGMENTS

This research was supported by a VIDI grant from the Netherlands Organization for Scientific Research (NWO) to Diane Pecher. We would like to thank Rafael Núñez for pleasant and useful exchanges of ideas.

REFERENCES

Bae, G. Y., Choi, J. M., Cho, Y. S., and Proctor, R. W. (2009). Transfer of magnitude and spatial mappings to the SNARC effect for parity judgments. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1506–1521.

Barsalou, L. W. (1999). Perceptual symbol system. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford: Oxford University Press), 245–283.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 129–163.

Bergen, B. K., Lindsay, S., Matlock, T., and Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cogn. Sci.* 31, 733–764.

Boot, I., and Pecher, D. (2010). Similarity is closeness: metaphorical mapping in a perceptual task. *Q. J. Exp. Psychol. (Colchester)* 63, 942–954.

Boot, I., and Pecher, D. (2011). Representation of categories: metaphorical use of the container schema. *Exp. Psychol.* 58, 162–170.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., and Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Front. Psychol.* 2:15. doi: 10.3389/fpsyg.2011.00015

Boroditsky, L., and Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychol. Sci.* 13, 185–189.

Casasanto, D. (2008). Similarity and proximity: when does close in space mean close in mind? *Mem. Cognit.* 36, 1047–1056.

Craver-Lemley, C., and Reeves, A. (1992). How visual imagery interferes with vision. *Psychol. Rev.* 99, 633–649.

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and the number magnitude. *J. Exp. Psychol. Gen.* 122, 372–396.

Di Luca, S., Lefèvre, N., and Pesenti, M. (2010). Place and summation coding for canonical and non-canonical finger numeral representations. *Cognition* 117, 95–100.

Dove, G. (2011). On the need for embodied and dis-embodied cognition. *Front. Cogn.* 1:242. doi: 10.3389/fpsyg.2010.00242

Estes, Z., Verges, M., and Barsalou, L. W. (2008). Head up, foot down: object words orient attention to the objects' typical location. *Psychol. Sci.* 19, 93–97.

Fischer, M., H., Castel, A. D., Dodd, M. D., and Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nat. Neurosci.* 6, 555–556.

Galfano, G., Rusconi, E., and Umiltà, C. (2006). Number magnitude orients attention, but not against one's will. *Psychon. Bull. Rev.* 13, 869–874.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., and Fias, W. (2006). Automatic response activation of implicit spatial information: evidence from the SNARC effect. *Acta Psychol. (Amst.)* 122, 221–233.

Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., and verguts, T. (2010). Verbal-spatial and visuospatial coding of number-space interactions. *J. Exp. Psychol. Gen.* 139, 180–190.

Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. New York: Cambridge University Press.

Gibbs, R. W. J. (2005). "The psychological status of image schemas," in *From Perception to Meaning*, ed. B. Hampe (Berlin: Mouton de Gruyter), 113–135.

Glenberg, A. M. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Ito, Y., and Hatta, T. (2004). Spatial structure of quantitative representation of numbers: evidence from the SNARC effect. *Mem. Cognit.* 32, 662–673.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Reason and Imagination*. Chicago: University of Chicago Press.

Keus, I. M., Jenks, K. M., and Schwarz, W. (2005). Psychophysiological evidence that the SNARC effect has its functional locus in a response selection stage. *Brain Res. Cogn. Brain Res.* 24, 48–56.

Keus, I. M., and Schwarz, W. (2005). Searching for the functional locus of the SNARC effect: Evidence for a response-related origin. *Mem. Cognit.* 33, 681–695.

Kranjec, A., and Chatterjee, A. (2010). Are temporal concepts embodied? A challenge for cognitive neuroscience. *Front. Psychol.* 1:240. doi: 10.3389/fpsyg.2010.00240

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.

Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago: Chicago University Press.

Lakoff, G., and Johnson, M. (1999). *The Philosophy in the Flesh: The Embodied Mind and its Challenge to the Western Thought*. New York, NY: Basic Books.

Lakoff, G., and Núñez, R. E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.

Landy, D. H., Jones, E. L., and Hummel, J. E. (2008). "Why spatial-numeric associations aren't evidence for a mental number line," in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Austin, TX, 357–362.

Lindemann, O., Abolafia, J. M., Pratt, J., and Bekkering, H. (2008). Coding strategies in number space: memory requirements influence spatial-numerical associations. *Q. J. Exp. Psychol.* 61, 515–524.

Machery, E. (2007). Concept empiricism: a methodological critique. *Cognition* 104, 19–46.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., and Schjeldahl, K. (2007). What's "up" with god? Vertical space as a representation of the divine. *J. Pers. Soc. Psychol.* 93, 699–710.

Meier, B. P., and Robinson, M. D. (2004). Why the sunny side is up: association between affect and vertical position. *Psychol. Sci.* 15, 243–247.

Müller, D., and Schwarz, W. (2007). Is there an internal association of numbers to hands? The task set influences the nature of the SNARC effect. *Mem. Cognit.* 35, 1151–1161.

Nathan, M. B., Shaki, S., Salti, M., and Algom, D. (2009). Numbers and space: associations and dissociations. *Psychon. Bull. Rev.* 16, 578–582.

Nicholls, M. E. R., Loftus, A. M., and Gevers, W. (2008). Look, no hands: a perceptual task shows that number magnitude induces shift of attention. *Psychon. Bull. Rev.* 15, 413–418.

Núñez, R., Doan, D., and Nikoulina, A. (in press). Squeezing, (striking), and vocalizing: is number representation fundamentally spatial? *Cognition.*

Pecher, D., Boot, I., and van Dantzig, S. (2011). "Abstract concepts: sensory-motor grounding, metaphors, and beyond," in *The Psychology of Learning and Motivation*, ed. B. Ross (Burlington: Academic Press), 217–248.

Pecher, D., Van Dantzig, S., Boot, I., Zanolie, K., and Huber, D. E. (2010). Congruency between word position and meaning is caused by task induced spatial attention. *Front. Cogn.* 1:30. doi: 10.3389/fpsyg.2010.00030

Proctor, R. W., and Cho, Y. S. (2006). Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychol. Bull.* 132, 416–442.

Reynvoet, B., and Brysbaert, M. (1999). Single-digit and two-digit Arabic numerals address the same semantic number line. *Cognition* 72, 191–201.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., and McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cogn. Sci.* 27, 767–780.

Ristic, J., Wright, A., and Kingstone, A. (2006). The number line effect reflects top-down control. *Psychon. Bull. Rev.* 13, 862–868.

Salillas, E., El Yagoubi, R., and Semenza, C. (2008). Sensory and cognitive processes of shifts of spatial attention induced by numbers: an ERP study. *Cortex* 44, 406–413.

Santens, S., and Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition* 108, 263–270.

Schwartz, W., and Keus, I. M. (2004). Moving the eyes along the mental number line: comparing SNARC effects with saccadic and manual responses. *Percept. Psychophys.* 66, 651–664.

Sell, A. J., and Kaschak, M. P. (2011). Processing time shifts affects the execution of motor responses. *Brain Lang.* 117, 39–44.

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., and Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *J. Physiol. Paris* 102, 106–119.

Spivey, M. J., Richardson, D. C., and Gonzalez-Marquez, M. (2005). "On the perceptual-motor and image-schematic infrastructure of language," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan (Cambridge: Cambridge University Press), 246–281.

Van Dantzig, S. (2009). *Mind the Body: Grounding Conceptual Knowledge in Perception and Action*. Rotterdam: Erasmus University.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci. (Regul. Ed.)* 7, 483–488.

Zanolie, K., Van Dantzig, S., Boot, I., Wijnen, J., Schubert, T. W., Giessner, S., and Pecher, D. (2011). Mighty metaphors: behavioral and ERP

evidence that power shifts attention on a vertical dimension. Manuscript submitted for publication.

Zanolie, K., and Pecher, D. (2011). Number-induced shifts in spatial attention: the necessity of magnitude information. Manuscript submitted for publication.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 July 2010; accepted: 27 May 2011; published online: 08 June 2011.

Citation: Pecher D and Boot I (2011) Numbers in space: differences between concrete and abstract situations. Front. Psychology ***2****:121. doi: 10.3389/fpsyg.2011.00121*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Pecher and Boot. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

APPENDIX

Sentences used in Experiments 1, 3, and 4. Approximate translations are given in brackets.

Few	Many
De oude man had 2 boeken in de kast staan.	De oude man las 2 boeken per dag.
[The old man had 2 books in his bookcase.]	*[The old man read 2 books a day.]*
Mijn oma woont 100 meter van mijn huis vandaan.	Er stond een rij van 100 meter voor de kassa.
[My grandma lives 100 meters from my house.]	*[There was a 100 meter line at the check-out.]*
De bouwvakker dronk op de snikhete dag 1 liter water.	Het meisje dronk in een keer 1 liter cola op.
[On the sweltering day the construction worker drank 1 liter of water.]	*[The girl drank 1 liter of cola in one go.]*
De visite at een halve taart op.	De man at een halve taart op na het avondeten.
[The guests finished half a cake.]	*[After his dinner the man finished half a cake.]*
Mijn werk is 1 km van mijn huis vandaan.	Mijn auto staat 1 km van mijn huis vandaan.
[My work is 1 km from my house.]	*[My car is parked 1 km from my house.]*
Haar middel heeft een omtrek van 60 cm.	Het hoofd van het kind heeft een omtrek van 60 cm.
[Her waist girth is 60 cm.]	*[The child's head girth is 60 cm.]*
Het TV-programma duurde 10 minuten.	De wasmiddelen-reclame duurde 10 minuten.
[The TV show lasted 10 minutes.]	*[The detergent commercial lasted 10 minutes.]*
De lift had 120 kilo draagkracht.	De man woog 120 kilo.
[The elevator had a capacity of 120 kilos.]	*[The man weighed 120 kilos.]*
Er zaten 80 mensen in het stadion.	Er zaten 80 mensen in de woonkamer.
[80 people were seated in the stadium.]	*[80 people were present in the living room.]*
De gang was 60 cm in de breedte.	De vader had een karper van 60 cm gevangen.
[The hallway was 60 cm wide.]	*[The father caught a 60 cm carp.]*
Met oud en nieuw had hij vier biertjes gedronken.	Hij stapte achter het stuur nadat hij vier biertjes had gedronken.
[On new year's eve he drank 4 beers.]	*[He got behind the wheel after drinking 4 beers.]*
De kippen van de boer hadden 5 eieren gelegd in een maand.	Mijn oma bakte 5 eieren voor mijn opa.
[In a month, the farmer's chickens had laid 5 eggs.]	*[My grandma baked 5 eggs for my granddad.]*
Hij had na een maand 40 km met zijn auto gereden.	Hij fietste elke dag 40 km.
[After a month he had driven 40 km in his car.]	*[He biked 40 km every day.]*
Zijn appartement was 16 vierkante meter.	Het balkon was 16 vierkante meter.
[His apartment was 16 square meters.]	*[The balcony was 16 square meters.]*
In anderhalf uur kan je in Ierland zijn.	Hij was anderhalf uur te laat op ons afspraakje.
[You can get to Ireland in one an a half hours.]	*[He was one and a half hours late on our date.]*
Hij was in een jaar 1 kilo aangekomen.	De bakker had een croissantje gebakken van 1 kilo.
[He had gained 1 kilo in a year.]	*[The baker had baked a croissant that was 1 kilo.]*
Er lagen 5 bananen in de kist bij de groenteboer.	Hij sneed 5 bananen in plakjes in zijn joghurt.
[The crate at the greengrocer had 5 bananas.]	*[He sliced 5 bananas in his yogurt.]*
De kast was 30 cm breed.	De poedel had een bot in zijn bek van 30 cm.
[The closet was 30 cm wide.]	*[The poodle held a 30 cm bone in it's mouth.]*
Hij dronk 4 koppen koffie per week.	Hij dronk 4 koppen koffie per dag.
[He drank 4 cups of coffee a week.]	*[He drank 4 cups of coffee a day.]*
De patatjes zijn 3 cm in de lengte.	De nagels van mijn vriendin zijn 3 cm.
[The chips were 3 cm long.]	*[My friend's nails are 3 cm.]*
Het vliegveld was 5 km van mijn huis vandaan.	De dichtstbijzijnde supermarkt was 5 km van mijn huis.
[The airport was 5 km from my house.]	*[The closest supermarket was 5 km from my house.]*
De kerk heeft een omtrek van 12 meter.	De boom heeft een omtrek van 12 meter.
[The church's circumference is 12 meters.]	*[The tree's circumference is 12 meters.]*
In een kwartier had ze het avondeten gemaakt.	In een kwartier had ze de vissen gevoerd.
[In a quarter of an hour she made dinner.]	*[In a quarter of an hour she fed the fish.]*
De oude vrouw kocht een zak met anderhalf kilo aardappels.	De oude vrouw kocht een zak met anderhalf kilo koekjes.
[The old lady bought a one and a half kilo bag of potatoes.]	*[The old lady bought a one and a half kilo bag of biscuits.]*
Er stonden 5 paar schoenen in de winkel.	Ze nam 5 paar schoenen mee op vakantie.

(Continued)

Sentences used in Experiments 1, 3, and 4. Approximate translations are given in brackets.

Few	Many
[There were 5 pairs of shoes in the shop.]	*[She brought 5 pairs of shoes on her vacation.]*
De file waar we in stonden was 1 km.	De lengte van de tuin was 1 km.
[The traffic jam we were in was 1 km.]	*[The length of the garden was 1 km.]*
De kok deed een lepeltje sambal in de pan met bami.	De jongen deed een lepeltje sambal in zijn kop soep.
[The chef put one teaspoon of sambal in the pot of noodles.]	*[The boy put one teaspoon of sambal in his cup of soup.]*
Na vijf minuten had ze besloten naar welk land ze zou gaan.	Na vijf minuten had ze besloten wat ze wilde drinken.
[After 5 minutes she had decided to which country she would go.]	*[After 5 minutes she had decided what she wanted to drink.]*
De gouden oorbellen kostten 10 euro.	De bloemkool kostte 10 euro.
[The golden earrings cost 10 euro.]	*[The cauliflower cost 10 euro.]*
Hij had 2 maanden in dat huis gewoond.	Een vriend van ons bleef 2 maanden logeren.
[He had lived in the house for 2 months.]	*[Our friend stayed at our home for 2 months.]*
De man werkte op woensdag 2 uur.	De student stond 2 uur op de bus te wachten.
[The man worked two hours on Wednesdays.]	*[The student had been waiting for the bus for two hours.]*
Voor het buurtfeest kocht hij 6 zakken chips.	Voor bij de film kocht hij 6 zakken chips.
[For the neighborhood party he bought 6 bags of crisps.]	*[For during the film he bought 6 bags of crisps.]*
Er zaten 7 mensen in de trein.	Er zaten 7 mensen in de taxi.
[7 people were on the train.]	*[7 people were in the taxi.]*
De man waste zich 2 keer per week.	De man waste zijn auto 2 keer per week.
[The man washed (himself) twice a week.]	*[The man washed his car twice a week.]*
De student spreekt zijn huisgenoot 1 keer in de maand.	De student spreekt zijn studiebegeleider 1 keer in de maand.
[The student talks to his housemate 1 time a month.]	*[The student talks to his school counselor 1 time a month.]*
De casiere van de supermarkt had die dag 35 klanten.	De chirurg had die dag 35 operaties.
[The supermarket cashier had 35 customers that day.]	*[The surgeon had 35 surgeries that day.]*
Mijn broer gaat 1 keer per maand naar de sportschool.	Mijn broer gaat 1 keer per maand naar de tandarts.
[My brother goes to the gym once a month.]	*[My brother goes to the dentist once a month.]*
Er staat 20 milimeter water in de badkuip.	Er staat 20 milimeter water in de huiskamer.
[The bathtub has 20 millimeters of water.]	*[The living room has 20 millimeters of water.]*
De topsporter wandelde 10 km.	Mijn oma liep met haar rollator 10 km.
[The sports star walked 10 km.]	*[My grandma walked 10 km with her rollator.]*
De man parkeerde de auto 1 cm over de witte lijn.	De zuster prikte 1 cm naast de ader.
[The man parked his car 1 cm over the line.]	*[The nurse injected 1 cm off the vein.]*

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 09 February 2011
doi: 10.3389/fpsyg.2011.00015

Manipulating objects and telling words: a study on concrete and abstract words acquisition

Anna M. Borghi[1,2]*, Andrea Flumini[1], Felice Cimatti[3], Davide Marocco[4] and Claudia Scorolli[1]*

[1] *Department of Psychology, University of Bologna, Bologna, Italy*
[2] *Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy*
[3] *Department of Philosophy, University of Calabria, Arcavacata di Rende, Italy*
[4] *School of Computing and Mathematics, University of Plymouth, Plymouth, UK*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Domenico Parisi, The National Research Council, Italy
Manuel De Vega, University of La Laguna, Spain

***Correspondence:**
Anna M. Borghi and Claudia Scorolli, Department of Psychology, University of Bologna, Viale Berti Pichat 5, 40100 Bologna, Italy.
e-mail: annamaria.borghi@unibo.it; claudia.scorolli2@unibo.it

Four experiments (E1–E2–E3–E4) investigated whether different acquisition modalities lead to the emergence of differences typically found between concrete and abstract words, as argued by the words as tools (WAT) proposal. To mimic the acquisition of concrete and abstract concepts, participants either manipulated novel objects or observed groups of objects interacting in novel ways (Training 1). In TEST 1 participants decided whether two elements belonged to the same category. Later they read the category labels (Training 2); labels could be accompanied by an explanation of their meaning. Then participants observed previously seen exemplars and other elements, and were asked which of them could be named with a given label (TEST 2). Across the experiments, it was more difficult to form abstract than concrete categories (TEST 1); even when adding labels, abstract words remained more difficult than concrete words (TEST 2). TEST 3 differed across the experiments. In E1 participants performed a feature production task. Crucially, the associations produced with the novel words reflected the pattern evoked by existing concrete and abstract words, as the first evoked more perceptual properties. In E2–E3–E4, TEST 3 consisted of a color verification task with manual/verbal (keyboard–microphone) responses. Results showed the microphone use to have an advantage over keyboard use for abstract words, especially in the explanation condition. This supports WAT: due to their acquisition modality, concrete words evoke more manual information; abstract words elicit more verbal information. This advantage was not present when linguistic information contrasted with perceptual one. Implications for theories and computational models of language grounding are discussed.

Keywords: categorization, concepts, embodied cognition, grounded cognition, language grounding, language acquisition

INTRODUCTION

How do children acquire abstract words? This paper presents a study on novel categories focusing on what differs in the acquisition of concrete and abstract words. One standard way of differentiating between concrete and abstract words is to refer to their perceivability. Concrete words refer to entities that can be perceived through the senses. Abstract words refer to entities more detached from physical experience (Paivio et al., 1986; Crystal, 1995; Barsalou et al., 2003). However, the distinction between concrete and abstract words cannot be conceived of as a dichotomy (Wiemer-Hastings et al., 2001). For example, words referring to social roles (e.g., "physician") might be more abstract than words referring to single objects (e.g., "bottle"), but less abstract than purely definitional words (e.g., "odd number") (Keil, 1989). In addition, words referring to emotions probably require special classification (Altarriba et al., 1999). Further, basic and subordinate words, such as "cat" and "Siamese cat," referring to single entities, can be seen as more concrete than superordinate words, such as "animal," that refer to sets of entities that differ in shape and other perceptual characteristics (e.g., Borghi et al., 2005). To summarize, the distinction between concrete and abstract words is not clear-cut, and should be intended as a continuum. However, we believe that this distinction captures some aspects of word meaning, and that it is important to understand how the process of abstraction occurs, from single instances to categories at different levels of abstraction. In particular, explaining the ways in which abstract words are represented constitutes a major challenge for embodied and grounded views of cognition, as well as for embodied computational models and robotics. The problem abstract words pose for embodied and grounded theories is clearly synthesized by Barsalou (2008, p. 634) as follows: "Abstract concepts pose a classic challenge for grounded cognition. How can theories that focus on modal simulations explain concepts that do not appear modal?" We will first clarify why explaining abstract concepts is a crucial challenge for embodied cognition, and later clarify its importance for research in robotics.

According to the standard propositional view (e.g., Fodor, 1998), the representation of both concrete and abstract concepts is abstract, symbolic, and amodal. In contrast, according to standard embodied accounts (e.g., Barsalou, 1999) both concrete and abstract concepts are grounded in perception and action systems, and therefore are modal. Notice that both standard propositional and embodied accounts evoke a single kind of representation, either amodal or modal, for both concrete and abstract concepts.

In contrast, recent views propose that multiple representational systems are activated during conceptual processing (e.g., Louwerse and Jeuniaux, 2010; for a non-embodied version of this view, see Dove, 2009). According to these views both sensorimotor and linguistic information play a role in conceptual representation. This idea is not entirely novel. The seminal dual coding theory by Paivio et al. (1986) applies two different kinds of representations, a linguistic and a sensorimotor code, to explain how concrete and abstract words are represented and recalled. Concrete words are recalled more easily because they activate both sensorimotor and linguistic information; differently abstract words are not "grounded," they only evoke linguistic information. Recent support to Paivio's theory comes from studies on brain imaging showing that abstract word processing is strongly lateralized toward the left hemisphere, while activation during processing concrete words is bilateral (for a review, see Sabsevitz et al., 2005). However, this might be due to the fact that the majority of the studies employ single words and tasks requiring a superficial level of processing. Recent studies requiring deeper processing, such as sentence sensibility evaluation tasks, do not provide evidence in favor of a pronounced laterality (e.g., Desai et al., 2010). The major difference between Paivio's view and the embodied accounts we will refer to is based on the concept of multiple representation; to elaborate, Paivio argues that abstract words are not "grounded" in perception and action systems, whereas according to the embodied perspective both concrete and abstract words activate both linguistic and perception–action information, even if these two kinds of information are differently distributed.

The language and situated simulation (LASS) theory is probably the most well-known of the multiple representation theories (Barsalou et al., 2008). In this view both the linguistic and the sensorimotor system are activated during word processing. The understanding of word meanings always implies activation of the sensorimotor system (simulation), but for tasks which do not require deep processing the linguistic system might suffice. While presenting the LASS theory, Barsalou et al. (2008) suggest that for abstract concepts, linguistic information might be more relevant than for concrete concepts, but they do not advance clear predictions pertaining the differences in processing between concrete and abstract concepts, independently from the task. Thus, they argue that "different mixtures of the language and simulation systems support the processing of abstract concepts under different task conditions." (Barsalou et al., 2008, p. 267).

More precise predictions concerning the difference between concrete and abstract words are advanced by the words as tools (WAT) proposal (Borghi and Cimatti, 2009, 2010), which assumes the existence of multiple representations. WAT is based on the idea, initially proposed by Wittgenstein (1953), that words are *tools* we use (see also Clark, 1998). Similarly to real tools, words can be considered as instruments to act in the social world, thus as social tools. The difference between concrete and abstract words is explained by WAT referring to the fact that, due to a different acquisition process, the role played by actions performed through words – by linguistic information – is more relevant for abstract than for concrete words. The present work aims to directly test the WAT proposal using novel categories and novel linguistic labels. According to WAT perception and action are crucial in the acquisition of concrete words. Instead to acquire the meaning of an abstract word children also rely on verbal explanations (for example, explaining the meaning of "democracy" requires many more other words than for explaining the meaning of "bread"). In this respect, the role played by words as social tools is more important for abstract than for concrete words. Evidence relevant to this issue was obtained by Wauters et al. (2003), who studied different modalities of acquisition (MOA) of words. They did not however, speak directly about concrete vs. abstract words. According to the authors, the meaning of a word like "ball" is acquired through perception, because every time the child hears the word, he/she sees a real ball, or a picture of it. The meaning of a word like "grammar," instead, has to be explained linguistically. Finally, the meaning of a word like "tundra" can be acquired in both ways, depending on the environment where it is learned. WAT predicts that this difference in the acquisition process can explain why, for concrete and abstract words both perception–action and linguistic information are activated. Linguistic and social information however, plays a more important role for abstract than for concrete words (e.g., Crutch and Warrington, 2005; Sabsevitz et al., 2005).

From a different perspective, an embodied and grounded account of the difference between concrete and abstract words is crucial in the process of developing intelligent machines capable of autonomously creating categories and using language. In computational cognitive science, robotics offers new opportunities for the design of artificial agents in which language is grounded on their ability to manipulate and experience the external world by means of physical interactions. The *symbol grounding problem* (Harnad, 1990) highlights the fact that, in traditional computational models, symbols are self-referential entities that require the interpretation of an external experimenter to identify the referential meaning of the lexical items. This issue has been widely discussed in the realm of cognitive science, and robotics offers a completely different way to solve the grounding problem. Indeed, in the last 20 years, many different models were created with the explicit aim of grounding symbols and language in perception (e.g., Steels, 2003) and, more recently, in action (Sugita and Tani, 2005; Marocco et al., 2010). Although the embodied approach to language in robotics is gaining increased interest, both in terms of cognitive modeling and applications, the current trend is strongly focused on systems capable of autonomously acquiring concrete concepts and words, that can be grounded on perception and action processes of the robot. Existing models do not focus on the acquisition of abstract words, except for highlighting that such abstract concepts and words permeate the entire domain of human language experience and cannot be neglected. Nevertheless, an extension of the actual grounding approach in robotics to abstract words is not automatic. In this regard, we believe that the WAT proposal offers an interesting way to incorporate abstract words in future cognitive robotic models without compromising the grounding and the embodied approach, which should be the milestone of the future robotics. On the other hand, a robotic model could be useful to complement traditional psychological experiments, and provide further evidence on the feasibility of a novel theory, such as the WAT proposal presented.

In this research we used novel categories to mimic the different ways in which concrete and abstract word meanings are acquired and then represented. Reported experiments are designed in a way that allows for replication with a computational model. Similar stimuli and training processes can be used to create a cognitive

based controller for a humanoid robot (Tikhanoff et al., 2008) that will be able to perform an identical categorization task. We defined concrete concepts as having a concrete, manipulable object as a referent. Abstract concepts, on the other hand, do not have a single, manipulable object as referent; instead they refer to rather complex relations between entities. We acknowledge that the distinction we made for operational simplicity is not exhaustive and that it covers only a subset of items. For example, it leaves out word meanings referring to perceivable but not manipulable objects or entities, such as "cloud," "mountain," and "moon." Even if the referents of these words cannot be manipulated, we would consider them as concrete, as their referents are clearly perceivable, can be scanned (acted upon) with the eyes, and are easy to imagine. We decided to address the distinction between concrete and abstract words starting from the extremes of the continuum: for this reason we decided to focus on concrete, manipulable objects. As for abstract word meanings, here we did not refer to purely definitional abstract word meanings, simply based on verbal explanations (as it might be the case for a word like "philosophy") but to word meanings that evoke complex relationships between entities; due to their complexity, we suspect applying a linguistic label and explaining their meaning is crucial in order to form categories. Consider that the referents of our abstract categories were interacting moving objects – thus they were perceivable, similarly to the referents of concrete categories. As a matter of fact, in our view the formation of abstract categories always starts with some form of perception, be it visual, acoustic, tactile, or otherwise.

Due to the difficulties involved in reproducing the acquisition of different kinds – concrete vs. abstract – of novel concepts/words in an artificial setting (i.e., laboratory), we operationalized the acquisition process considering two phases – the experience and the word acquisition – as follows:

a – Novel concepts acquisition: Training 1 (Experience) was designed to mimic the acquisition of concrete and abstract concepts. The idea underlying these two different acquisition processes is that, where typically concrete concepts refer to category members which are perceptually similar or elicit similar actions, abstract concepts refer to entities that show complex interactions, or do not share an evident perceptual similarity (i.e., common features are not perceptually salient). We showed participants 3D figures of novel objects vs. 3D figures of objects interacting in novel ways. Then participants were tested (TEST 1: Categorical Recognition).

b – Novel labels acquisition: during Training 2 (Words Acquisition) participants were taught the category name; in some conditions a verbal explanation of the category meaning was added. Then participants were tested (TEST 2: Words–Objects Match). We predicted that in both tests participants would produce less errors with concrete than with abstract categories, as the first can be formed more easily on perceptual and motor basis. This difference should be reduced when a category label and a linguistic explanation of what the category members had in common were given.

The manipulation of TEST 3 in the different experiments allowed us to check for the effectiveness of our operationalization of acquisition process (Experiment 1), as well as to test if the verbal labeling, possibly strengthened by a verbal explanation, reinforces learning of both concrete and abstract categories in different ways (Experiments 2, 3, and 4).

c – Real words evidence match: TEST 3 of Experiment 1 consisted of a feature production task. We predicted that the pattern of produced properties would match that typically obtained in feature generation tasks with concrete and abstract words.

d – Linguistic vs. Manual Information: in Experiments 2, 3, and 4, TEST 3 consisted of a property verification task. We chose to ask participants to respond to the objects' color because color was not relevant to the motor response. In one condition participants were required to provide a manual response (i.e., to press a key on the keyboard), and in another a verbal response (i.e., to respond "yes" with the microphone; see Scorolli and Borghi, 2007). We predicted facilitation for manual responses with concrete words and for mouth responses with abstract words. This would demonstrate that language is part of the representation of abstract words meanings. The rationale is the following: if linguistic information is more relevant for the representation of the meaning of abstract compared to concrete words, with abstract words phono-articulatory aspects should be accessed more easily compared to sensorimotor manual ones. Therefore, a linguistic response (even a simple "yes" response) should be facilitated compared to a manual one.

EXPERIMENT 1

The experiment was designed to mimic the acquisition of concrete and abstract categories and to verify whether the novel categories we used reproduced the acquisition process that occurs with real world categories. As anticipated, in Experiment 1, TEST 3 consisted of a production task. Before starting the experiment, participants were randomly assigned to two groups. One group was first shown the category and then tested on concrete items; later participants were shown and then tested on abstract items; the other group first learned and then was tested on the two kinds of items in reverse order. Across the experiments the order of presentation of the two blocks (concrete block; abstract block) was counterbalanced. The same methodological choice was applied to all the other three experiments.

METHOD

Participants

Sixteen students of the University of Bologna took part in the study (three men; mean age = 20.31 years; SD = 1.62). All were native Italian speakers, both right- and left-handed (two left-handed) and all had normal or corrected-to-normal vision. The study was approved by the local ethics committee.

Materials

3D figures of novel objects and related new labels. We invented four novel words (*calona*, *fusapo*, *norolo*, *tocesa*) all having the same number of syllables and letters. We avoided using new words with ambiguous accents. Two of the four words ended with the vocal "a," which in Italian characterizes the female gender; the remaining two words ended with the vocal "o," which in Italian characterizes the male gender. The new words corresponded to four new categories of objects, composed of 12 exemplars each (4 × 12). The criteria we followed to construct the "original" three new objects were the following:

1. CALONA was a 3D concave figure ("C" shaped). The colors we used were sky-blue and light-gray;
2. FUSAPO was a 3D figures with five protuberances ("*"shaped). The colors we used were blue and yellow (**Figure 1**);
3. NOROLO was a 3D figure with small convex nooks ("N" shaped). The colors we used were red and gray;
4. TOCESA was a 3D figure shaped as wavy slash, without internal convexities or concavities ("I" shaped). The colors we used were violet and beige.

The other nine exemplars for each category were both built by inverting the surface and depth colors (3 × 2), and by rotating the original figures by 180° (6 × 2). Finally, we built 40 3D figures that were used as fillers: they did not belong to a category and were not assigned a name.

3D figures of novel relations and related new labels. We invented four new words (cofiro, latofo, panifa, rodela) by following the same criteria as described for the linguistic labels used for the 3D figures of novel objects. These new words referred to new categories of relations between two 3D figures; each of these categories was composed by 12 exemplars (4 × 12). We used the following criteria to construct the "original" three new relations (that is, novel groups of 3D interacting objects):

a. COFIRO: two 3D moving figures. After the contact just one 3D figure remained, and it moved in a straight line or in a curved line;
b. LATOFO: one 3D static figure and two 3D moving figures. After the contact two 3D figures appeared at the opposite diagonal sides of the computer screen (e.g., one at the top right of the screen and the other at the bottom left of the screen), and they moved converging toward the central point of the screen;
c. PANIFA: two 3D moving figures. After the contact one of them moved in a straight line; the other one executed a turning movement with a different velocity (**Figure 2**);
d. RODELA: one 3D static figure and two 3D moving figures. After the contact the two 3D figures moved in a same (straight) line and with the same velocity, but in an opposite direction, as if the figures were pushed away from each other.

All the 3D figures were sky-blue cylinders; they were arranged horizontally, one came from one part of the screen and the other from the other side. For LATOFO and RODELA we added a 3D static figure to the two interacting ones. This aimed to reproduce real-life abstract word acquisition: some abstract words can evoke both relations between entities and static visual images (e.g., "freedom" can evoke a bird flying in the sky as well as an image of the Statue of Liberty). In other words, it can happen that objects which would be first categorized as exemplars of a concrete category (e.g., a statue) can be re-categorized and evoked by abstract words.

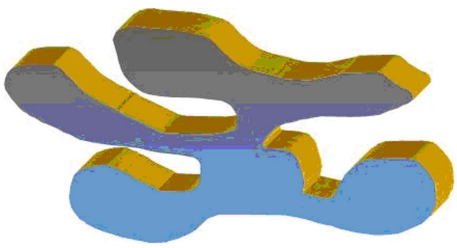


FIGURE 1 | An exemplar of the concrete category FUSAPO; all other category members were perceptually similar to the shown exemplar.

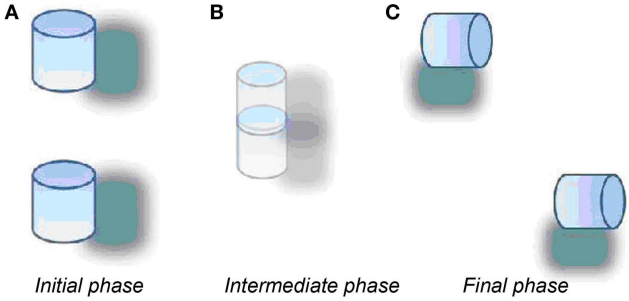


FIGURE 2 | An exemplar of the abstract category PANIFA; the figure shows three phases – initial (A), intermediate **(B)** and final **(C)** – of the interacting movement. All the other category members were not perceptually similar, but showed similar complex interactions.

The other nine exemplars for each category were built by using parallelepipeds (3 × 2) instead of cylinders; the movement of the 3D figures followed a vertical instead of a horizontal direction (6 × 2). Finally, we built 40 3D figures to use as fillers, and we constructed 40 relations between 3D figures to use as fillers. They did not belong to a category and were not assigned a name. The duration of each relation was the same for both the categories' exemplars and the fillers (4 s).

Procedure

Across all experiments, participants were trained and tested individually in a quiet laboratory room. They sat on a comfortable chair in front of a computer screen. All participants were submitted to two training phases (Experience; Word Acquisition) and to three different tests (Categorical Recognition; Word–Object Match; Production).

Training 1: experience. Training 1 aimed to reproduce the different processes underlying the acquisition of concrete and abstract concepts. Whereas typically concrete words refer to category members which are perceptually similar or elicit similar actions, abstract words refer to entities that show complex interactions or do not share an evident perceptual similarity (i.e., common features are not perceptually salient). For example, the word "truth" binds experiences and situations that might be rather complex and different. During this training session participants were sitting in front of the computer screen. They were exposed to 20 trials. In each trial either three 3D figures (in the concrete concept acquisition condition) or three relations between 3D figures (in the abstract concept acquisition condition), were shown. Both the 3D figures and the relations were novel, i.e., participants had never experienced them before. In order to mimic the acquisition of concrete concepts (e.g.,

BOTTLE), participants were presented with 3D figures of novel objects as previously described. They were instructed to verify whether the objects could be inserted inside a donut shaped 3D figure. The experimenter invited them to manipulate the objects with the mouse for 12 s each. In order to simulate the acquisition of abstract concepts (e.g., TRUTH), participants were instructed to observe the groups of dynamic objects until the end of their interaction (12 s). The 3D figures interacted in ways that revealed the existence of a common structure. For example, two objects moved toward each other, then only one of them remained on the screen, moving in a straight line (COFIRO).

TEST 1: Categorical Recognition. Training 1 was followed by a categorical recognition task (TEST 1). Participants were instructed to look at a fixation cross that remained on the screen for 500 ms. Then they were shown two exemplars of the same or different categories, and were asked to judge whether the stimuli belonged to the same category or not by pressing two different keys (left, right). The key–response mapping was counterbalanced. They were shown 24 randomly ordered trials, with different combinations of the exemplars or of the exemplars and fillers, that is:

1. two exemplars of the same category;
2. two exemplars belonging to two different categories;
3. one exemplar of a category and one filler, that did not belong to any learned category.

Concrete concepts' exemplars remained on the screen for 2 s, while abstract concepts' exemplars were displayed for 10 s. The 24 experimental trials were preceded by two training trials.

The Categorical Recognition task aimed to verify whether the training phase allowed participants to form a category on a purely sensorimotor basis, and to contrast it with a different category. We collected and analyzed errors, as this is the more reliable and informative measure for this particular task. Across all studies, percentages of errors are reported. We predicted that participants would produce less errors with concrete than with abstract categories, as the first can be formed more easily without the aid of language.

Training 2: Words Acquisition. After TEST 1, participants were trained to associate a linguistic label to each learned exemplar. Five exemplars from each category were randomly selected and they were presented once to participants together with the appropriate linguistic label. In order to mimic the acquisition of concrete words participants were shown 20 3D figures together with the related linguistic labels ("calona," "fusapo," "norolo," "tocesa"), presented in random order. Each trial lasted 2 s. Symmetrically, in order to simulate the acquisition of abstract words, participants observed the 20 relations together with the related linguistic labels ("cofiro," "latofo," "panifa," "rodela"), presented in random order. Each trial lasted 4 s. Participants were instructed to learn the linguistic labels associated with the 3D figures and with the relations.

TEST 2: Words–Objects Match. After the Training 2 participants had to perform a Words–Objects Match task. They were presented with 24 trials. One of the learned names and two figures/relations were displayed on the computer screen: the target object, corresponding to the label, and another nearby, which in half of the trials was novel and in the remaining 12 trials was an exemplar already associated with a different label. One of the two figures/relations was located on the left of the screen, the other on the right; the figure location was counterbalanced. Participants were required to decide by pressing a different key (left, right) on the keyboard which of the two was named with the shown label. This second test aimed to verify whether participants had associated a label with a category, and whether they were able to generalize it to a different category. We predicted that participants would produce fewer errors with concrete than with abstract categories, as the first rely more than the second on perception and action. However, the difference between concrete and abstract categories should be reduced compared to TEST 1, given that participants could now rely on linguistic labels as well.

TEST 3: production task. After TEST 2, TEST 3 consisted of a feature production task with novel category names. The experimenter told participants each category name (in four random orders) asking them to produce the first properties that came to their mind. They were prompted to produce properties until they stopped for about 15 s. Properties produced were transcribed; both their frequency and production order was recorded. We predicted that the pattern of produced properties would match that typically obtained in production tasks with concrete and abstract words. Behavioral studies with production tasks, such as word association and property generation tasks, have shown that, whereas concrete words activate mainly perceptual and thematic relations, abstract words typically elicit more taxonomic relations (Borghi and Caramelli, 2001); in addition, they elicit more situations and introspective relations compared with concrete words (Barsalou and Wiemer-Hastings, 2005).

RESULTS

Across all experiments, significant results will be reported.

TEST 1: Categorical Recognition

We performed a one-way ANOVA on errors produced in the categorical recognition task, in which the factor Concept (Concrete vs. Abstract) was manipulated within participants. As predicted, Abstract Concepts ($M = 5.21\%$) elicited more errors than Concrete Ones ($M = 2.34\%$), $F(1,15) = 12.70$, MSE = 5.17, $p < 0.005$ (see **Table 1**).

TEST 2: Words–Objects Match

An ANOVA was performed on the errors produced. Consider that two objects were presented on the screen, the target one and another object. Therefore in the ANOVA two factors were entered, both manipulated within participants: the factor Word (Concrete vs. Abstract) and the factor Other Exemplar (Novel vs. Learned). Both factors reached significance; Abstract Words ($M = 5.01\%$) elicited more errors than Concrete Ones ($M = 1.37\%$), $F(1,15) = 11.96$, MSE = 17.79, $p < 0.005$, and more errors were produced when the target exemplar was presented with a Learned ($M = 4.17\%$) than with a Novel Other Exemplar ($M = 2.21\%$), $F(1,15) = 15.70$, MSE = 3.89, $p < 0.005$ (see **Table 1**).

Table 1 | Errors percentages and standard deviations (in parenthesis) for TEST 1 and TEST 2 of Experiment 1. For TEST 3 we reported results on ratings' scores.

EXPERIMENT 1		
CATEGORICAL RECOGNITION		
Concept	Concrete	Abstract
	2.34 (2.62)	5.21 (2.95)
WORDS–OBJECTS MATCH		
Word	Concrete	Abstract
	1.37 (1.64)	5.01 (4.90)
Other Exemplar	Novel Exemplar	Learnt Exemplar
	2.21 (3.66)	4.17 (4.27)
PRODUCTION		
Scaled ratings (1 concrete referent → 7 abstract referent) applied to the individual protocols		
Word	Concrete	Abstract
	3.04 (0.29)	4.14 (0.29)
Normalized position in which each property was produced $(n+1-p)/(n-1)\times r$		
Word	Concrete	Abstract
	3.11 (0.25)	4.48 (0.46)

TEST 3: production task

Different analyses were performed on the production task. The number of produced properties did not differ significantly between Concrete ($M = 4.18$) and Abstract Words ($M = 3.73$); $p = 0.29$. The properties produced with each word were put together, organized in two different random orders, and 12 participants were asked to rate the produced properties on a 7-point scale. They were required to select 1 if they believed that the property was typical of words having "concrete" referents, such as bottles, screwdriver, building, cellular, and cat, and 7 if they thought the property was typical of words having "abstract" referents, such as happiness, philosophy, risk, fantasy, democracy. The raters did not know which situation the properties had been produced in. We performed an ANOVA on the ratings of the properties produced with concrete and abstract words. As predicted, we found that abstract words elicited significantly higher scores than concrete words ($M = 3.93$; $M = 3.13$), $F(1,11) = 27.51$, MSE = 0.14, $p < 0.001$. In addition, the scaled ratings were applied to the individual protocols in order to verify whether the properties produced and the production order of the properties for each word reflected the properties typically produced for concrete or for abstract words (the same method was used by Borghi and Barsalou, 2001; Borghi, 2004; Wu and Barsalou, 2009). The average rating of each property was multiplied by the frequency of the produced property for each of the participants. A one-way ANOVA was performed on the obtained mean values, with participants as the random factor. The only factor manipulated was significant, $F(1,11) = 27.51$, MSE = 0.14, $p < 0.001$, as the mean values obtained with Abstract Words ($M = 4.14$) were higher than those produced with Concrete Words ($M = 3.04$), indicating that the novel Abstract Words we created elicited properties typical of real-life abstract words (e.g., "singularity"; "variation"; "linear motion"); this was symmetrically true for the novel Concrete Words which elicited a higher number of properties such as "hole in the middle," "stick-shaped," "crab-shaped." In addition, the average rating on each property was multiplied by the position of the property produced for each participant according to the formula $(n+1-p)/(n-1)\times r$, where n is the total number of properties produced by each participant for each word, p the position in which each property was produced, and r the average rating on that particular property (for a similar procedure, see Wu and Barsalou, 2009). This normalized p is the position in which each property was produced, in relation to n, the total number of properties produced by each participant. One ANOVA was performed on the obtained mean values, with participants as random factor; the factor manipulated was the kind of Word (Abstract vs. Concrete Words). The ANOVA again revealed lower mean values for Concrete ($M = 3.11$) than for Abstract Words ($M = 4.48$), $F(1,15) = 55.38$, MSE = 0.27, $p < 0.00001$. This indicates that with our novel Concrete Words properties typically elicited by real concrete words were elicited earlier, and the same was symmetrically true for our novel Abstract Words (see **Table 1**).

DISCUSSION

Results of Experiment 1 indicate that with our training with novel categories and words we were able to recreate the real-life situation in which concrete and abstract words are learned.

Results for TEST 1 (categorical recognition) indicated that it is more difficult to form abstract categories than concrete ones. In addition, results of TEST 3 (property generation task) showed that the properties produced for the concrete and abstract words we created corresponded to those typically obtained with existing concrete and abstract words. Results of TEST 1 and TEST 3 revealed that abstract categories are more difficult to form, and that abstract words are represented differently from concrete ones, as they elicit less perceptual properties, such as properties related to shape, and more abstract and relational properties.

The higher difficulty of abstract words compared to concrete ones was also maintained in TEST 2 (Words–Objects Match), when participants learned to associate a novel word to a category. Results on TEST 2 showed that the use of linguistic labels did not further facilitate the acquisition of abstract in comparison to concrete words. This reveals that the higher complexity of abstract concepts is not reduced thanks to the use of linguistic labels. A possibility is that, in order to reduce the complexity of abstract words, a verbal explanation of the category meaning is needed.

EXPERIMENT 2

Given our results on Words–Objects Match (TEST 2) in Experiment 1, in Experiment 2 we decided to add a verbal explanation to the linguistic label used for abstract categories. This should mirror the way the acquisition process works. Abstract words differ from concrete words insofar that the first refer to a variety of situations, states, events. Due to this complexity, linguistic labels should be more relevant for abstract than for

concrete words acquisition, and the first might also require a verbal explanation of their meaning. This is often not the case for concrete words, for which the linguistic label is usually associated with the presence of the object. Experiment 2 aimed to test whether there is a facilitation effect when the meaning of abstract words is explained linguistically, compared to when only the linguistic label is provided.

In addition, the aim of Experiment 2 is to verify whether the different acquisition modality has an impact on the response modality. We designed a property verification task (TEST 3), to be performed in substitution of the production task of Experiment 1 in order to address this aim. We chose to use color as the target property as color was not relevant to the motor response and to the response device that we used.

Specifically, we predicted that, given that for concrete words manual information is more relevant than for abstract ones, participants should be faster to perform a property verification task with concrete words when they had to respond using a keyboard instead of a microphone. Symmetrically, if it is true that linguistic information is more important for the acquisition of abstract word meanings than for concrete ones, faster responses should be noted with regard to abstract words while responding with the microphone than with the keyboard. We expect a stronger effect when abstract words are presented not only with novel verbal labels but with the explanations as well.

METHOD

Participants

Thirty-two students of the University of Bologna took part in the study (eight men; mean age = 20.44 years; SD = 1.41). All were native Italian speakers and right-handed.

Procedure

All participants were submitted to two training phases (Experience; Word Acquisition) and to three different tests (Categorical Recognition; Word–Object Match; Property verification task). Training 1 and TEST 1 were identical to Experiment 1. However, Training 2 varied, as participants were randomly assigned to two different conditions, the Explanation or No Explanation condition. In the Explanation condition with abstract words half of the participants were told the name of the abstract category and were given an explanation clarifying the similarities of the members of a given category; in the No Explanation condition only the name was associated to the category. Training 2 for concrete categories was the same of Experiment 1.

In TEST 3 participants took part in a color verification task. Questions appeared on the screen, for example, "Is LATOFO yellow?" To respond "yes" or "no" they had to press two keys on the keyboard in one block (24 trials), or to pronounce the word "yes" or "no" in the microphone in another block (24 trials). The block order was counterbalanced. Both response times and errors were recorded. Forty-eight responses were recorded; "yes" responses corresponded to questions on five different colors (blue, red, violet, yellow for concrete words and sky-blue for abstract), and "no" responses corresponded to questions about five wrong colors (black, brown, green, orange, white).

RESULTS

TEST 1: Categorical Recognition

In the one-way ANOVA conducted on error rates the factor Concept (Concrete vs. Abstract), which was manipulated within participants, was highly significant. As predicted and as in Experiment 1, Abstract Concepts ($M = 6.18\%$) elicited more errors than Concrete Ones ($M = 1.82\%$), $F(1,31) = 51.32$, MSE = 5.92, $p < 0.0000001$ (see **Table 2**).

TEST 2: Words–Objects Match

We performed two different ANOVAs on the errors produced, one for the No Explanation group (A) and another for the Explanation group (B). In the first ANOVA two factors were manipulated within participants, Word (Abstract vs. Concrete, both without explanation) and Other Exemplar (Novel vs. Learned). In the second ANOVA the same factors were manipulated but, as far as the Word factor is concerned, we contrasted Abstract Words *with* Explanation vs. Concrete Words *without* Explanation. In the first ANOVA, Abstract Words ($M = 4.04\%$) elicited more errors than Concrete Ones ($M = 1.17\%$), $F(1, 15) = 12.01$, MSE = 10.93, $p < 0.005$, and more errors were produced when the target exemplar was associated with a Learned ($M = 3.52\%$) than with a Novel Other Exemplar ($M = 1.69\%$), $F(1,15) = 13.35$, MSE = 3.98, $p < 0.005$ (see **Table 2**). In addition, the interaction between Word and Other Exemplar was significant, $F(1,15) = 5.46$, MSE = 3.19, $p < 0.04$. *Post hoc* LSD showed that all differences were significant ($p < 0.05$), with the exception of the difference between Concrete Words accompanied with a Learned vs. Novel Exemplar. With Abstract Words, instead, a Target Exemplar presented together with a Learned Exemplar elicited more errors than a Target Exemplar associated with a Novel Exemplar ($p < 0.0005$). In the second ANOVA both main effects were significant: Abstract Words with Explanation ($M = 3.19\%$) elicited more errors than Concrete Words without explanation ($M = 0.98\%$), $F(1, 15) = 6.09$, MSE = 12.87, $p < 0.05$, and more errors were produced when the target exemplar was associated with a Learned ($M = 2.67\%$) than with a Novel Other Exemplar ($M = 1.50\%$), $F(1,15) = 6.09$, MSE = 12.87, $p < 0.05$ (see **Table 2**).

TEST 3: property verification task with keyboard vs. microphone

In TEST 3 we collected both RTs and errors, for a number of reasons. First, previous work on the influence of action sentences on keyboard and microphone response devices was performed recording response times (e.g., Scorolli and Borghi, 2007). Second, differently from TEST 1 and TEST 2, no figures were presented, and participants had to read and respond to verbal questions. Thus there were no differences in the presentation timing of concrete categories (static figures) and abstract ones (videos). We will report results based on LSD test ($p < 0.05$) and discuss the results crucial for our hypotheses. Even though we collected RTs as well, we believe that, given that we study word acquisition, accuracy probably represents the most important measure of participants' performance.

About 24.77% of the trials were removed as errors. RTs above or below two standard deviations from each participant's mean values for correct trials were excluded from this analysis. This trimming method leads to the removal of further 3.39% of the data. The mean RTs for correct responses for true trials for each

Table 2 | Errors percentages and standard deviations (in parenthesis) for each TEST of Experiment 2.

EXPERIMENT 2			
CATEGORICAL RECOGNITION			
Concept		Concrete	Abstract
		1.82 (2.17)	6.18 (3.45)
WORDS–OBJECTS MATCH			
Group A	Word	Concrete without explanation	Abstract without explanation
		1.17 (1.75)	4.04 (3.99)
	Other Exemplar	Novel Exemplar	Learned Exemplar
		1.69 (2.62)	3.52 (3.83)
Group B	Word	Concrete with explanation	Abstract with explanation
		0.98 (1.75)	3.19 (3.55)
	Other Exemplar	Novel Exemplar	Learned Exemplar
		1.05 (2.44)	2.67 (3.40)
PROPERTY VERIFICATION TASK: KEYBOARD VS. MICROPHONE			
Group A	Word	Concrete without explanation	Abstract without explanation
		15.69 (8.47)	10.55 (8.52)
Group B	Word	Concrete without explanation	Abstract with explanation
		15.04 (5.77)	7.75 (5.63)

participant were submitted to two ANOVAs, one for the No Explanation group (A) and another for the Explanation group (B). In the first ANOVA two factors were manipulated within participants: Word (Abstract vs. Concrete, both without explanation) and Response Device (Keyboard vs. Microphone). In the second ANOVA we manipulated the same factors but, with the factor Word, we contrasted Abstract Words *with* Explanation vs. Concrete Words *without* Explanation. In both ANOVAs the factor Word was significant. Abstract Words (M = 958 and 950 ms, respectively) were responded to significantly faster than Concrete ones (M = 1192 ms; M = 1200 ms), $F(1,15) = 12.52$, MSE = 69871.63, $p < 0.005$; $F(1,15) = 57.04$, MSE = 17525.17, $p < 0.0001$ (see **Table 2**). Crucially in the second ANOVA we found an interaction between the kind of Words and the kind of Device, $F(1,15) = 11.18$, MSE = 91173.10, $p < 0.005$: Concrete Words were responded to significantly faster with the keyboard (M = 1057 ms) than with the microphone (M = 1343 ms; LSD *post hoc*, $p < 0.05$); symmetrically Abstract Words were responded to faster with the microphone (M = 841 ms) than with the keyboard (M = 1059 ms; LSD *post hoc*, $p < 0.06$; see **Figure 3**).

The main effect of Word on both the analyses is of marginal interest, as it is probably due to the fact that the task was easier to perform when using Abstract Words, as the figures/entities referred to through abstract words were always light blue colored, whereas objects referred to by concrete words differed in colors. Much more crucial for our hypotheses is the interaction between Word and Response Device found in the second ANOVA (group B): as predicted, with Abstract Words provided by a verbal Explanation RTs were faster with the microphone than with the keyboard; symmetrically with Concrete Words RTs were slower with the microphone than with the keyboard (see **Figure 3**). Finally it is interesting to notice the difference between Abstract and Concrete Words, still present *without* the Explanation (group A, 234 ms), was increased by the introduction of the verbal Explanation (group B, 250 ms), particularly in case of mouth responses.

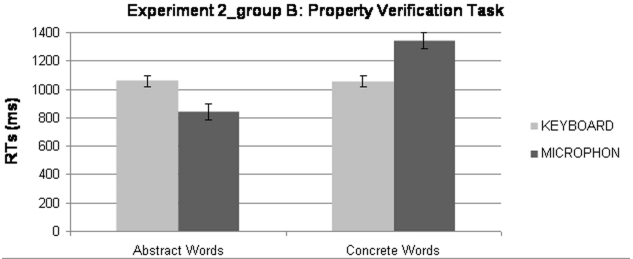


FIGURE 3 | Experiment 2, group B: interaction between Words (Abstract with Explanation, Concrete) and Response Device (Keyboard, Microphone).

Two further ANOVAs on errors were performed, in which the same factors were manipulated. In both analyses the factor Word reached significance: Concrete Words (group A: M = 15.69%; group B: M = 15.04%) elicited more errors than Abstract Words (group A: M = 10.55%, $F(1,15) = 4.49$, MSE = 94.38, $p < 0.05$; group B: M = 7.75%, $F(1,15) = 26.04$, MSE = 32.69, $p < 0.0005$), probably due to the different difficulty level involved in processing the color property. Crucially, the introduction of the explanation strongly reduced errors with Abstract Words (10.55 vs. 7.75%; see **Table 2**).

DISCUSSION

Results of Experiment 2 confirmed and extended those obtained in Experiment 1. Results on the recognition test confirm the results of Experiment 1, indicating that it is more difficult to form abstract categories than concrete ones. As in Experiment 1, TEST 2 showed that when participants learned to associate a novel word with a category, abstract words caused more difficulty in comparison to concrete words. Interestingly, abstract words without Explanation (group A) produced a significantly higher frequency of errors when the exemplar nearby has already been learned: this suggests that the categorical boundaries are less marked with exemplars

referred to by abstract rather than by concrete nouns. By adding an Explanation to the label (group B), the categorical boundaries with exemplars referred to by abstract nouns become marked as the ones referred to by concrete nouns.

More crucial to our hypotheses are the results of TEST 3. As predicted, we found that Abstract Words produced faster responses with the microphone than with the keyboard; by introducing the Explanation (group B) this difference becomes significant. Symmetrically, Concrete Words (group B) were responded to more quickly with the keyboard than with the microphone. This clearly supports the WAT proposal, as it suggests that concrete words evoke more manual information, whereas abstract words elicit more verbal information.

EXPERIMENT 3

A potential problem of Experiment 2 was that TEST 3 (the property verification task) was submitted separately for concrete and abstract words. It is possible that, because abstract words always referred to blue objects, participants did not have to retrieve the perceptual properties of the single categories to respond, whereas this was necessary for concrete words. This could explain why RTs were faster with abstract than with concrete words. Experiment 3 is very similar to Experiment 2, with some modifications. First, given the interesting results obtained with explanations, we decided to use only the explanation condition with abstract words. Second, we balanced color information of objects referred to by both concrete and abstract categories, coloring the abstract figures. We used both concrete and abstract figures of different colors. We introduced this variation in order to solve the potential limitations of Experiment 2, thus to avoid any facilitation with abstract words in responding to the property verification task due to the fact that all abstract words' referents were blue in color. Third, in order to precisely control for the influence of learning the new labels of categorization we decided to perform the category recognition task both before and after learning the category labels. Fourth, and most importantly, we decided to perform the property verification task at the end of the experiment, so that both concrete and abstract words were presented. This modification was introduced in order to be sure that participants referred to the learned category names to respond.

METHOD

Participants

Eighteen students of the University of Bologna took part in the study (nine men; mean age = 23.00 years; SD = 2.30). All were native Italian speakers, both right- and left-handed (one left handed).

Procedure

All participants were submitted to two training phases (Experience; Word Acquisition) and to four different tests (Categorical Recognition without labels; Categorical Recognition with labels; Word–Object Match; Property verification task). The procedure was identical to that of Experiment 2. We only introduced three variations: (1) all abstract words were presented using both the noun and the explanation, thus the No Explanation condition for abstract words was eliminated; (2) we added a further categorical recognition task after Training 2, in order to verify whether using category labels (for both concrete and abstract words) and explanations (for abstract words) would facilitate recognition; (3) the entities to which the abstract words referred to were presented in different colors. Similarly to what we did with concrete ones, we assigned to each abstract category a specific color (light blue, light green, orange, and pink).

RESULTS

TEST 1: Categorical Recognition

In an ANOVA conducted on errors two factors were manipulated within participants, the factor Concept (Concrete vs. Abstract), and the factor Label Studied (Before vs. After learning the label designating the category). Only the factor Label Studied was significant, showing that more errors were produced before ($M = 1.01\%$) than after learning the label ($M = 0.29\%$), $F(1,17) = 36.26$, MSE = 0.26, $p < 0.00005$. Thus, both concrete and abstract category formation appears to benefit from language (see **Table 3**).

TEST 2: Words–Objects Match

An ANOVA was performed on errors produced in the word–object match. Both the factors Word and Other Exemplar were significant. Abstract words ($M = 4.46\%$) elicited more errors

Table 3 | Errors percentages and standard deviations (in parenthesis) for each TEST of Experiment 3.

EXPERIMENT 3			
CATEGORICAL RECOGNITION: WITHOUT/WITH LABEL			
Label Studied		Before learning the label	After learning the label
		1.01 (0.68)	0.29 (0.53)
WORDS–OBJECTS MATCH			
Word		Concrete	Abstract
		2.20 (1.99)	4.46 (3.63)
Other Exemplar		Novel Exemplar	Learned Exemplar
		2.08 (2.49)	4.57 (3.22)
PROPERTY VERIFICATION TASK: KEYBOARD VS. MICROPHONE			
		Device	
		Keyboard	Microphone
Word	Abstract	4.17 (1.87)	2.87 (1.34)
	Concrete	2.35 (1.72)	3.56 (1.92)

than concrete ones ($M = 2.20\%$), $F(1,17) = 8.42$, MSE = 10.89, $p < 0.01$, and more errors were produced when the exemplar nearby had already been learned ($M = 4.57\%$) than when it had not ($M = 2.08\%$), $F(1,17) = 61.85$, MSE = 1.80, $p < 0.000001$ (see **Table 3**).

TEST 3: property verification task with keyboard vs. microphone

In TEST 3 we collected both RTs and errors for the reasons we previously explained (see Experiment 2, TEST 3: Property Verification Task with Keyboard vs. Microphone). 12.93% of the trials was removed as errors. The same trimming method of Experiment 2 was used; this lead to the removal of 3.22% of the data. An ANOVA was performed with the two factors Words (Abstract vs. Concrete) and Response Device (Keyboard vs. Microphone) manipulated within participants. As expected, the difference between Abstract and Concrete Words found in Experiment 2 disappeared (mean values were respectively $M = 1150$ and 1151 ms): this demonstrates that this difference was due to the fact that processing color was easier in Experiment 2 for abstract words, as the entities they referred to were all of the same color. Crucial to our aims, the interaction between Word and Response Device was significant, $F(1,17) = 5.69$, MSE = 6173.39, $p < 0.03$ (see **Figure 4**). LSD *post hoc* showed that responses with the keyboard were slower than responses with the microphone for both Abstract and Concrete Words; however, with the first the difference was more marked ($p = 0.000005$) than with the second ($p = 0.005$). In addition, responses with the Microphone in trend were faster with Abstract than with Concrete Words ($p = 0.09$).

The interaction was also significant in a further ANOVA we performed on errors with the same factors, $F(1,17) = 35.62$, MSE = 0.80, $p < 0.00005$. *Post hoc* LSD showed that, as predicted, Abstract Words ($M = 4.17\%$) elicited more errors than Concrete Words ($M = 2.35\%$) with the Keyboard ($M = 2.87$, 3.56% respectively, $p < 0.0005$), while they elicited less errors than Concrete Words with the Microphone ($p < 0.04$). Responses to Abstract Words with the Keyboard produced more errors than all other conditions except responses to Concrete Words with the Microphone. Responses to Concrete Words with the keyboard elicited fewer errors than all other conditions except responses to Abstract Words using the Microphone (see **Table 3**).

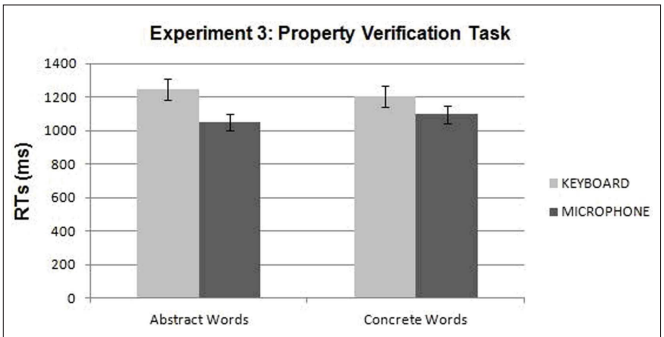


FIGURE 4 | Experiment 3: interaction between Word (Abstract, Concrete) and Response Device (Keyboard, Microphone).

DISCUSSION

Results of Experiment 3 confirmed and extend those of Experiment 2, eliminating some potential problems. Differently from Experiments 1 and 2, in TEST 1 (Categorical Recognition Task) we found no difference between abstract and concrete words, probably due to the fact that adding a property (color) to referents of abstract words increased their difference from contrast categories. Interestingly for us, in this experiment results of TEST 1 allowed us to conclude that the introduction of category labels facilitated categorization. The comparison between the same tasks performed before and after the linguistic training reveals this.

In TEST 2, the same pattern of results as Experiments 1 and 2 emerged: abstract words elicited more errors than concrete ones, thus confirming their higher complexity as well as the fact that their borders are not so clearly marked as those observed between referents of concrete words.

In TEST 3, as expected, the advantage of abstract words over concrete ones disappeared. This confirms that it was due to the modifications we made: we introduced color differences between the entities to which abstract categories referred to, in order to be certain that the task did not differ in difficulty for concrete and abstract words. The interaction between Response device and Words revealed that responses with the keyboard were always slower than responses with the microphone but that the discrepancy between microphone and keyboard was more marked with abstract than with concrete words. The pattern was complemented by the results on errors, which were fully in line with our predictions: more errors were elicited by abstract words using the keyboard, and by concrete words when using the microphone.

EXPERIMENT 4

The two last experiments left two issues unsolved. In Experiment 2 we manipulated the presence of explanations, but only for abstract words. In Experiment 3 participants were given explanations to clarify abstract word meanings because this would mirror the typical acquisition process of abstract categories. However, manipulating explanations only for abstract words did not allow us to precisely determine if there is an effect of explanation also for concrete words. Therefore, in Experiment 4 we presented only the category label or the label and the explanation for both concrete and abstract words. In addition, in this experiment for abstract words the information provided by perceptual input and that provided by the verbal label plus explanation were disentangled. To dissociate these two sources of information we used different colors for the members of abstract categories, in order to induce participants to categorize them on the basis of color, but the labels and explanations for these items still rested on items' reciprocal interaction, rather than on their perceptual features. Therefore, with concrete items the label and the explanation converged with the category formed on the basis of perceptual Experience (Training 1), whereas with abstract items the verbal and perceptual experience did not match. This manipulation was introduced in order to verify whether the advantage of the microphone responses was simply due to phono-articulatory aspects of the words or to their conceptual content as well. Our major predictions concerned TEST 3: (1) If the mouth activation found in Experiment 3 (TEST 3, vocal responses) is due to a motor phono-articulatory activation pertaining to the

superficial linguistic information, in Experiment 4 (TEST 3) we should find an advantage of vocal responses both with concrete and abstract words, as well as a main effect of the verbal explanation. (2) If, consistent with the WAT proposal, the previously found advantage for vocal responses pertains also the category content, then it should play a major role if it complements information given by perception and action, not if it contrasts with it. Therefore we should find a difference with results of Experiment 3: there should be an advantage of the microphone over the keyboard only when the label and the explanation do not contrast with perceptually based categories. In this experiment, this contrast characterizes abstract categories.

METHOD

Participants

Eighteen students of the University of Bologna took part in the study (seven men; mean age = 24.55 years; SD = 3.66). All were native Italian speakers and right-handed.

Procedure

The procedure was similar to that of Experiment 3, except for two variations. First, during Training 2 (Words Acquisition) half of the participants were taught the linguistic labels (Label group) vs. the linguistic labels plus the verbal explanation (Label + Explanation group), both for abstract and concrete items. The verbal explanations for abstract items were the same used in Experiments 2 and 3, so they basically described the kind of interaction. For concrete items the verbal explanations focused on the figure shape, avoiding any reference to its color (e.g., CALONA: "a figure having a concavity"). The number of words for each explanation across both the abstract and the concrete blocks was equal.

Second, in Experiment 4 we used different colors for each category member: for both concrete and abstract items, two members of each category had the same color as two members of another category. For example, FUSAPO surface could be yellow, blue, red, or sky-blue; its thickness was always the same, i.e., dark blue. NOROLO surface shared with FUSAPO surface yellow and blue colors, but it could be also green or violet; the color of the thickness was always dark blue.

RESULTS

TEST 1: Categorical Recognition

We performed two different ANOVAs on errors: one for the Label group and another for the Label + Explanation group. In the first ANOVA two factors were manipulated within participants, the factor Concept (Concrete vs. Abstract), and the factor Label Studied (Before vs. After learning the category label). In the second ANOVA we manipulated the same factors, but the levels of Label Studied factor differed (Before vs. After learning the category label with explanation). In the first ANOVA, both main effects were significant: more errors were produced with Abstract (M = 7.41%) than with Concrete Concepts (M = 3.36%), $F(1,8) = 7.73$, MSE = 19.12, $p < 0.03$, and more errors were produced before (M = 6.54%) than after learning the label (M = 4.22%), $F(1,8) = 17.31$, MSE = 2.79, $p < 0.005$. The factor Concept was also significant in the second ANOVA: more errors were produced for Abstract (M = 11.17%) than for Concrete Concepts (M = 3.60%), $F(1,8) = 32.61$, MSE = 15.38, $p < 0.0005$. The factor Label Studied did not reach significance, but we found a significant interaction between Concept and Label Studied, $F(1,8) = 7.26$, MSE = 1.83, $p < 0.05$ (see **Table 4**), due to the fact that after learning label + explanation errors decreased with concrete words (LSD *post hoc*, $p > 0.005$), but not with abstract ones.

Table 4 | Errors percentages and standard deviations (in parenthesis) for each TEST of Experiment 4.

EXPERIMENT 4			
CATEGORICAL RECOGNITION: WITHOUT/WITH LABEL			
Group A: label		Concept	
		Concrete	Abstract
		3.36 (2.59)	7.41 (5.04)
Group B: label + explanation		Concept	
		Concrete	Abstract
		3.70 (2.35)	11.17 (4.47)
WORDS–OBJECTS MATCH			
Group A: label		Word	
		Concrete	Abstract
		2.55 (4.08)	6.48 (4.57)
Group B: label + explanation		Word	
		Concrete	Abstract
		2.66 (4.45)	7.20 (4.36)
PROPERTY VERIFICATION TASK			
		Device	
		Keyboard	Microphone
Word	Abstract	18.23 (8.37)	21.79 (6.17)
	Concrete	17.36 (6.38)	13.89 (5.19)

TEST 2: Words–Objects Match

We performed two different ANOVAs on errors: one for the Label group and another for the Label + Explanation group. In the first ANOVA two factors were manipulated within participants: Word (Concrete vs. Abstract) and Other Exemplar (Exemplar already learned, with only linguistic label vs. Exemplar not learned). In the second ANOVA the same factors were manipulated, but the levels of the Other Exemplar factor differed (Exemplar already learned, with label + explanation vs. Exemplar not learned).

In both ANOVAs we found a significant main effect of the factor Word: fewer errors were produced with Concrete than With Abstract Words (group A: M = 2.55% and 6.48% respectively, $F(1,8) = 8.31$, MSE = 16.77, $p < 0.05$; group B: M = 2.66% and 7.29% respectively, $F(1,8) = 22.13$, MSE = 8.71, $p < 0.005$; see **Table 4**).

TEST 3: property verification task with keyboard and microphone

In TEST 3 for RTs 35.63% of the trials was removed as errors. We used the same trimming method as in previous experiments; this lead to the removal of 2.38% of the data. An ANOVA was performed with three factors: Word (Abstract vs. Concrete), Response Device (Keyboard vs. Microphone), and Verbal Explanation (Without vs. With), the last one manipulated between participants. We found that vocal responses (M = 1128.73 ms) were 147.57 ms faster than manual responses (M = 1276.30 ms), even if the factor Response

Device did not reach significance, $F(1,16) = 3.48$, MSE = 112633, $p < 0.08$. The interaction between the factors Word and Response Device was significant, $F(1,16) = 4.58$, MSE = 47804.8, $p < 0.05$. The advantage of the microphone over the keyboard did not reach significance with abstract words ($M = 1221.67$ vs. $M = 1184.37$, respectively), while with concrete words responses with the microphone ($M = 1073.09$) were faster than responses with the keyboard ($M = 1330.93$ ms; LSD, $p > 0.01$; see **Figure 5**).

Finally in the ANOVAs on errors with the same factors, we found that abstract words ($M = 20.01\%$) elicited more errors than concrete ones ($M = 15.63\%$), $F(1,16) = 7.84$, MSE = 44.13.08, $p < 0.05$. The significant interaction between Word and Response Device, $F(1,16) = 5.87$, MSE = 37.90, $p < 0.05$, was due to the fact that abstract words with the microphone ($M = 21.79\%$) elicited more errors than concrete words with both the keyboard ($M = 17.36\%$) and the microphone ($M = 13.89\%$; LSD *post hoc*, $p > 0.05$; see **Table 4**).

DISCUSSION

Results of TEST 1 indicate that the difference between the condition Label and No-Label increases when an explanation is added to the category name. Thus explanations facilitate categorization, as they render category boundaries more marked and clearer. However, the contribution of explanations is relevant only for concrete categories. For abstract categories, explanations do not help, as the information they provide is in contrast with perceptually based categorization.

Results of TEST 3 are the most intriguing. As predicted, participants were faster to respond with the microphone than with the keyboard with all words: this suggests that the phono-articulatory aspect of the words pronounced during acquisition affects performance. It is unclear, however, why no effect of explanation was present. The most important result is the interaction showing that the advantage of the microphone over the keyboard is more marked with concrete than with abstract words, both in RTs and accuracy. This suggests that not only phono-articulatory but also conceptual information is at play in explaining the advantage of responses with the microphone. In fact this advantage shows up only when there is a convergence between the linguistic information (label and explanation) and the category formed on sensorimotor basis, that is only with concrete words. One could object that the effect is due to the fact that explanations used with concrete words might have reduced the importance of manual interaction with objects. However, this does not account for the advantage of the microphone with concrete over abstract words.

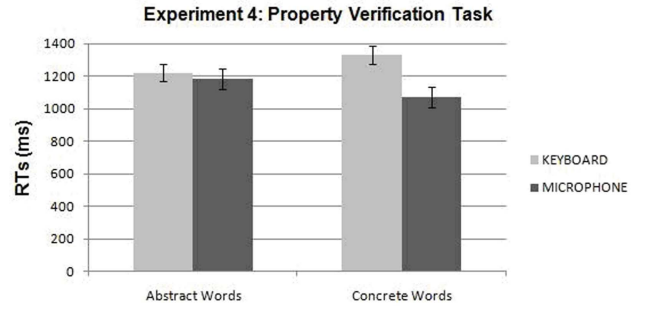


FIGURE 5 | Experiment 4: interaction between Word (Abstract, Concrete) and Response Device (Keyboard, Microphone).

GENERAL DISCUSSION

Four experiments were designed to study the acquisition of concrete and abstract categories and words. We chose to use novel categories, in order to avoid confounds often associated with research on concrete and abstract words. We identified some characteristics which are typical of abstract but not of concrete categories, and we created novel categories according to these criteria. First, abstract categories do not refer to a single object but rather to a complex relationship between different objects. In addition, the entities to which abstract categories refer are not manipulable, even though they are perceivable, as they are interacting moving objects. Notice that our distinction does not cover the whole continuum ranging from abstract to concrete categories. Further work is needed for a thorough investigation of different typologies of abstract words (for attempts in this direction, see Rüschemeyer et al., 2007; Setti and Caramelli, 2005). Here we used two different examples of concrete and abstract words and have shown that different processes are involved in their acquisition.

In Experiment 1 we ascertained, using a production task, that the pattern of produced properties with our novel concrete and abstract categories was similar to that typically elicited by concrete and abstract words.

In Experiments 2, 3, and 4 we introduced a modification: abstract words were not only learned by associating a label with the entities/relations they referred to, but also when an explanation of their meaning was provided. This learning situation should resemble the learning process of children, as studies on MOA show. We found that this learning process influenced a later property verification task: participants responded earlier to concrete words while using the keyboard, while responses with abstract words were faster while using the microphone. Similar results with action words and effectors showed that, while comprehending sentences referring to mouth-related actions, response times were faster with the microphone than with the keyboard (Scorolli and Borghi, 2007). In addition, in line with WAT, participants' performance with abstract words was improved when provided with a verbal explanation (Experiment 2, group B; Experiment 3). This effect was not observed in concrete words. The fact that the advantage of the explanation was confined to abstract words revealed that the difference is not simply due to phono-articulatory aspects, but that for accessing the meaning of abstract words linguistic information plays a major role. This was confirmed in Experiment 4, in which we found that, due to the fact that with abstract words the verbal label and explanation were in contrast with the already formed perceptually based category, the advantage of the microphone over the keyboard was reduced compared to the other experiments.

Our results are in line with embodied and grounded theories of categorization and language comprehension. Namely, both the concrete and the abstract categories we used are embodied and grounded (e.g. Barsalou and Wiemer-Hastings, 2005; Casasanto and Boroditsky, 2008; Glenberg et al., 2008; Boot and Pecher, 2009), as they have objects or relations as referents. We were able to demonstrate that they are not only grounded in perception and action systems, but that for forming them language is important. This leads to the prediction that abstract

words would not only activate linguistic areas in the brain, but also classic motor and sensorimotor areas (Scorolli et al., 2010; for initial fMRI results with existing concrete and abstract word combinations, see work performed within the project www.rossiproject.eu).

Results are in line with the predictions advanced by the WAT proposal. They reveal that a different learning process might lead to differences in performance on different tasks, such as a production task vs. a property verification task. In addition, the present study provides evidence that for representing abstract concepts motor linguistic information is more important than manual information, whereas for representing concrete concepts the pattern is opposite.

One effect was not predicted by the WAT proposal. Our results show that the formation of both concrete and abstract categories benefits from learning a linguistic label. As it emerges from the categorical recognition task in Experiment 3, language is relevant because it helps to better differentiate between categories (Mirolli and Parisi, in press). The recognition test in Experiment 4 (TEST 1a,b) shows that labeling is mostly helpful when an explanation of the category meaning is added. As shown in TEST 2 and TEST 3 of Experiment 2, the benefit provided by language is higher in the case of abstract categories when a verbal explanation (group B) supports the linguistic label. Nevertheless, when no explanation is provided, labeling is useful for both concrete and abstract categories. In sum: labeling helps categorization, independently of category complexity. However, even when no explanation is provided, given that abstract words do not refer to manipulable objects and are linked by complex relational properties, language plays a major role in their representation.

This opens an interesting scenario. Language is relevant for both concrete and abstract words because it helps better differentiate between categories. However, in tasks for which categorization is not relevant, such as the color verification task, it is more accessible in the representation of abstract than of concrete word meanings. This might occur because: (a) the members of abstract categories are not manipulable; and (b) more linguistic information is typically associated with the acquisition of abstract word meanings.

Further work should address unsolved issues in this research. One important expansion could be to introduce the social development component implied in word acquisition. We used language in a very simple way, through adding labels or explanations to read and to associate with the relevant categories. Thus, language was not associated with a real social experience, as in real life. Further work should take into account aspects of social development which characterize language acquisition.

Finally, we believe this work has important implications for modeling. The design of the task is particularly suitable for further modeling applications and replications. We succeeded in isolating some properties we believe to be relevant in real-life categories and built novel categories based on our assumptions. We could verify that the behavioral responses produced within these novel categories were similar to the ones produced within real-life categories and settings. This procedure demands an additional modeling development. We believe that computational models can integrate and generate a more general description of the experimental results. For example, a robotic model, as discussed in the introduction, can benefit from a psychological theory that provides a possible way to tackle a new and complex problem for robotics itself, such the theory described focusing on the grounding and acquisition of abstract words. On the other hand, the same robotic model can be tested in many different experimental situations, some of them not even applicable to human subjects. Experiments of this nature can complement and integrate experiments with human participants and can offer new insights and hypotheses to test. Moreover, the process of developing artificial cognitive models always requires a profound articulation of the theory implemented. This fact forces the researcher to well define and to operationally describe every aspect of the theory and, at the same time, it emphasizes the importance of the central aspects of the theory, that can be fully exploited and validated by the model, at least as a preliminary proof of concept.

ACKNOWLEDGMENTS

Thanks to people of the EMCOlab for discussions and useful suggestions on this paper. Thanks to Kate Burke for revision of the English text. This work was supported by the European Community, project ROSSI: Emergence of communication in RObots through Sensorimotor and Social Interaction (Grant agreement n. 216125).

REFERENCES

Altarriba, J., Bauer, L. M., and Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behav. Res. Methods Instrum. Comput.* 31, 578–602.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford, UK: Oxford University Press), 245–284.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thought*, eds D. Pecher and R. Zwaan (New York: Cambridge University Press), 129–163.

Boot, I., and Pecher, D. (2009). Similarity is closeness: metaphorical mapping in a conceptual task. *Q. J. Exp. Psychol.* 29, 1–13.

Borghi, A. M., and Barsalou, L. W. (2001). *Perspective in the Conceptualization of Categories, Abstracts*. Orlando: Psychonomic Society.

Borghi, A., and Caramelli, N. (2001). "Taxonomic relations and cognitive economy in conceptual organization," in *Proceedings of 23rd Meeting of the Cognitive Science Society*, eds J. D. Moore and K. Stenning (London: Erlbaum), 98–103.

Borghi, A. M. (2004). Objects concepts and action: extracting affordances from objects' parts. *Acta Psychol. (Amst)* 115, 1, 69–96.

Borghi, A. M., Caramelli, N., and Setti, A. (2005). Conceptual information on objects' locations. *Brain Lang.* 93, 140–151.

Borghi, A. M., and Cimatti, F. (2009). "Words as tools and the problem of abstract words meanings," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. Taatgen and H. van Rijn (Amsterdam: Cognitive Science Society), 2304–2309.

Borghi, A. M., and Cimatti, F. (2010). Embodied cognition and beyond: acting and sensing the body. *Neuropsychologia* 2010, 48, 763–773.

Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593.

Clark, A. (1998). "Magic words: how language augments human computation," in *Language and Thought. Interdisciplinary Themes*, eds P. Carruthers and J. Boucher (Oxford: Oxford University Press), 162–183.

Crutch, S. J., and Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain* 128, 615–627.

Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language.*

Cambridge: Cambridge University Press.

Desai, R. H., Binder, J. R., Conant, L. L., and Seidenberg, M. S. (2010). Activation of sensory-motor areas in sentence comprehension. *Cereb. Cortex* 20, 468–478.

Dove, G. (2009). Beyond perceptual symbols: a call for representational pluralism. *Cognition* 110, 412–431.

Fodor, J. A. (1998). *Concepts. Where Cognitive Science Went Wrong.* Cambridge, MA: Harvard University Press.

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., and Buccino, G. (2008). Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* 61, 905–919.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Keil, F. C. (1989). *Concepts, Kinds, and Cognitive Development.* Cambridge, MA: MIT Press.

Louwerse, M. M., and Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition* 114, 96–104.

Marocco, D., Cangelosi, A., Fischer, K., and Belpaeme, T. (2010). Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Front. Neurorobot.* 4:7. doi: 10.3389/fnbot.2010.00007

Mirolli, M., and Parisi, D. (in press). Towards a vygotskyan cognitive robotics: the role of language as a cognitive tool. *N. Ideas Psychol.*

Paivio, A., Yuille, J. C., and Madigan, S. A. (1986). Concreteness, imagery and meaningfulness values for 925 words. *J. Exp. Psychol.* 76(Suppl.), 1–25.

Rüschemeyer, S. A., Brass, M., and Friederici, A. D. (2007). Comprehending prehending: neural correlates of processing verbs with motor stems source. *J. Cogn. Neurosci.* 19, 855–865.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., and Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage* 27, 188–200.

Scorolli, C., and Borghi, A. (2007). Sentence comprehension and action: effector specific modulation of the motor system. *Brain Res.* 1130, 119–124.

Scorolli, C., Jacquet, P., Binkofski, F., Nicoletti, R., Tessari, A., and Borghi, A. M. (2010). "Processing abstract and concrete sentences: a role of primary motor cortex?" in *Fourth Annual Rovereto workshop on Concepts Actions and Objects, Functional and Neural perspective* (Rovereto: CAOS).

Setti, A., and Caramelli, N. (2005). "Different domains in abstract concepts," in *Proceedings of the XXVII Annual Conference of the Cognitive Science Society,* eds B. Bara, L. Barsalou, and M. Bucciarelli (Mahwah, NJ: Erlbaum), 1997–2002.

Steels, L. (2003). Evolving grounded communication for robots. *Trends Cogn. Sci.* 7, 308–312.

Sugita, Y., and Tani, J. (2005). Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adapt. Behav.* 13, 211–225.

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008). "An open-source simulator for cognitive robotics research: the prototype of the iCub humanoid robot simulator," in *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS*08), Gaithersburg, MD.

Wauters, L. N., Tellings, A. E. J. M., Van Bon, W. H. J., and Van Haaften, A. W. (2003). Mode of acquisition of word meanings: the viability of a theoretical construct. *Appl. Psycholinguist.* 24, 385–406.

Wiemer-Hastings, K., Krug, J., and Xu, X. (2001). "Imagery, context availability, contextual constraint and abstractness," in *Proceedings of the XXIII Annual Conference of the Cognitive Science Society. Edinburgh, Scotland, August 1–4, 2001* (Mahwah, NJ: Lawrence Erlbaum Associates.

Wittgenstein, L. (1953). *Philosophische Untersuchungen.* Oxford: Blackwell.

Wu, L. L., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol.* 132, 173–189.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 July 2010; accepted: 17 January 2011; published online: 09 February 2011. Citation: Borghi AM, Flumini A, Cimatti F, Marocco D and Scorolli C (2011) Manipulating objects and telling words: a study on concrete and abstract words acquisition. Front. Psychology ***2****:15. doi: 10.3389/fpsyg.2011.00015*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Borghi, Flumini, Cimatti, Marocco and Scorolli. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 27 July 2011
doi: 10.3389/fpsyg.2011.00162

Architecture of explanatory inference in the human prefrontal cortex

Aron K. Barbey[1,2,3]* and Richard Patterson[4]

[1] *Decision Neuroscience Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[2] *Center on Health, Aging, and Disability, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[3] *Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, IL, USA*
[4] *Department of Philosophy, Emory University, Atlanta, GA, USA*

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Diane Pecher, Erasmus University Rotterdam, Netherlands
Chad Edward Forbes, University of Delaware, USA

****Correspondence:***
Aron K. Barbey, Decision Neuroscience Laboratory, 110 Huff Hall, 1206 South Fourth Street, Champaign, IL 61820 MC-586, USA.
e-mail: barbey@Illinois.edu
http://www.DecisionNeuroscienceLab.org/

Causal reasoning is a ubiquitous feature of human cognition. We continuously seek to understand, at least implicitly and often explicitly, the causal scenarios in which we live, so that we may anticipate what will come next, plan a potential response and envision its outcome, decide among possible courses of action in light of their probable outcomes, make midstream adjustments in our goal-related activities as our situation changes, and so on. A considerable body of research shows that the lateral prefrontal cortex (PFC) is crucial for causal reasoning, but also that there are significant differences in the manner in which ventrolateral PFC, dorsolateral PFC, and anterolateral PFC support causal reasoning. We propose, on the basis of research on the evolution, architecture, and functional organization of the lateral PFC, a general framework for understanding its roles in the many and varied sorts of causal reasoning carried out by human beings. Specifically, the ventrolateral PFC supports the generation of basic causal explanations and inferences; dorsolateral PFC supports the evaluation of these scenarios in light of some given normative standard (e.g., of plausibility or correctness in light of real or imagined causal interventions); and anterolateral PFC supports explanation and inference at an even higher level of complexity, coordinating the processes of generation and evaluation with further cognitive processes, and especially with computations of hedonic value and emotional implications of possible behavioral scenarios – considerations that are often critical both for understanding situations causally and for deciding about our own courses of action.

Keywords: explanatory inference, causal reasoning, executive control, lateral prefrontal cortex

PREFRONTAL CONTRIBUTIONS TO EXPLANATORY INFERENCE

The human mind is driven toward understanding. We wonder why events unfold in particular ways, why objects have specific properties and why people behave the way they do. The capacity to infer the causal structure of experience and to generate explanations is central to our sense of understanding, making possible the formation of conceptual representations that constrain inference, guide generalization, and provide the basis for goal-directed, intelligent behavior. Accordingly, extensive research in social psychology and philosophy is dedicated to the study of explanation, with social psychology focusing on explanations of behavior (e.g., Heider, 1958; Gilbert, 1998; Malle, 2004) and philosophy on explanation in science (e.g., Salmon, 1998; Kitcher and Salmon, 1989). Only recently, however, has cognitive science addressed such questions as what constitutes an explanation, what makes some explanations better than others, and the principles that determine when we seek explanations and how we generate them (e.g., Keil and Wilson, 2000; Keil, 2006).

Two recent developments have spurred the emergence and growth of research on explanation within cognitive psychology. First, prominent theories of conceptual knowledge accord a central role to explanation (e.g., Carey, 1985; Murphy and Medin, 1985; Murphy, 2002; Keil, 2003, 2006). Explanations facilitate category learning, influence judgments of the typicality of category members and foster conceptual coherence. Second, explanations and the causal representations they entail exert a profound influence on learning and reasoning (e.g., Barbey and Wolff, 2006, 2007, under review; Tenenbaum et al., 2006; Sloman et al., 2009; Wolff et al., 2010). Cognitive psychologists have therefore increasingly recognized the importance of investigating the psychology of explanation. We suggest that category learning, typicality judgments, reasoning, and conceptual coherence are strongly interconnected, and that our beliefs about the causal powers of objects, events, and agents – and about the rule-like causal relationships among them – are central to the generation and evaluation of the myriad ways in which we interpret, understand and explain ourselves and our environment.

Parallel developments in cognitive neuroscience have fostered the study of the neural mechanisms underlying explanation. For instance, the resurgence of cognitive simulation theories has motivated neuroscience models of explanatory inference based on the simulation of modality-specific components of experience (e.g., Damasio, 1989; Barsalou, 1999; Barbey and Barsalou, 2009; Barbey and Grafman, in press, 2011; Barbey et al., 2011a,b). According to this framework, a highly integrative, multimodal representation system in the brain supports simulation mechanisms for explanatory inference across the spectrum of cognitive activities, including high-level perception, implicit memory, working memory, long-term memory, and conceptual knowledge (for a recent review, see Barbey and Barsalou, 2009). Although the functional topography

of modality-specific representations and their role in these processes have become increasingly well understood, remarkably little is known about the simulation mechanisms that encode the higher-level structure of experience, representing causal relationships that support explanatory inference and establish the proper mappings between situations, actions and consequences necessary for coordinated, purposeful behavior. The absence of such data represents a substantial gap in understanding both the neural architecture of cognitive simulations and their role in higher cognitive functions.

Here we introduce an integrative cognitive neuroscience theory for understanding the mechanisms that enable the top-down control and coordination of modality-specific representations, drawing upon recent theoretical developments in cognitive psychology and emerging neuroscience evidence indicating that the lateral prefrontal cortex (PFC) supports the generation, evaluation, and coordination of representations that encode the causal structure and probable causal implications of events, and thus provide the basis for causal understanding of our environment and for our day-to-day navigation through that environment. We regard explanation featuring a specific causal mechanism as the central or prototypical case of causal explanation. Less central are cases in which one must act without having settled confidently on one mechanism or another, or again, while simply assuming that there is some causal mechanism at work but without having any good idea what it might be. At the ragged edge of understanding we sometimes have to be content with statistical correlations. Here we note another significant gap in current understanding of causal explanations: statistical correlations do in many circumstances give rise to the formulation of causal generalizations (including causal mechanisms) that we then apply to new cases in relevantly similar circumstances. And although we will review evidence that different brain systems support inductive reasoning to probabilistic generalizations on the one hand, and deductive causal reasoning (from a "major premise" asserting a general causal relationship and another premise bringing some particular event under that causal rule) on the other, the circumstances under which, and the processes by which, we move from statistical correlation to causality remain to be investigated.

We begin by reviewing psychological research on the structure of explanations, surveying contemporary research and theory from cognitive psychology suggesting that explanatory inference accommodates novel information in the context of background beliefs, as it enables generalizations and predictions about self, others, and the environment. We then review the biology, evolution and ontogeny of the human PFC, and introduce a cognitive neuroscience framework for causal inference based on a functional division of labor within the lateral PFC. Our review examines a broad range of evidence from the social and decision neuroscience literatures demonstrating that computational mechanisms for the generation and evaluation of causal simulations are mediated by functionally specialized regions of the lateral PFC (ventrolateral PFC and dorsolateral PFC, respectively), and that at yet higher levels of complexity, where these and other cognitive processes must be coordinated, causal inference is further supported by the anterolateral PFC. We show how this framework supports the integration of a diverse body of neuroscience evidence concerning human reasoning not just about basic physical and social contexts, but also within the context of moral, ethical, and legal systems of value and belief.

PSYCHOLOGY OF EXPLANATION

Psychological evidence supports the predominance of causation in explanation (e.g., Barbey and Wolff, 2006, 2007, under review; Tenenbaum et al., 2006; Sloman et al., 2009; Wolff et al., 2010). Explanations typically appeal to causes, along with knowledge of general patterns that constrain which causes are judged to be probable (Einhorn and Hogarth, 1986) and relevant (Lombrozo and Carey, 2006; Wellman and Liu, 2007). Explanations recruit a great deal of prior knowledge, establishing relevant causal mechanisms that provide a source of constraint for reasoning and a basis for generalizing from known to novel cases (reviewed in Lombrozo, 2006). As a consequence, the top-down control and coordination of behavior depends on the capacity to generate causal explanations and understanding of the physical and social world.

CAUSAL REPRESENTATIONS

A major function of the PFC is to extract statistical regularities across experience in an effort to infer general patterns and causal relationships that establish the proper mappings between situations, actions and consequences necessary for goal-directed behavior (for reviews, see Miller, 2000; Miller and Cohen, 2001; see also Barbey et al., in press). By extracting the higher-level structure of experience, representations within the PFC enable the top-down control and coordination of multiple brain mechanisms across diverse brain areas and networks (for a recent review of the neurobiological mechanisms underlying PFC function, see Miller and Phelps, 2010).

CAUSAL REPRESENTATIONS IN THE LATERAL PFC

Behaviorally relevant causal information about causal powers and causal associations and patterns, and the causal inferences these support, are encoded by diverse areas of lateral PFC (**Figure 1**; for reviews, see Miller, 2000; Miller and Cohen, 2001). We will not discuss here the issue of whether our representations of events, or of agents and objects and their causal powers, are through-and-through a matter of modal simulations (as when we see or imagine billiard balls colliding, levees breaking, etc.), or whether more schematic and abstract – and perhaps even amodal – representations are involved. Elsewhere we suggest and defend a pluralistic approach within which modal simulation is the evolutionarily oldest – and probably still the default – medium of causal reasoning, but on which one employs more or less schematic, and more or less abstract modes of reasoning depending on one's circumstances (Patterson and Barbey, in press). As will become evident, the framework we propose here will accommodate a wide variety of views about the ground-level nature of causal representations.

While the lateral PFC is a site of convergence for the synthesis of multimodal information from a wide range of brain systems (see below on connectivity of these regions), we propose that the neural architecture of the lateral PFC entails two pathways for explanatory inference. The ventrolateral PFC supports the generation and maintenance of causal simulations, relying upon computational mechanisms for detecting and encoding causal relationships. Within this framework, as a causal event is experienced repeatedly, its simulated components and the causal relationships linking them increase in potency. Thus when one component is perceived

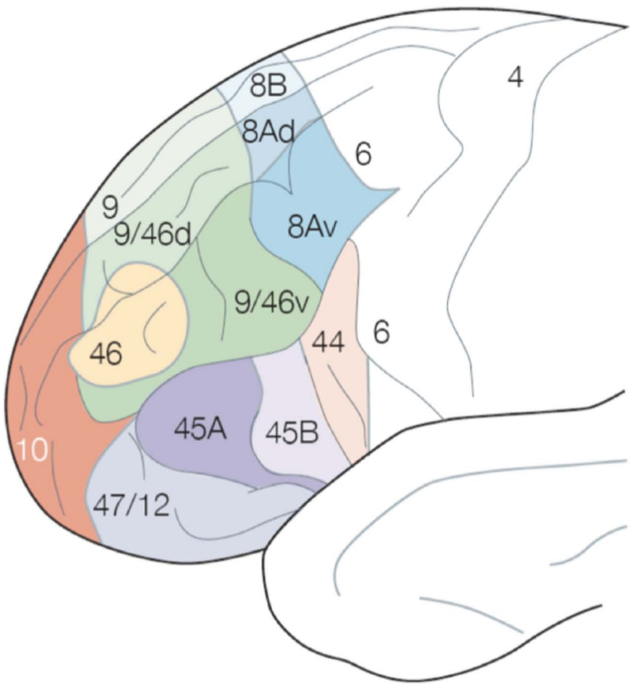


FIGURE 1 | Brodmann map of the lateral PFC. Reproduced with permission from Ramnani and Owen (2004).

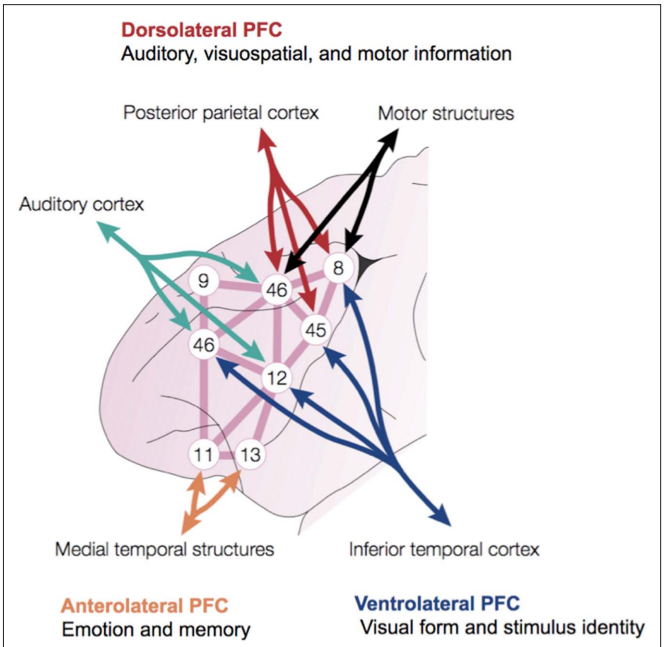


FIGURE 2 | Integrative anatomy of the macaque monkey PFC. Numbers refer to subregions within the lateral PFC defined by Brodmann. Modified with permission from Miller (2000).

initially, these strong associations complete the pattern automatically, supporting inferences about the underlying cause(s) and their resulting effect(s).

On the basis of this same experience one also forms representations of the causal powers, active and receptive, of the agents, objects, and events involved in causal situations. These representations underlie the implicit and explicit production of novel and counterfactual simulations critical for planning, monitoring, and adjusting behavior.

These first level explanations and inferences must often be evaluated and re-evaluated as we make our way in the world – e.g., by devising, imagining, or performing an intervention to find out whether or not an effect is present in the absence of the candidate cause. Such evaluations are supported by computational mechanisms in the dorsolateral PFC. This framework of causal evaluation operating over representations of causal patterns and powers is supported in the first instance by classic neuroscience research on working memory, which demonstrates that the ventrolateral PFC supports the maintenance of cognitive representations and the dorsolateral PFC is additionally recruited for monitoring and manipulating items (e.g., Petrides, 2005; D'Esposito et al., 1999). Further, and more direct, evidence is reviewed below.

ANATOMICAL CONNECTIVITY, EVOLUTION, AND DEVELOPMENT OF THE LATERAL PFC

The inferential architecture of the lateral PFC derives from the anatomical connectivity, evolution, and development of this region. The lateral PFC consists of three major subregions that emphasize processing of particular information based on their interconnections with specific cortical regions (**Figure 2**).

Ventrolateral areas are more heavily interconnected with cortical regions for processing information about visual form and stimulus identity (inferior temporal cortex) that supports the detection of causal relationships and the categorization of environmental stimuli. Dorsal portions of the lateral PFC are heavily interconnected with cortical areas for processing visuospatial and motor as well as auditory information. It is primarily the capacity to manipulate visuospatially arrayed representations of objects and events in causal scenarios that makes possible the evaluation and adjustment of causal understanding to serve one's short and long-term needs (Barbey et al., 2009a). Finally, the anterolateral PFC is indirectly connected (via the ventromedial PFC) with limbic structures that process internal information, such as emotion, memory, and reward (Goldman-Rakic, 1987; Pandya and Barnes, 1987; Fuster, 1989; Barbas and Pandya, 1991). The lateral PFC therefore enables the synthesis of information across this broadly distributed network of modal brain regions (for pertinent reviews, see Kringelbach, 2005).

Research investigating the evolution and ontogeny of the PFC suggests that the lateral PFC initially emerged from ventrolateral prefrontal regions, followed by dorsolateral and then anterolateral cortices (**Figure 3**; Flechsig, 1901, 1920; Fuster, 1997).

From an evolutionary perspective, the emergence of lateral PFC subregions reflects their relative priority for goal-directed behavior, with the ventrolateral PFC enabling the capacity to maintain basic causal beliefs and generate explanatory inferences. The fine details of the human capacity to represent a range of possible causal outcomes or antecedents of a given situation are not yet understood, but we suggest that these go hand-in-hand with the capacity to evaluate causal explanations and to plan, monitor, and adjust causal behavior in light of our causal understandings – abilities supported by the dorsolateral PFC.

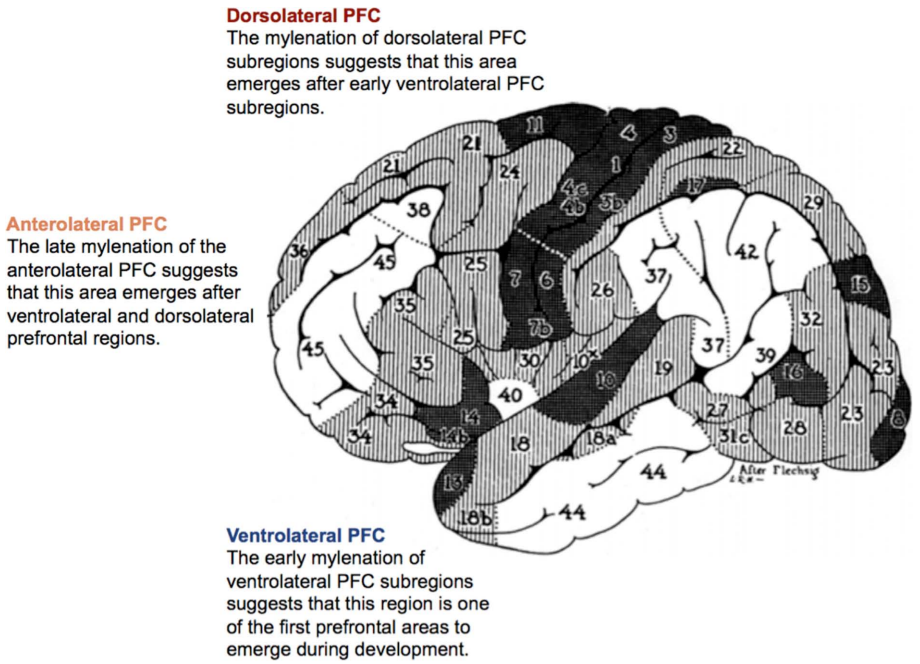


FIGURE 3 | Ontogenetic map of the prefrontal cortex according to Flechsig (1901, 1920). The numeration of the areas indicates the order of their myelination. Modified with permission from Flechsig (1920).

Finally, the evolution of the anterolateral PFC enabled processing of higher-order relations and reasoning about complex forms of goal-directed behavior involving both the generation and evaluation of explanatory inferences (for a review, see Ramnani and Owen, 2004), but also the integration of these processes with hedonic and emotional information associated with different causal scenarios, and especially with different possible causal outcomes. Consistent with its evolutionary development, the ontogeny of the lateral PFC reflects the importance of first establishing explanations for understanding the physical and social world, followed by the capacity to manipulate and evaluate these explanations, and finally high-order inferences involving both sorts of activity – along with the coordination of these processes with further relevant information and computation including the assessment of hedonic outcomes of possible actions. This coordination of multiple processes will routinely characterize human inferences about the multifaceted (causal) outcomes of actions, and is in general supported by the anterolateral PFC (Ramnani and Owen, 2004). We focus here on the anterolateral PFC, but with some reference to its connections with the larger anterior PFC and orbitofrontal cortex (OFC), among other regions. Hedonic valences of rewards and punishers along with their connections to specific stimuli are represented largely in OFC, and this information will typically be incorporated into human calculations of outcomes and decision making (Kringelbach, 2005). For a recent review of pertinent developmental evidence, see Rochat (2009).

This is not to suggest that any of such functions (e.g., stimulus identification via spatial properties, spatial mapping of the environment, control of behavior, engagement in social transactions, etc.) are carried out solely by any one region of PFC. On the contrary, physical, biological, emotional, social, and other information will be used in performing the functions we associate with all three areas of PFC focused on here. The suggestion is rather that these functions and their anatomical correlates are integrated, and that there is substantial evidence that one useful way of distinguishing functions of the PFC with regard to causal explanation and inference in particular coincides with the anatomical division into ventrolateral, dorsolateral, and anterolateral PFC.

An empirical case for the functional specificity of ventrolateral PFC for the maintenance of information and dorsolateral PFC in the manipulation of representations has been established in the functional neuroimaging literature on working memory, providing evidence that broadly supports the proposed functional organization of lateral PFC (for meta-analytic reviews, see Wager and Smith, 2003; Wager et al., 2004; Owen et al., 2005). It makes good sense that dorsolateral PFC is, on the one hand, heavily involved in the evaluation of causal explanations and, on the other hand, is strongly interconnected with visuospatial processing areas, because such evaluation is typically carried out via manipulation of visuospatial arrays of representations of potential causes, background conditions, enablers, etc. This holds for theories emphasizing the use of modal simulations as well as for those centering on much more schematic, abstract, and even amodal representations (see Barbey and Wolff, 2006, 2007; Patterson and Barbey, in press, under revision; Sloman et al., 2009). Meanwhile anterolateral PFC has been shown in a range of research reviewed above to be involved on the one hand in higher-order reasoning and, on the other, in social cognition. This correlation, too, makes good sense in that human social understanding and decision making are frequently complex, drawing on and integrating multiple cues of diverse types

into coherent explanatory scenarios. "Immediate" situations are in turn imbedded in larger causal scenarios and narratives that one must take into account, where these sometimes reach as far as overarching life goals, and where one wants to consider at many points the likely hedonic and emotional impact of possible actions. And of course, such representations are subject to both top-down and bottom-up influences, as we consider "what difference it might make" for our pursuit of some larger goal if we undertake one immediate action rather than another, or for what we should do here and now if we are to further one long-term goal rather than another. This further suggests that for the anterolateral PFC to fulfill its role not just with regard to higher-order inference in general, but social and emotional life in particular, its strong connectivity to OFC and, via ventromedial PFC, with the limbic system is critical, since our inferences about what will result from a given action, and for whom, and how, will have to include much information about, and computations of, hedonic and emotional values. But the evolutionary history of the connectivity of the anterolateral PFC with OFC and subcortical limbic areas remains to be written.

We note finally that the picture given just above dovetails with additional evidence that the anterior-to-posterior axis of the lateral PFC is organized hierarchically, whereby progressively anterior subregions are associated with higher-order processing requirements for planning and the selection of action (for recent reviews, see Ramnani and Owen, 2004; Koechlin and Summerfield, 2007; Badre, 2008; Botvinick, 2008). Thus, processes within the lateral PFC respect the hierarchical organization of this region, with progressively anterior regions representing causal simulations that support higher-order inferences based on computational mechanisms for generating and evaluating explanations.

INFERENTIAL ARCHITECTURE OF THE LATERAL PFC

We now turn to a review of recent evidence from the social and decision neuroscience literatures demonstrating (1) the involvement of the ventrolateral PFC for the generation and maintenance of explanatory inferences, (2) the recruitment of the dorsolateral PFC for evaluating possible explanations in light of normative criteria, and (3) activation in the anterolateral PFC for manipulation and utilization of higher-order inferences that incorporate both types of process and also coordinate these with other relevant processes, such as computation of hedonic values of predicted outcomes of potential actions (**Figure 4**). The representational architecture underlying these forms of inference further predicts the recruitment of broadly distributed neural systems, incorporating medial prefrontal (Barbey et al., 2011a; for reviews, see Kringelbach, 2005; Amodio and Frith, 2006) and posterior knowledge networks (e.g., Simmons et al., 2010) representing unimodal and multimodal components of experience.

VENTROLATERAL PFC

Social neuroscience studies have shown that explanatory inference is mediated by the ventrolateral PFC (areas 44, 45, and 47; **Figure 4B**). Fiddick et al. (2005), observed recruitment of ventrolateral PFC when participants drew explanatory social inferences based on normative beliefs concerning reciprocal altruism and social exchange. Speaking more generally, this region is recruited when representing normative rules that guide social behavior (Barbey et al., 2009a). It is particularly sensitive to norm violations that motivate explanatory inferences about the cause(s) of deviant (i.e., non-normative) behavior (for recent behavioral evidence, see Uttich and Lombrozo, 2010). Berthoz et al. (2002), for

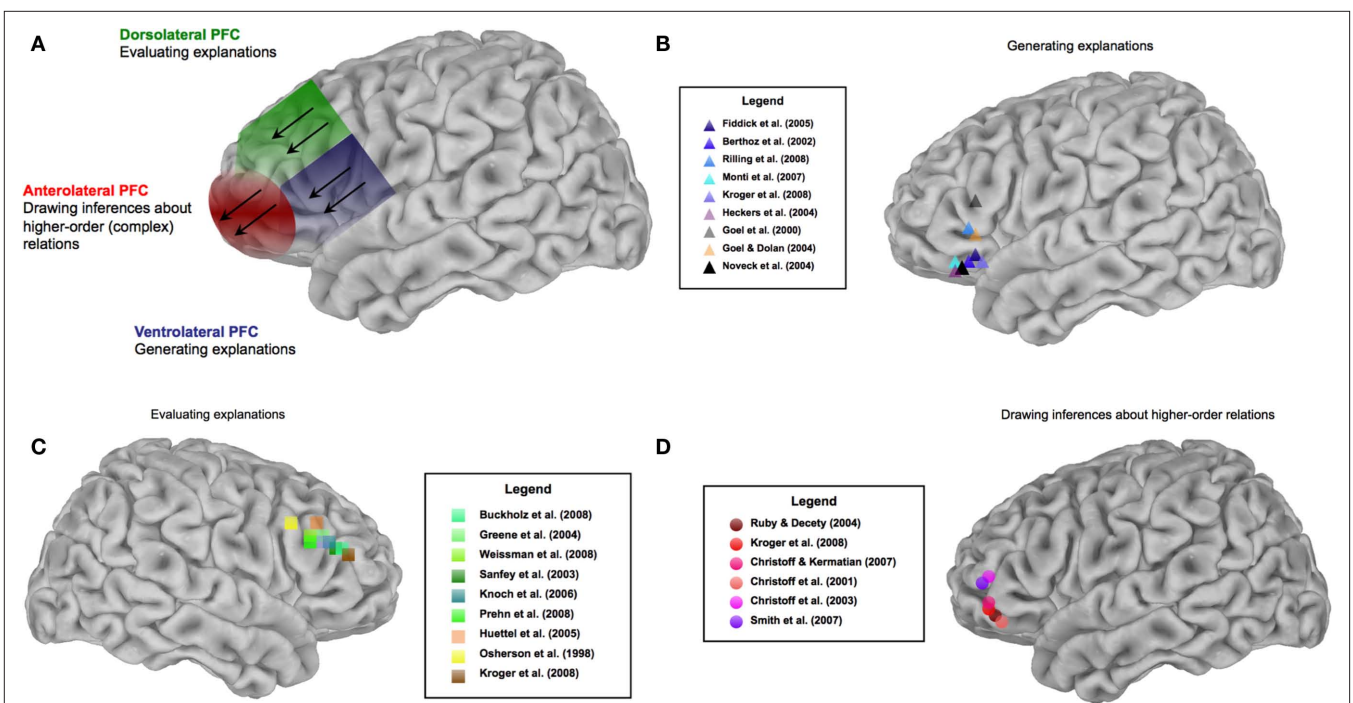


FIGURE 4 | Neural architecture of explanatory inference. (A) Summarizes the functional organization of the lateral PFC, and **(B–D)** illustrate supportive evidence.

example, demonstrated recruitment of left ventrolateral PFC (area 47) when participants detected violations of social norms stories representing obligatory and prohibited courses of action (e.g., the decision to "spit out food made by the host"). Similarly, Rilling et al. (2008) reported activation within left ventrolateral PFC (area 47) when participants detected the failure to cooperate in a Prisoner's dilemma game. Here the need for explanation is especially pressing, for we want to understand why the usual explanation or cause (as set forth in a social norm or rule) does not hold in a particular case. Does some "higher" or super-ordinate rule – or simply a conflicting coordinate rule – come into play in this particular situation? Or is there some other explanation altogether? In the food-spitting example, the explanation is presumably not a matter of a conflict among social rules, but a visceral response (perhaps to food that a particular person finds intensely nauseating). The research here does not establish whether ventrolateral PFC can "handle" the testing and comparison of alternative explanations (as opposed to the detection of a violation of some salient and usually explanatorily adequate social rule or norm). This needs to be investigated further, as part of making as clear as possible the conditions under which dorsolateral PFC, or even anterolateral PFC, must be recruited.

Meanwhile the decision neuroscience literature supports the proposed tripartite framework by suggesting a reason why, from a wider perspective, the ventrolateral PFC should be involved in the generation of basic social (and other) sorts of causal explanation. One very common type of explanation in everyday life and in scientific contexts depends essentially on deductive inference, as when some "covering law," or a behavioral rule or norm, combines with a statement of some particular facts to entail that some other fact must obtain, or that some specific action is obligatory, etc. For example, if books burn at Fahrenheit 451, and this book is heated to Fahrenheit 500, it will burn – that is, given an implicit ceterus paribus clause ("other things being equal"), or given the appropriate "enabling" or "background" conditions, such as that the book is not sopping wet, etc. Similarly, a great many everyday explanations in social and psychological, as well as physical, domains will explicitly or implicitly take the form of very simple deductions. For example, Why does one not spit out this bite of food, even if it tastes bad? Because one must not offend one's host. Phrased as a very simple deduction, we have: one must not do something that offends the host; spitting out the host's home-cooked food would offend the host; therefore one must not spit out the host's home-cooked food. This gives a general explanation for a more specific, but still general, rule. To get a prohibition against *this* person's spitting out *this* host's food on *this* occasion, we simply note that this person is a guest of that person, and that this food was home-cooked by the latter. This casts the explanation of why one doesn't spit out the host's food as a natural and simple deduction appealing to a general premise about what is impermissible. The explanation of why someone on a particular occasion did spit out the host's food would appeal to a different premise and deduction. (Perhaps, One involuntarily spits out food that is rotten and intensely nauseating; *this* person finds *this* food intensely nauseating.) Many everyday deductions are so intuitively obvious that they can be carried out automatically, but others will require conscious attention. The important point here is that a substantial body of neuroscientific research, if not a complete consensus in the field, strongly indicates that when common sources of variability are controlled (regarding the linguistic content, linguistic complexity, and deductive complexity of reasoning problems), simple deductions in general are supported by ventrolateral PFC. A recent series of experiments by Monti et al. (2007) controlled for these sources of variability and provided evidence that the left ventrolateral PFC (area 47) mediates representations of the logical structure of a deductive argument (e.g., If *P* or *Q*, then *Not-R*/*P*/Therefore, *Not-R*), supporting the generation of explanatory inferences within this region. Furthermore, a recent study by Kroger et al. (2008) controlled for the complexity and type of calculations that were performed and also observed activation within the left ventrolateral PFC (areas 44 and 45) for deductive reasoning (see also Heckers et al., 2004). Converging evidence is provided by Goel and colleagues (Goel et al., 2000; Goel and Dolan, 2004), who have consistently observed activation within the left ventrolateral PFC (areas 44 and 45) for deductive conclusions drawn from categorical syllogisms (e.g., All humans are mortal/Some animals are human/Therefore, some animals are mortal). Finally, Noveck et al. (2004) demonstrated recruitment of left ventrolateral PFC (area 47) for drawing deductive conclusions from conditional statements (e.g., If *P* then *Q*/*P*/Therefore, *Q*), consistent with the role of this region for generating explanatory inferences. In sum, this evidence indicates that generating a broad array of physical, social, and other explanations are supported by the ventrolateral PFC.

DORSOLATERAL PFC

Social neuroscience evidence demonstrates that the dorsolateral PFC (areas 46 and 9) represents computational mechanisms for evaluating explanations and causal scenarios based on some normative criterion, where that may involve testing an attribution of correctness to a causal scenario (as in thinking about pertinent causal interventions), or about an attribution of fairness or permissibility – or the opposite – concerning causal outcomes of alternative possible actions (**Figure 4C**). An early study by Sanfey et al. (2003) reported activity within the right dorsolateral PFC (area 46) when participants were presented with an unfair offer. Knoch et al. (2006) further demonstrated that deactivating this region with repetitive transcranial magnetic stimulation reduced participants' ability to reject unfair offers in an ultimatum game. In these cases the making of an offer is in itself in accordance with the norms or rules of the game, but a norm of fairness has been violated – perhaps egregiously, as when one is offered only 1 dollar out of the 10 to be divided. One's response goes beyond detection of the unfairness, since one must then decide what to do about accepting or rejecting the offer, and this will typically involve weighing the outcomes of these options in light of multiple goals – the goal of maximizing one's money in the game, or the social goal of asserting and maintaining one's status (Rilling et al., 2008), or the goal of defending a commitment to fairness and to punishing unfairness, and perhaps other "higher" or coordinate goals.

Buckholtz et al. (2008) observed activity within the right dorsolateral PFC (area 46) when participants evaluated the causal role of specific factors, assigning responsibility for crimes and making judgments about appropriate (e.g., equitable or fair) forms of punishment in a legal decision making task. The work of Greene et al. (2004) further suggests that this region is involved

in normative evaluations involving conflicting moral goals. These authors employed moral scenarios similar to the trolley problem (Thomson, 1976) and assessed trials in which participants acted in the interest of greater aggregate welfare at the expense of personal moral standards. This contrast revealed reliable activation within the right dorsolateral PFC (area 46), suggesting that this region is critical for normative evaluations involving conflicting moral goals. (For additional evidence for the role of this region in such evaluative processes, see Prehn et al., 2008; Weissman et al., 2008).

We suggest that the relevant common factor in all these various cases is a "second order" reflection upon an initial or "first order" causal scenario (whose formulation was supported by ventrolateral PFC), where reflection on that scenario is motivated by some need to "think again" or "think twice," in order to decide whether to attribute or withhold attribution of some normative property – e.g., moral permissibility, fairness, social utility – to some given causal scenario(s). We suggest that the main reason studies with widely differing orientations have found involvement of dorsolateral PFC is that the kinds of norms involved show a corresponding variety.

ANTERIOR PFC

Additional support for the general framework proposed here derives from the decision neuroscience literature, which demonstrates that progressively anterior subregions of the lateral PFC (area 10) are associated with higher-order processing requirements for thought and action (Koechlin and Summerfield, 2007; Badre, 2008; Botvinick, 2008). Ramnani and Owen (2004) reviewed contemporary research and theory investigating the cognitive functions of the anterior PFC, concluding that this region is central for integrating the outcomes of multiple cognitive operations, consistent with the predicted role of the anterior PFC for representing higher-order inferences that depend on the generation and evaluation of explanatory inferences (for representative findings, see Christoff et al., 2001, 2003; Christoff and Keramatian, 2007; Smith et al., 2007; Kroger et al., 2008; Barbey et al., 2011a).

This is to frame the issue once again in terms of levels of processing, with complexity increasing with anteriority. A large body of social neuroscience evidence supports this picture, although it may at first glance appear in some cases to invoke anterior PFC on the basic of content or subject matter (social and emotional) rather than level of complexity. It is well-established that anterior PFC (areas 10 and 11) – and the OFC more broadly – are central for explanatory social inference (**Figure 4D**). Studies of patients with lesions confined to the OFC have reported impairments in a wide range of social functions, including the regulation and control of social responses, the perception and integration of social cues, theory of mind and perspective taking (Rolls et al., 1994; Bechara et al., 2000; Ruby and Decety, 2004; LoPresti et al., 2008). Recent evidence from Stone et al. (2002) further demonstrates that patients with OFC damage show selective impairments in reasoning about normative social behavior and drawing explanatory social inferences. Bechara et al. (2000) observed profound deficits in the ability of OFC patients to represent and integrate social and emotional knowledge needed to generate mental state ascriptions and explanatory inferences about the causes of observed social behavior. Converging evidence is provided by LoPresti et al. (2008), who demonstrated that the left anterolateral PFC (area 11) mediates the representation and assessment of multiple social cue (i.e., emotional expression and personal identity), further suggesting that this region broadly supports the generation and evaluation of explanatory social inferences (for additional neuroscience evidence, see Moll et al., 2006).

It is, however, not merely the social nature of such inference that calls for involvement of anterior PFC, for as noted earlier, some social inferences (involving both adherence to and violations of social norms) are primarily supported by ventrolateral or dorsolateral PFC. What marks the involvement of anterior PFC, and explains the strong connection to social reasoning, is that typical "real life" social inference involves coordination and integration of multiple cognitive or computational tasks in the service of a larger goal. In particular, social reasoning that guides actual interactions with others will routinely integrate emotional and hedonic considerations into the evaluation of potential explanations of past actions and into deliberation about potential courses of future action. Thus anterior PFC draws on the resources of limbic areas (with which it is strongly connected via ventromedial PFC) and on hedonic representations in OFC (with which it is strongly interconnected; Kringelbach, 2005), as well as on the explanatory scenarios and the evaluative reflections and manipulations supported by ventrolateral and dorsolateral PFC.

CONCLUSION

INFERENTIAL ARCHITECTURE OF LATERAL PFC

We have reviewed converging lines of evidence to support the central role of the lateral PFC in explanatory inference, drawing upon recent theoretical developments in cognitive psychology and neuroscience bearing on the biology, evolution, ontogeny, and cognitive functions of this region. We have surveyed a broad range of evidence from social and decision neuroscience demonstrating that the lateral PFC mediates the generation and evaluation of explanations, with the ventrolateral PFC recruited when constructing explanatory inferences, engagement of the dorsolateral PFC for the evaluation of explanations, and the anterolateral PFC recruited when we utilize both of these processes – and additional ones with which they must be coordinated (such as the calculation of hedonic values of possible outcomes or actions; **Figure 4A**). The reviewed findings set the stage for new approaches to understanding the architecture of cognitive understanding suggesting that neural mechanisms within the lateral PFC detect and encode the higher-level structure of experience, representing causal relationships that guide the selection and control of modality-specific knowledge and provide the basis for explanatory inference.

Our findings raise further questions for future neuroscience research. One challenge is to address how neural mechanisms for generating and evaluating explanatory inferences are represented within dual process theories that distinguish between automatic versus controlled cognitive processes (e.g., Barbey and Sloman, 2007). Future research should further investigate the cognitive operations that are performed within the lateral PFC to support human inference. Does this region contain mechanisms that control the recruitment of representations stored in posterior cortices (e.g., Barbey et al., 2009a,b, 2011a,b, in press; Barbey and Grafman, in press, 2011), serve as an integrative hub for synthesizing modality-specific representations (e.g.,

Pessoa, 2008), or store unique forms of knowledge (e.g., Wood and Grafman, 2003)? Future research should also address the biological, developmental and evolutionary principles that account for the observed lateralization of processes for generating (left hemispheric) versus evaluating (right hemispheric) explanations (**Figure 4**). Research should further investigate the computational mechanisms underlying hierarchical cognitive representations with particular emphasis on the computational principles that enable the top-down control and coordination of modality-specific representations. Further research is needed also to define more precisely the functional and neural boundaries between ventrolateral, dorsolateral, and anterolateral PFC, especially given their intraPFC interconnections and their cooperative functioning in all but the most simple situations. The evidence surveyed here supports a broadly drawn tripartite framework for PFC involvement in causal explanation and understanding, but much of this evidence is itself indirect, and not designed precisely to investigate involvement of subregions of lateral PFC in explanatory reasoning. Finally, future research should investigate the larger role of the lateral PFC in the formation of human belief systems, investigating the neural mechanisms that integrate networks of causal knowledge to construct complex systems of value and belief, providing the foundations for explanatory inference and our sense of understanding.

REFERENCES

Amodio, D., and Frith, C. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn. Sci. (Regul. Ed.)* 12, 193–200.

Barbas, H., and Pandya, D. (1991). "Patterns of connections of the PFC in the rhesus monkey associated with cortical architecture," in *Frontal Lobe Function and Dysfunction*, eds H. S. Levin, H. M. Eisenberg, and A. L. Benton (New York: Oxford University Press), 35–58.

Barbey, A. K., and Barsalou, L. W. (2009). "Reasoning and problem solving: models," in *Encyclopedia of Neuroscience*, Vol. 8, ed. L. Squire (Oxford: Academic Press), 35–43.

Barbey, A. K., Krueger, F., and Grafman, J. (2009a). An evolutionarily adaptive neural architecture for social reasoning. *Trends Neurosci.* 32, 603–610.

Barbey, A. K., Krueger, F., and Grafman, J. (2009b). Structured event complexes in the prefrontal cortex support counterfactual representations for future planning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1291–1300.

Barbey, A. K., and Sloman, S. A. (2007). Base-rate respect: from ecological rationality to dual processes. *Behav. Brain Sci.* 30, 241–297.

Barbey, A. K., and Wolff, P. (2006). "Causal reasoning from forces," in *Proceedings of the 28 Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum), 2439.

Barbey, A. K., and Wolff, P. (2007). "Learning causal structure from reasoning," in *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum), 713–718.

Barbey, A. K., and Grafman, J. (in press). "The prefrontal cortex and goal-directed social behavior," in *The Handbook of Social Neuroscience*, eds J. Decety and J. Cacioppo (Oxford University Press).

Barbey, A. K., and Grafman, J. (2011). An integrative cognitive neuroscience theory for social reasoning and moral judgment. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 55–67.

Barbey, A. K., Koenigs, M., and Grafman, J. (2011a). Orbitofrontal contributions to human working memory. *Cereb. Cortex* 21, 789–795.

Barbey, A. K., Krueger, F., and Grafman, J. (2011b). "Architecture of counterfactual thought in the prefrontal cortex," in *Predictions in the Brain: Using our Past to Prepare for the Future*, ed. M. Bar (New York: Oxford University Press), 40–57.

Barbey, A. K., Solomon, J., Colom, R., Krueger, F., Forbes, C., and Grafman, J. (in press). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain*.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Bechara, A., Damasio, H., and Damasio, A. (2000). Emotion, decision making, and the orbitofrontal cortex. *Cereb. Cortex* 10, 1047–3211.

Berthoz, S., Armony, J. L., Blair, R. J. R., and Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* 125, 1696–1708.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci. (Regul. Ed.)* 12, 201–208.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., and Marois, R. (2008). The neural correlates of third-party punishment. *Neuron* 60, 930–940.

Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books, MIT Press.

Christoff, K., and Keramatian, K. (2007). "Abstraction of mental representations: theoretical considerations and neuroscientific evidence," in *The Neuroscience of Rule-Guided Behavior*, eds S. A. Bunge and J. D. Wallis (New York: Oxford University Press), 107–128.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., and Gabrieli, J. D. E. (2001). Rostrolateral PFC involvement in relational integration during reasoning. *Neuroimage* 14, 1136–1149.

Christoff, K., Ream, J. M., Geddes, L. P. T., and Gabrieli, J. D. E. (2003). Evaluating self-generated information: anterior prefrontal contributions to human cognition. *Behav. Neurosci.* 117, 1161–1168.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33,25–62.

D'Esposito, M. D., Postle, B. R., Ballard, D., and Lease, J. (1999). Maintenance versus manipulation of information held in working memory: an event-related fMRI study. *Brain Cogn.* 41, 66–86.

Einhorn, H. J., and Hogarth, R. M. (1986). Judging probable cause. *Psychol. Bull.* 99, 1–19.

Fiddick, L., Spampinato, M. V., and Grafman, J. (2005). Social contracts and precautions activate different neurological systems: an fMRI investigation of deontic reasoning. *Neuroimage* 28, 778–786.

Flechsig, P. (1901). Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. *Lancet* 2, 1027–1029.

Flechsig, P. (1920). *Anatomie des Menschlichen Gehirnsund Ruckenmarks auf Myelogenetischer Grundlage*. Leipzig: Thieme; New York: Basic Books.

Fuster, J. M. (1989). *The PFC*. New York: Raven.

Fuster, J. M. (1997). *The PFC–Anatomy Physiology, and Neuropsychology of the Frontal Lobe*. Philadelphia: Lippincott-Raven.

Gilbert, D. T. (1998). "Ordinary personology," in *The Handbook of Social Psychology*, 4th Edn, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York: Oxford University Press), 89–150.

Goel, V., Buchel, C., Frith, C., and Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *Neuroimage* 12, 504–514.

Goel, V., and Dolan, R. J. (2004). Differential involvement of left PFC in inductive and deductive reasoning. *Cognition* 93, B109–B121.

Goldman-Rakic, P. S. (1987). "Circuitry of primate PFC and regulation of behavior by representational memory," in *Handbook of Physiology: The Nervous System*, ed. F. Plum (Bethesda, MD: American Physiology Society), 373–417.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400.

Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., and Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus* 14, 153–162.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: John Wiley & Sons, Inc.

Keil, F. C. (2003). Folkscience: coarse interpretations of a complex reality. *Trends Cogn. Sci. (Regul. Ed.)* 7, 368–373.

Keil, F. C. (2006). Explanation and understanding. *Annu. Rev. Psychol.* 57, 227–254.

Keil, F. C., and Wilson, R. A. (2000). *Explanation and Cognition*. Cambridge, MA: MIT Press.

Kitcher, P., and Salmon, W. (eds). (1989). *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, Vol. 13. Minneapolis: University of Minnesota Press.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right PFC. *Science* 314, 829–832.

Koechlin, E., and Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends Cogn. Sci. (Regul. Ed.)* 11, 229–235.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: linking reward

to hedonic experience. *Nat. Rev. Neurosci.* 6, 691–702.
Kroger, J. K., Nystrom, L. E., Cohen, J. D., and Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Res.* 1243, 86–103.
Lombrozo, T. (2006). The structure and function of explanations. *Trends Cogn. Sci. (Regul. Ed.)* 10, 464–470.
Lombrozo, T., and Carey, S. (2006). Functional explanation and the function of explanation. *Cognition* 99, 167–204.
LoPresti, M. L., Schon, K., Tricarico, M. D., Swisher, J. D., Celone, K. A., and Stern, C. E. (2008). Working memory for social cues recruits orbitofrontal cortex and amygdala: a functional magnetic resonance imaging study of delayed matching to sample for emotional expressions. *J. Neurosci.* 28, 3718–3728.
Malle, B. F. (2004). *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. Cambridge, MA: MIT Press.
Miller, E. K. (2000). The PFC and cognitive control. *Nat. Rev. Neurosci.* 1, 59–65.
Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
Miller, E. K., and Phelps, E. (2010). Preface: current opinion in neurobiology – cognitive neuroscience 2010. *Curr. Opin. Neurobiol.* 20, 1–2.
Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). "Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628.
Monti, M. M., Osherson, D. N., Martinez, M. J., and Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: a language-independent distributed network. *Neuroimage* 37, 1005–1016.
Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychol. Rev.* 92, 289–316.
Noveck, I. A., Goel, V., and Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex* 40, 613–622.
Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59.
Pandya, D. N., and Barnes, C. L. (1987). "Architecture and connections of the frontal lobe," in *The Frontal Lobes Revisited*, ed. E. Perecman (New York: The IRBN Press), 41–72.
Patterson, R., and Barbey, A. K. (in press). "Causal simulation theory: an integrative cognitive neuroscience framework for causal reasoning," in *The Neural Representation of Belief Systems*, eds J. Grafman and F. Krueger (New York, NY: Psychological Press).
Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158.
Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 781–795.
Prehn, K., Wartenburger, I., Meriau, K., Scheibe, C., Goodenough, O. R., Villringer, A., van der Meer, E., and Heekeren, H. R. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Soc. Cogn. Affect. Neurosci.* 3, 33–46.
Ramnani, N., and Owen, A. M. (2004). Anterior PFC: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* 5, 184–194.
Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., Murdock, C. D., and Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 465, 1256–1266.
Rochat, P. (2009). *Others in Mind – Social Origins of Self-Consciousness*. New York, NY: Cambridge University Press.
Rolls, E. T., Hornak, J., Wade, D., and McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J. Neurol. Neurosurg. Psychiatr.* 57, 1518–1524.
Ruby, P., and Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Neurosci.* 16, 988–999.
Salmon, W. (1998). *Causality and Explanation*. New York: Oxford University Press.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2003). The neural basis of decision making in the ultimatum game. *Science* 300, 1755–1758.
Simmons, W. K., Reddish, M., Bellgowan, P. S. F., and Martin, A. (2010). The selectivity and functional connectivity of the anterior temporal lobes. *Cereb. Cortex* 20, 813–825.
Sloman, S. A., Barbey, A. K., and Hotaling, J. (2009). A causal model theory of the meaning of "cause," "enable," and "prevent." *Cogn. Sci.* 33, 21–50.
Smith, R., Keramatian, K., and Christoff, K. (2007). Localizing the rostrolateral PFC at the individual level. *Neuroimage* 36, 1387–1396.
Stone, V., Cosmides, L., Tooby, J., Kroll, N., and Knight, R. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11531–11536.
Tenenbaum, J. B., Griffiths, T., and Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci. (Regul. Ed.)* 10, 309–318.
Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *Monist* 59, 204–217.
Uttich, K., and Lombrozo, T. (2010). Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition* 116, 87–100.
Wager, T. D., Jonides, J., and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage* 22, 1679–1693.
Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274.
Weissman, D. H., Perkins, A. S., and Woldorff, M. G. (2008). Cognitive control in social situations: a role for the dorsolateral PFC. *Neuroimage* 40, 955–962
Wellman, H. M., and Liu, D. (2007). "Causal reasoning as informed by the early development of explanations," in *Causal Learning: Psychology, Philosophy, and Computation*, eds A. Gopnik and L. E. Shultz (New York, NY: Oxford University Press), 261–279.
Wolff, P., Barbey, A. K., and Hausknecht, M. (2010). For want of a nail: how absences cause events. *J. Exp. Psychol. Gen.* 2, 191–221.
Wood, J., and Grafman, J. (2003). Human prefrontal cortex: processing and representational perspective. *Nat. Rev. Neurosci.* 4, 139–147.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 August 2010; accepted: 28 June 2011; published online: 27 July 2011.
Citation: Barbey AK and Patterson R (2011) Architecture of explanatory inference in the human prefrontal cortex. Front. Psychology ***2****:162. doi: 10.3389/fpsyg.2011.00162*
This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.
Copyright © 2011 Barbey and Patterson. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 31 January 2011
doi: 10.3389/fpsyg.2011.00005

The mechanics of embodiment: a dialog on embodiment and computational modeling

Giovanni Pezzulo[1,2]*, Lawrence W. Barsalou[3], Angelo Cangelosi[4], Martin H. Fischer[5], Ken McRae[6] and Michael J. Spivey[7]

[1] *Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Roma, Italy*
[2] *Istituto di Linguistica Computazionale "Antonio Zampolli," Consiglio Nazionale delle Ricerche, Pisa, Italy*
[3] *Department of Psychology, Emory University, Atlanta, GA, USA*
[4] *School of Computing and Mathematics, University of Plymouth, Plymouth, UK*
[5] *School of Psychology, University of Dundee, Dundee, Scotland, UK*
[6] *Department of Psychology, University of Western Ontario, London, ON, Canada*
[7] *School of Social Sciences, Humanities and Arts, University of California, Merced, CA, USA*

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Rob Ellis, University of Plymouth, UK
Serge Thill, University of Skövde, Sweden

***Correspondence:**
Giovanni Pezzulo, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Via S. Martino della Battaglia, 00185, Roma, Italy.
e-mail: giovanni.pezzulo@istc.cnr.it

Embodied theories are increasingly challenging traditional views of cognition by arguing that conceptual representations that constitute our knowledge are grounded in sensory and motor experiences, and processed at this sensorimotor level, rather than being represented and processed abstractly in an amodal conceptual system. Given the established empirical foundation, and the relatively underspecified theories to date, many researchers are extremely interested in embodied cognition but are clamoring for more mechanistic implementations. What is needed at this stage is a push toward explicit computational models that implement sensorimotor grounding as intrinsic to cognitive processes. In this article, six authors from varying backgrounds and approaches address issues concerning the construction of embodied computational models, and illustrate what they view as the critical current and next steps toward mechanistic theories of embodiment. The first part has the form of a dialog between two fictional characters: *Ernest*, the "experimenter," and *Mary*, the "computational modeler." The dialog consists of an interactive sequence of questions, requests for clarification, challenges, and (tentative) answers, and touches the most important aspects of grounded theories that should inform computational modeling and, conversely, the impact that computational modeling could have on embodied theories. The second part of the article discusses the most important open challenges for embodied computational modeling.

Keywords: grounded cognition, embodiment, simulation, cognitive robotics, computational modeling, situated cognition, embodied cognition

INTRODUCTION

Embodied cognition is a theoretical stance which postulates that sensory and motor experiences are part and parcel of the conceptual representations that constitute our knowledge. This view has challenged the longstanding assumption that our knowledge is represented abstractly in an amodal conceptual network of formal logical symbols. There now exist a large number of interesting and intriguing demonstrations of embodied cognition. Examples include changes in perceptual experience or motor behavior as a result of semantic processing (Boulenger et al., 2006; Meteyard et al., 2008), as well as changes in categorization that reflect sensory and motor experiences (Smith, 2005; Ross et al., 2007). These demonstrations have received a great deal of attention in the literature, and have spurred many researchers to take an embodied approach in their own work. There are also a number of theoretical accounts of how embodied cognition might work (Clark, 1998; Lakoff and Johnson, 1999). One influential proposal is "perceptual symbols system" theory (Barsalou, 1999), according to which the retrieval of conceptual meaning involves a partial re-enactment of experiences during concept acquisition. However, to a large extent, embodied theories of cognition are still developing, particularly in terms of their computational implementations, as well as their specification with regard to moment-by-moment online processing.

Given the established empirical foundation, and the relatively underspecified theories to date, many researchers are extremely interested in embodied cognition but are clamoring for more mechanistic implementations. *What is needed at this stage is a push toward explicit computational models that implement sensorimotor grounding as intrinsic to cognitive processes.* With such models, theoretical descriptions can be fleshed out as explicit mechanisms, idiosyncratic patterns across experiments may be explained, and quantitative predictions for new experiments can be put forward.

In this article, six authors from varying backgrounds and approaches address issues concerning the construction of embodied computational models, and illustrate what they view as the critical current and next steps toward mechanistic theories of embodiment. We propose the use of *cognitive robotics* to implement embodiment, and discuss the main prerequisites for a fruitful cross-fertilization between empirical and robotics research. Cognitive robotics is a broad research area, whose central aim is realizing complete robotic architectures that, on the one hand, include principles and constraints derived from animal and human cognition and, on the other hand, learn to operate autonomously in complex, open-ended scenarios (possibly interacting with humans) and have realistic embodiment, sensors, and effectors.

The relationship between theories of grounded cognition and cognitive robotics is twofold. On the one hand, theories and findings in research on grounded cognition imply that robot design should take into account the fact that a robot's cognitive capacities should not be independent of its design and the modalities it uses for interacting with the external environment. This poses opportunities and challenges for robotics research (Pfeifer and Scheier, 1999). On the other hand, computational modeling in cognitive robotics can contribute to the development of better theories of embodied cognition by clarifying and testing some of its critical aspects, such as the extent to which embodied phenomena exert a causal influence on cognitive processing, thereby suggesting new avenues of research. Note that we are interested in computational models of embodied cognition in general, and not only for modeling human cognition, although we often use human cognitive abilities as examples in this article.

The article is structured as follows. We begin by clarifying the usage of some terms. The second section takes the form of a dialog between two fictional characters: Ernest, the "experimenter," and Mary, the "computational modeler." The dialog consists of an interactive sequence of questions, requests for clarification, challenges, and (tentative) answers. The dialog touches on the most important aspects of grounded theories that should inform computational modeling and, conversely, the impact that computational modeling could have on grounded theories. In the final section, we discuss the most important open challenges for embodied computational modeling, and suggest a roadmap for future research.

The use of terms such as "grounded" and "situated" is somewhat arbitrary, and these terms are used often interchangeably with "embodied." Because of this issue in the current literature, we introduce some definitions at the outset of this article (cf. Myachykov et al., 2009; Fischer and Shaki, 2011). Together with these definitions, we also provide examples that specifically pertain to numerical cognition because this area of knowledge representation has traditionally been considered as a domain par excellence for abstract and amodal concepts, a view we wish to challenge.

GROUNDING

At the most general level, cognition has a physical foundation and it is, first and foremost, *grounded* in the physical properties of the world, such as the presence of gravity and celestial light sources, and constrained by physical principles (at least until we have evidence of life and cognition in a virtual reality). One example of grounding in the domain of numerical cognition is the fact that we associate smaller numbers with lower space and larger numbers with upper space (Ito and Hatta, 2004; Schwarz and Keus, 2004). This association is presumably universal because it reflects the physical necessity that the aggregation of more objects leads to larger piles. The recognition of the physical foundation of cognition has led researchers to challenge traditional theories of cognitive science and AI, in which cognitive operations were conceived as unconstrained manipulations of arbitrary and amodal symbols. The philosophical debate on how concepts and ideas have any meaning and are linked to their referents was revitalized by Searle's (1980) Chinese room argument and by Harnad's (1990) paper on symbol grounding, in which he argued that language-like symbols traditionally used in AI are meaningless because they lack grounding and reference to the external world. Harnad argued that the solution to this problem lies in the grounding of symbols in sensorimotor states; in this way, internal manipulations are constrained by the same laws that govern sensory and motor processes. Successively, *grounded cognition* has become the label for a methodological approach to the study of cognition that sees it as "grounded in multiple ways, including simulations, situated action, and, on occasion, bodily states" (Barsalou, 2008a, p. 619). As such, grounded cognition is different from, and wider than, embodied or situated cognition, because on occasion "cognition can indeed proceed independently of the specific body that encoded the sensorimotor experience (Barsalou, 2008a)." Rather, embodied and situated effects on representation and cognition can be conceptualized as a cascade.

EMBODIMENT

On top of conceptual grounding, embodied representations are shaped by sensorimotor interactions, and consequently by the physical constraints of the individual's body. Thus, embodiment is a consequence of the filtering properties of our sensory and motor systems, but this input is already structured and shaped in accord with physical principles, and these provide the grounding of cognition. For an example of embodiment in the domain of numerical cognition, consider the ubiquitous fact that small numbers are responded to faster with the left hand and larger numbers with the right hand – the spatial–numerical association of response codes (SNARC) effect. This effect is weaker in people who start counting on the fingers of their right hand compared to those who start counting on their left hand (Fischer, 2008; Lindemann et al., inpress), presumably because right-starters associate small numbers with their right side. This shows that the systematic use of one's body influences the cognitive representation of numbers. Note that "embodied cognition" is often used metonymically so as to refer to "grounded cognition;" the former label is much more popular than the latter, and there is nothing wrong with its use providing that one keeps in mind that its literal meaning is restrictive.

SITUATEDNESS

Finally, situated cognition refers to the context dependence of cognitive processing and reflects the possibility that embodied signatures in our performance are context-specific and can be modified through experience. This can be a simple change of posture, as in the crossing of arms that reveals the dominance of allocentric over egocentric spatial coding in the Simon effect (Wallace, 1971). The SNARC effect offers two illustrations of this idea. First, a given number can be associated with either left or right space depending on the range of other numbers in the stimulus set (Dehaene et al., 1993; Fias et al., 1996). Second, turning one's head alternatingly to the left and right while generating random numbers leads to a bias, such that left turns evoke more smaller numbers than do right turns (Loetscher et al., 2008). Both examples illustrate how the specific situation modulates the grounded and embodied representation of numbers (see also Fischer et al., 2009, 2010).

Although most contemporary theories of grounded cognition focus only on a subset of the phenomena that we have described here, future theories should tell a coherent story of how all of the relevant grounded, embodied, and situated phenomena constitute and constrain cognition.

EMBODIED COGNITION AND COMPUTATIONAL MODELING: A DISCUSSION

TOPIC 1. WHAT QUALIFIES AS AN "EMBODIED" COMPUTATIONAL MODEL AND WHAT ARE ITS MOST IMPORTANT REQUIREMENTS?

Recent research has shown that grounded, embodied, and situated phenomena have a great impact on cognitive processing at all levels rather than being confined to the sensory and motor peripheries. In particular, beyond basic response production, intelligent action coupled with perception epitomizes embodied approaches. This poses significant challenges for computational models in all traditions (symbolic, connectionist, etc.). The first and foremost challenge is that cognition cannot be studied as a module independent from other modules (sensory and motor), as suggested by the "cognitive sandwich" metaphor. Instead, cognition is deeply interrelated with sensorimotor action and affect. Evidence indicates that even complex cognitive operations such as reasoning and language rely on and recruit perceptual and motor brain areas, and that imposing interference in these sensorimotor areas significantly impairs (or enhances) a person's ability to execute cognitive tasks. Embodiment plausibly exerts its influence also by shaping development; thus, complex cognitive operations are learned based on simpler sensorimotor skills, which provide a ready neural and functional substrate. This implies that cognitive processes cannot be divorced from the sensorimotor processes that provided the scaffold for their development.

Consider a few examples of embodiment signatures in cognition. Spatial associations are frequently used to ground abstract conceptual knowledge, such as numerical magnitudes. This has been documented extensively in the SNARC effect (for a recent meta-analytic review, see Wood et al., 2008). Briefly, smaller magnitudes are associated with left space and larger magnitudes with right space, but this mapping is sensitive to contextual and cultural factors. More recently, the manipulation of magnitudes (arithmetic) has been shown to be mapped onto space, with addition inducing right biases and subtraction inducing left biases (Pinhas and Fischer, 2008; Knops et al., 2009). Another significant example of embodiment signatures in cognition is attention deployment, which plays a central role in forming concepts and directing reasoning within a grounded cognition framework (Grant and Spivey, 2003). In line with the embodied cognition approach, bodily constraints impose corresponding constraints on cognitive functioning and vice versa. Consider first how body postures affect attentional processing. With regard to one's own postures, attention cannot be cued more laterally if the observer's eyes are already at their biomechanical limit (Craighero et al., 2004). Similarly, preshaping one's hand influences the selection of large or small objects in a visual search task (Symes et al., 2008), and planning to either point or grasp modulates the space- and object-based deployment of attention (Fischer and Hoellen, 2004) as well as the selection of object features (Bekkering and Neggers, 2002). With regard to perceiving other people's postures, a large body of work on joint attention has discovered behavioral and neural evidence of rapid and automatic ability to process another person's gaze direction (Frischen et al., 2007), head orientation (Hietanen, 2002), and hand aperture (Fischer et al., 2008) to deploy one's own attention to a likely action goal. Body postures also affect one's higher-level cognition. For example, adopting a particular posture will improve one's recollection of events that involved similar postures, such as reclining on a chair and the associated experience of a dental visit (Dijkstra et al., 2008).

This body of work seems to confirm the tight coupling between sensory and motor maps on the one hand and conceptual processing on the other hand, as postulated by the embodied cognitive approach. It does, however, also raise an architectural challenge for computational modeling because it seems to require persistent cross-talk between domain-specific systems to account for the wide range of embodiment effects on performance. In terms of computational modeling, the main implication of this view is that the specific way in which robots act, perceive their external environment, and strive to survive and obtain reward, must have a significant impact on their cognitive representations and skills, and on how they develop. Indeed, this insight has close relations with a limitation that has been widely recognized in AI research, namely that cognitive processes were implemented by manipulating abstract symbols that were not "grounded" in the external world, and were unrelated to the robot's action repertoire and perception (Harnad, 1990).

Research in grounded cognition makes an even stronger case for the influence of embodiment on cognition. *Not only should representations be grounded, but their processing should essentially be fully embodied as well*, such that there is no central processing independent of sensorimotor processes and/or affective experience. For instance, if we consider again the examples mentioned for spatial reasoning and attention, this leads to difficult issues for modelers, such as how spatial relations could be transferred to other domains (e.g., the temporal domain), or how spatial associations of abstract concepts could be simulated. The grounded perspective opens new avenues in robotics research, although a number of open research issues remain to be addressed.

One general set of issues relative to the realization of embodied cognitive models concerns the computational architecture, aside from whether it takes the form of neural networks, Bayesian approaches, production systems, classic AI architectures, or another form. Ernest, an experimenter, and Mary, a computational modeler, discuss these topics.

Requirement 1. Modal versus amodal representations

Mary: What are the most important constraints that grounded cognition pose for computational modeling and robotics? And, conversely, what are the essential features that computational models should have for them to be recognized as being "grounded cognitive models"?

Ernest: Perhaps the first and foremost attribute of a grounded computational model is the implementation of cognitive processes (e.g., memory, reasoning, and language understanding) as depending on *modal* representations and associated mechanisms for their processing [e.g., Barsalou's (1999) simulators] rather than on amodal representations, transductions, and abstract rule systems.

Mary: This is a key departure point from most models in AI, which use *amodal* representations. However, don't you think that modal representations, such as for instance visual representations, are too impoverished to support cognition? Take for example the visual representation of an apple. It could be sufficient to support decisions on how to grasp the apple, but maybe not for processing the word "apple" or for reasoning about apple market prices.

Ernest: Well, there is a big gap between the visual representation of an apple and the kind of reasoning you have in mind. However, note that embodied cognition does not claim that the brain is a recording system, or that objects are represented as "pictures" in the brain. Rather, the key idea is that the format of representations is still modal when they are manipulated in reasoning, memory, and linguistic tasks. For instance, it is now increasingly recognized that linguistic objects are stored as sensorimotor codes (Pulvermüller, 2005). Since language processing recruits the same circuits as action representation and planning (bidirectional) interference effects occur that have been observed experimentally. There is increasing evidence that the modal nature of representations creates "interferences" with memory and reasoning tasks as well (see Barsalou, 2008a for a review). Note also that to implement a truly embodied cognitive system, *multiple modalities* are essential. In addition to sensory and motor modalities, internal modalities, including affect, motivation, and reward, are essential from the embodied perspective.

Mary: Then, one can ask: what modalities are critical to include in a model of grounded cognition?

Ernest: Well, the answer depends, I suppose, on the empirical phenomena you want to model and on the specific embodiment of the robots you use. In general, however, it seems obvious that an embodied model of human cognition has to include the perceptual modalities, including, if possible, various aspects of vision, hearing, taste, smell, and touch, as all of them are relevant to human cognition. Note that the different modalities could be organized differently in the brain, and could contribute differently to action control. Excluding task-specific weighting of relevance, it seems important to implement visual dominance over the other modalities because it has frequently been demonstrated in sensory conflict paradigms, as in Calvert et al. (2004), for example. Embodied models of non-human cognition should also consider that non-human animals use multiple modalities as well.

Mary: Implementing multimodality in robots is not simple. There are nowadays many robotic platforms in the market; if we also consider that some of them can be customized, this offers (at least in principle) an ample choice of modalities to be included. However, simply including more modalities does not guarantee better performance, because an important issue concerns the associations among them. *How should representations of different modalities be associated into patterns?*

Ernest: In principle, this could be done directly, via connections from one modality to another, or indirectly, via association areas that function as hubs, linking modalities.

Mary: In computational terms, a simple, but certainly not unique, method for implementing direct connections from one modality to another is designing robot controllers composed of multiple, interlinked modal "maps" [e.g., Kohonen's (2001) self-organizing maps], such as for instance motor maps, visual maps, and auditory maps, and see how they become related so as to support cognitive processing. For instance, a robot controller composed of multiple maps can learn coordinated motor programs such as looking at objects, pointing at objects, hearing their sound, etc., and the maps could develop strong associations between object-specific (motor, perceptual) features. Association areas can be designed as well within the same framework, by introducing maps that group the outputs of multiple maps. However, I don't see any principled criterion for deciding what association areas should be included and how should they be organized.

Ernest: This problem is complicated because modalities seem to have a *hierarchical* rather than a flat structure. However, we still have an incomplete knowledge of the hierarchical structure of feature areas and association areas, and the connectivity patterns among them. Also important are the unique areas associated with *bottom-up activation versus top-down simulation*, along with shared areas. Regarding perception, most researchers believe that at least one association area or convergence zone is required for integrating information from different modalities (Damasio, 1989; Simmons and Barsalou, 2003). These ideas offer a starting point for computational modeling, which could be useful for answering many open questions. One question is whether a single area exists in which all types of information are integrated, as proposed by Patterson et al. (2007), or whether there are many, possibly hierarchically organized convergence zones. This question seems ripe for modeling because we have only begun to explore the consequences of various configurations in experimental work. For example, is more than one association area required to account for patterns of conceptual impairment?

Mary: I believe that the computational methodology could help in this regard by assessing the computational advantages that association areas and hierarchies provide, and assessing their costs (in computational terms). In addition, by using computational modeling it becomes possible to investigate the factors that regulate the patterns of connectivity among modalities. For instance, it has been proposed that "far-" senses, such as vision, are often predictive of "near-" senses, such as touch (Verschure and Coolen, 1991), and this would constrain associations and hierarchies. A third important issue concerns investigating the relations between sensory and motor codes. Indeed, in complete robot architectures, not only sensory information, but also motor information, in terms of both planning and execution of movements, would be required, and an important issue for computational modeling is how to integrate them.

Ernest: In cognitive neuroscience and psychology, there is a wide interest on how sensory and motor information is integrated in the brain. Traditional theories of planning tend to see sensory, cognitive, and motor codes as different; this implies a transduction from (modal) sensory to (amodal) cognitive codes, in which the latter guides cognitive processing and activates motor codes (Newell and Simon, 1972). By contrast, ideomotor theories of action provide support for the *common coding* of action and perception (Prinz, 1997; Hommel et al., 2001), which requires no transduction and could provide a better substrate for computational modeling of embodied phenomena. Similar ideas have rapidly gained importance in (social) cognitive neuroscience, due to the discovery of multimodal neurons, such as mirror neurons (Rizzolatti and Craighero, 2004). Many researchers believe that both perception and action rely on a principle of feature binding, whose anatomical and functional aspects are only partly understood. At the functional level, Körding and Wolpert (2004) proposed that the central nervous system combines, in a nearly optional fashion, multiple sources of information, such as visual, proprioceptive and predicted sensory states, to overcome sensory and motor noise. One advantage here

is that neuroscience research is advancing rapidly and continues to provide useful information on how to represent and integrate these types of information.

Mary: This is indeed a very relevant point for computational modeling of embodiment. In traditional AI and vision research, internal representations are typically defined as functions of the input, and perceptual learning is formulated as the problem of extracting useful "features" from passively received perceptual (mainly visual) stimuli. Even in robotics, most studies (e.g., using reinforcement learning) use fixed sets of representations, which define for instance the current location and pose of the robot. Not only are these representations predefined (by the programmer), but also they are "generic," or not specifically tied to the motor repertoire of the robot. Conversely, researchers in cognitive robotics increasingly recognize that perception and action form a continuum, and perceptual learning cannot disregard what is behaviorally relevant for the robot (see, e.g., Weiller et al., 2010); or, in other words, that representations should be shaped by the motor repertoire of the robot rather than being generic descriptions of the external world. This has led to a renaissance of the construct of object *affordances* (Gibson, 1979). Yet another formulation of the same idea is that learning is not a passive process, but is governed by the properties of the learning system. Because robots can actively control their inputs by means of their motor commands, their perceptual representations become dependent on motor skills and imbued with motor information as they explore their environment.

Ernest: Robotics seems to be a good starting place for investigating the mutual relations among the sensory modalities, and between sensory and motor modalities. Going even further in this direction, internal representations should be imbued with value representation, or information from the "internal" modalities, such as affect, interoception, motivation, and reward.

Mary: In robotic scenarios, adding internal motivations to robot architectures offers a natural way to link actions and value. The study of motivational systems is recently re-gaining importance in robotic settings (see, e.g., Fellous and Arbib, 2006).

Ernest: This stream of research could have additional advantages. Indeed, not only is value information essential for indicating significance to a robot's actions, but current research in embodied cognition is revealing that affect, emotion, and the internal states that result from them could play a key role in shaping high-level cognition in a fully embodied system. One possibility is that understanding and producing abstract concepts, such as love or fear, depends on knowledge acquired from introspection (Craig, 2002; Barsalou and Wiemer-Hastings, 2005).

Requirement 2. From sensorimotor experience to cognitive skills: abstraction and abstract thought on the top of a modal system

Mary: This leads us to another important topic. Even if we understand how the modalities interact, from a computational viewpoint, we do not know how modal representations can support cognitive processing, nor the wide range of cognitive tasks embodied theories can potentially tackle. Is there any theory of how grounded representations could do that?

Ernest: One of the most "organic" proposals so far is Barsalou's (1999) idea of simulation. A simulation is "the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008a, pp. 618–619). The ability to map simulations onto sensorimotor states by using overlapping systems is essential, and permits implementing the top-down construal that characterizes all cognitive activity. A big challenge for computational modeling is realizing simulation mechanisms. This permits testing whether and how they can support cognitive operations ranging from memory tasks to categorization, action planning and symbolic operations, and producing both abstractions and exemplars, both of which are central to cognition.

Mary: Then, it seems to me that an embodied cognition picture of cognitive processing could be the following (see **Figure 1**). First, grounded models are formed based on situated interaction of the robot with its environment (including other robots or humans). These symbols are multimodal and link perceptual, motor and valence information related to the same learning episodes. Second, cognitive processing is performed through the re-enactment of grounded symbols: a process that is called "situated simulation." During situated simulation, what becomes active anew includes not only the relevant episode-specific representations, but also the associated bodily resources and sensorimotor strategies, and so cognition operates under the same constraints and situatedness of action.

Ernest: This seems to me quite an appropriate blueprint. How different is it from standard computational models?

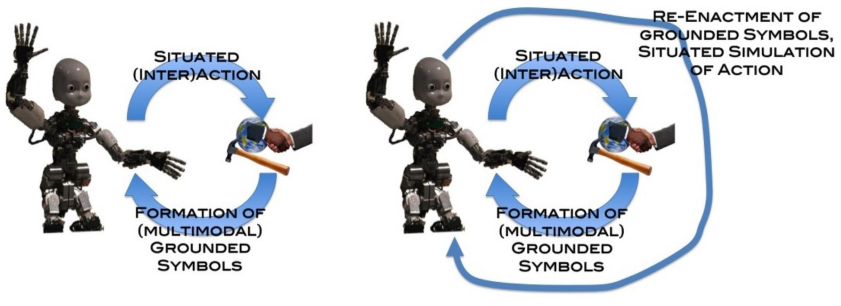


FIGURE 1 | A grounded cognition perspective on how grounded (modal) symbols are firstly formed based on situated interactions with the external environment, and therefore re-enacted as situated simulations that afford higher-level cognitive processing (having the same characteristics and constraints as embodied and situated action).

Mary: Well, some existing systems, for instance connectionist architectures, already encode sensorimotor patterns in some form. What is more novel is how sensorimotor patterns are reused, and the idea that grounded symbols can be *re-enacted* so as to produce grounded cognitive processes. From a computational viewpoint, one interesting aspect of this theory is that a single mechanism, simulation, could underlie a wide range of cognitive phenomena. However, despite the potentialities of this idea, it raises many feasibility issues, such as how quick and accurate simulations should be to be really useful, how many computational resources are required to run simulations in real time, and how simulations of different aspects of the same situation can merge. Feasibility issues are of primary importance for computational modeling; if the idea of situated simulation successfully permeates cognitive robotics, a lot of effort will be required to bridge the gap between its current conceptualization and the full specification of efficient simulated mechanisms. In addition, we still have an incomplete picture of how simulation works. Even if we have a complete architecture provided with multiple modalities, it is still unclear what should be re-enacted that constitutes a simulation.

Ernest: As a first approximation, simulations could be automatic processes that simply re-enact the content of previously stored perceptual symbols, although there could be deliberate uses of simulations as well.

Mary: This simply shifts the problem from the re-enactment to the formation of simulators, and more in general to how the different modalities contribute to specific cognitive tasks. Take categorization as an example. It is difficult to see how individual concepts are extracted from rich multimodal experience, and which mechanisms are responsible for their formation. How should these mechanisms work in practice?

Ernest: Psychologists and neuroscientists have often focused on pattern association in associative areas, which could encode increasingly "abstract" concepts. Nevertheless, how (and which) patterns are associated and classified is only partly understood. One recent idea (Barsalou, 1999) is that categorical representations might emerge when attention is focused repeatedly on the same kind of thing in the world, by utilizing associative mechanisms among modalities, which, in turn, might permit re-enactment and simulation when needed. To the best of my knowledge, this mechanism has never been tested in computational models and would certainly be a valuable contribution to embodied cognition research because it would represent the development of alternative computational mechanisms.

Mary: This is a very good example of what grounded models can offer concerning longstanding questions, like the acquisition of abstract concepts and abstract thought. Also, your example highlights the "style" of embodied cognitive models compared to traditional computational modeling. What seems to me to be crucial here is that the acquisition of representations and skills is itself an embodied and situated process, is grounded in the sensorimotor abilities and bodily resources of the learner, and thus is modulated by the same environmental and cultural circumstances.

Ernest: You are right. Not only should grounded models refrain from using amodal symbols, but also from modeling the acquisition and use of concepts and reasoning skills as abstract processes, or processes that are not subject to the same constraints and laws that govern sensorimotor skills, as has been done in traditional computational modeling of psychological phenomena. This is not to say that all concepts originate in experience, given that there could be nativist contributions as well, only that the empirical contributions to concepts reflect the constraints of actual experience.

Mary: This is a good starting point for a research program in embodied computational modeling. However, computer scientists also have to deal with the soundness and feasibility of their approach; and unfortunately, from a computational viewpoint, the powers and limitations of simulations are still unclear. For instance, similar to traditional theories of conceptual representations, simulations could be too rigid to account for the variety of experience. If simulations and concepts collect (or perhaps average on) experience, how do they adapt to novel simulations and how do they get framed around background situations?

Ernest: One possible answer to this question is that simulations are not expected to replay all collected information; instead, they merge with perception to form completely situated experiences and can re-enact different content depending on current goals, sensorimotor, social, and affective states, all of which make (only) some content relevant. In a series of articles, Barsalou (1999, 2008a, 2009) presented various arguments and data indicating that simulators can be considered dynamical systems that produce simulations in a context-specific way that changes continually with experience.

Mary: A second possible limitation of simulations and re-enactment is that they seem to be prima facie related to the here-and-now. Therefore, it is difficult to see how they relate to something outside the present situation.

Ernest: According to Barsalou (1999), simulations re-enact perceptual experience, and the same neural codes implied in the initial experience with the actual objects. However, this does not preclude running simulations that represent *future* states of affairs. The re-enactment notion should not be confused with passive recollection of past states. Many recent studies have highlighted the importance of simulating *future* states of affairs to coordinate with the world as it will be, not simply with how it is now, and have argued that preparing for future action, not just recalling it, could be the main adaptive advantage of re-enactment, simulation, and memory systems (Glenberg, 1997; Schacter et al., 2007; Pezzulo, 2008; Bar, 2009; Barsalou, 2009). The richness and multimodality of simulations is useful to produce predictions across domains. A study by Altmann and Kamide (1999) shows that subjects started to look at edible objects more than inedible objects when listening to "the boy will *eat*" but not "the boy will *move*," indicating that people can combine linguistic and non-linguistic cues to generate predictions. (Note that here the terminology is somewhat ambiguous because sometimes "simulation" is used as a synonym of re-enactment and sometimes as a synonym of long-term prediction.). Studies that involve imagination of *future* states of affairs also report that (visual and motor) simulations and imagery share neural circuits with actual perceptions and actions (Kosslyn, 1994; Jeannerod, 1995) and are subject to the same timing and general constraints. For instance, visual images and perception have the same metric spatial information and are subject to illusions in the same way. Performed and imagined actions respect Fitts's law of motor control and its occasional violations (Eskenazi et al., 2009; Radulescu et al., 2010). It has been proposed that detachment from

the here-and-now of experience, which, for example, is required for planning actions in the future, could be realized as a sophistication of the predictive and prospective abilities required in motor control and could recruit the same brain areas, rather than being processed in segregated brain areas with abstract representations (Pezzulo and Castelfranchi, 2007, 2009). This view is supported by the close relationship between the neural circuits that underlie motor imagery and motor preparation (Cisek and Kalaska, 2004).

Mary: Still, we have been talking about non-present circumstances related to the senses and the modalities (future or past states of affairs). It is even harder to imagine how simulations might relate to non-observable circumstances, such as, for example, "transcendental" concepts like space and time. How could space and time be implemented in a grounded system? And how would these implementations allow the system to run simulations of non-present experience with some fidelity to the representation of space and time in actual perception and action? A second issue concerns abstract concepts, including how they can support a full-fledged symbolic system.

Ernest: Well, these are all difficult questions, and I believe that cross-fertilization between empirical research and computational modeling could contribute to clarifying them. One tenet of grounded cognition is that *all* processes are situated and use modality-specific information rather than being processed in an abstract, amodal, logical space. This means that the representations of space and time in grounded systems, in all their manifestations, draw significantly on the processing of space and time in actual experience. Perception, cognition, and action must be coupled in space and time, and simulations of non-present situations must be implemented in space and time, perhaps using overlapping systems. Internal simulations do not escape this rule; so if abstract concepts and symbolic manipulations are grounded in introspective simulations, they should be sensitive to external space and time, too, and retain sensorimotor aspects. Although realizing how to implement a full-fledged symbolic system is a complex issue, some ideas useful for modelers have been presented. For example, Barsalou (2003) argued that selective attention and categorical memory integration are essential for creating a symbolic system. Once these functions are present, symbolic capabilities can be built upon them, including type-token propositions, predication, categorical inference, conceptual relations, argument binding, productivity, and conceptual combination.

Requirement 3. Realistic linkage of cognitive processes with the body, the sensory and motor surfaces, the environments in which cognition happens, and brain dynamics

Mary: These ideas at least provide some initial direction for creating novel grounded architectures and models. However, we have mostly discussed the modality of representations: do you think that there are additional factors that embodied models should include?

Ernest: According to grounded cognition theories, not only the modalities, but also sensorimotor skills and bodily resources influence cognitive processing, even in abstract domains. For instance, visuomotor strategies and eye movements are reused for abstract thinking; finger movements can be employed for counting; spatial navigation skills can be reused for reasoning in the temporal domain; motor planning processes can be re-enacted for imagining future events, understanding actions executed by others, or as an aid to memory. In these cases, expressions such as "taking a perspective on a problem" or "putting oneself into another's shoes" or "grasping a concept" have to be taken more literally than normally assumed. Overall, due to their coherent learning processes and to re-enactment, grounded cognitive processes have the same powers, but also the same constraints, as bodily actions.

Mary: I wonder how your examples could be treated in a sound cognitive robotics design methodology. Ideally, rather than focusing on the abstract nature of cognitive problems, modelers should ask first what sensorimotor processes could support them in embodied agents. An emblematic example is temporal reasoning via spatial skills: a somewhat novel way to approach this topic could be learning spatial navigation first, and then reasoning in the temporal domain on top of the spatial representations, by re-enacting similar bodily processes. Clearly, this should be done within an embodied and situated research program rather than generically via computational modeling. I am increasingly more inclined to propose *cognitive robotics* as the primary methodology with which this and other grounded phenomena could be studied, as it emphasizes the importance of sensorimotor processes, situated action, and the role of the body. Still, it is unclear to me how realistic the bodies of robots should be. What kind of embodiment is necessary to study grounded phenomena? Is *the specific embodiment* of our models really important for embodied phenomena to happen? Can we study embodied cognition in agents that do not have their own "body" (as in general cognitive agents) or that are just computer simulations of robots' bodies?

Ernest: Well, we know that most embodied effects are not only due to the way task-related information is represented, re-enacted, and processed, but are also due to the fact that *the body* is the medium of all cognitive operations, whether they are as simple as situated action or as complex as reasoning. Contrary to traditional cognitive theory, researchers in embodied cognition (e.g., Lakoff and Johnson, 1999) have argued that the body shapes cognition during development and continues to exert an influence at all stages of cognitive processing. Embodiment could have subtle and unexpected effects on cognitive processing. For instance, Glenberg and Kaschak (2002) showed that the action system influences sentence comprehension [the action–sentence compatibility (ACE) effect] and that subjects needed more time to understand sentences when the action required to signal comprehension was in the opposite direction than the target sentence (e.g., upward direction when the sentence referred to downward actions). From all these considerations, I would say that being realistic about embodiment is a must, at least in certain domains. Then, I see the point of your last question: a paradoxical consequence of taking embodiment claims literally is that cognitive robots could *not* be good models of human cognition because their bodies are too different from the body of humans, and have different computational, sensory, memory, and motor resources. What are the currently available platforms in cognitive robotics, and how well embodied are they?

Mary: Within embodiment research, cognitive models can be based on a variety of tools and platforms ranging from general cognitive agent systems (including multi-agent systems), to robot simulation models, up to physical robot platforms in cognitive robotics.

- *Cognitive agents.* Through these models we can typically simulate only selected features of the agent's embodied system. For example an agent can have a retina-like visual system, and a motor control system to navigate the environment. This is the case for models of environment navigation as in foraging tasks. Moreover, these models are suitable for multi-agent simulation where we also investigate social and interaction aspects of cognitive processing. For example, Cangelosi (2001) implemented a multi-agent model of the evolution of communication. In it, a population of simulated abstract agents have to perform a foraging task. They can perceive the visual properties of objects ("mushrooms") that determine their category of edible and inedible objects. Moreover, agents have a motor system to navigate the 2D world and approach/avoid foods. The perceptual and motor systems are implemented through a connectionist network, which also includes information relevant to the agent's basic drives, such as hunger. Through this essential modeling of the agent's sensorimotor system, it has been possible to investigate the symbol grounding problem in language learning and evolution (see also Cangelosi, 2010).
- *Simulated robotic agents.* These include realistic models of an existing robot, such as simulation models of the iCub humanoid platform (Tikhanoff et al., 2008), which is an open source robotic platform specifically developed for cognitive research, and of mobile robots such as khepera (Nolfi and Floreano, 2000). Moreover, it is possible to build physics-realistic models that do not correspond to living systems, such as in studies of the evolution of morphology (Pfeifer and Bongard, 2006). Typically these simulation models are based on physics engines that simulate the physics of object–object interaction dynamics with a high degree of fidelity. Despite the fact that the use of a simulation might not provide a full model of the complexity present in the real environment and might not assure fully reliable transferability of the controller from the simulation environment to the real one, robotic simulations are of great interest for cognitive scientists (Ziemke, 2003). For example, a simulator for the iCub robot (Metta et al., 2008; Tikhanoff et al., 2008) magnifies the value a research group can extract from the physical robot by making it more practical to share a single robot between several researchers. The fact that the simulator is free and open makes it a simple way for people interested in the robot to begin learning about its capabilities and design, with an easy "upgrade" path to the actual robot due to the protocol-level compatibility of the simulator and the physical robot. And for those without the means to purchase or build a humanoid robot, such as small laboratories or hobbyists, the simulator at least opens a door to participate in this area of research.
- *Physical robot platforms in cognitive robotics.* This is for embodied models of cognitive capabilities directly implemented and tested in the physical platform such as the iCub robot (Metta et al., 2008; Macura et al., 2009). Physical robot models are important when one wants to study the detailed physics of interaction dynamics of specific configurations of sensors and actuators. The main field of cognitive modeling relying on physical robot platforms is that of cognitive robotics (Metta and Cangelosi, in press). In particular, cognitive robotics regards the use of bio-inspired methods for the design of sensorimotor, cognitive, and social capabilities in autonomous robots. Robots are required to learn such capabilities (e.g., attention and perception, object manipulation, linguistic communication, social interaction), through interaction with their environment, and via incremental developmental stages. Cognitive robotics, especially approaches that focus on the modeling of developmental stages (aka developmental and epigenetic robotics), can be very beneficial in investigating the role of embodiment, from the early stages of cognitive development to well developed cognitive systems, and to study how bodies and cognitive abilities co-evolve and exert significant influences on one another.

The choice of the most suitable modeling approach from the three methodologies listed above depends on the specific aims of the research, the availability of resources, and the consideration of the technical issues specific to the chosen methodology. For example, the first two approaches based on cognitive agents and on simulated robot agents are useful when the details of the whole embodiment system are not crucial, but rather it is important to investigate the role of specific sensorimotor properties in cognitive modeling. Moreover, the practical and technical requirements of the first two methods are limited as they are mostly based on software simulations. Instead, the work with physical robot platforms, as in cognitive and developmental robotics, has the important advantage of considering the constraints of a whole, integrated embodiment architecture. In addition, robotics experiments can demonstrate that what has been observed in simulation models can actually be extrapolated to real robot platforms. This enhances the potential scientific and technological impact of the research, as well as further demonstrating the validity of cognitive theories and hypotheses.

Ernest: I see that there is a range of possibilities here. Do you think that, to study cognition, the bodies of robots should be the same, or very similar to, the bodies of humans or of other animals?

Mary: The kind of embodiment and the constraints that have to be taken into consideration depend on what you expect from embodied computational modeling. On the one hand, modeling can help to find novel ways of understanding phenomena that are potentially applicable to cognition in general, such as the idea that sensory processing, categorization, and action planning are interdependent rather than separate processing stages. On the other hand, if one aims to produce specific predictions about, say, humans, then she should aim at replicating the same bodies and the same constraints (e.g., environmental and social), or at least a useful approximation – which could be difficult to define *a priori*.

Ernest: One could argue that this is not the whole story, though, since there are additional constraints that could be potentially central to embodied cognitive modeling, such as brain dynamics and their peculiarities and limitations, which are could also be part of the robot embodiment in some sense.

Mary: I see your point here. Modeling in general is about finding useful abstractions, but it is difficult to define *a priori* which constraints should be included in embodied computational models, and which should not. There is a debate on this topic in the cognitive robotics community, with positions that range from defenders of biologically constrained methods to the less demanding artificial

life approach. Although this is still an open point, it is necessary to recognize that, compared to traditional AI methodology that focuses on "abstract" or "general" intelligence or problem solvers, embodied cognitive modeling suffers more from the idiosyncrasies of what is meant to be modeled, be it a human or another living organism, because it takes a theoretical stance on the role of the "substratum" of cognition and the body.

Ernest: I see another problem of embodied computational models compared to traditional ones. Indeed, one important aspect of embodied models is that they should be coherent at the architectural level; or, in other terms, that their functions should not be developed by fully encapsulated models that work in isolation. This is especially true for the realization of higher-level cognitive abilities, such as reasoning, language, and categorical thinking, which cannot be totally disjoint from the neural systems that, say, direct eye movements and attention, regulate posture, or prepare actions to be executed.

Mary: This seems to rule out the hybrid approaches that are popular in robotics, in which complex cognitive skills are juxtaposed to basic sensorimotor skills, with a minimal (and predefined) interface. In addition, this poses a challenge to any kind of modular design in which functionalities are partly or completely encapsulated and do not interact, calling overall for a truly integrative theory. Understanding to what extent modelers can use modular design and what functionalities actually interact in any given cognitive process is both an important research aim and is crucial for the realization of working robotic systems. Indeed, to achieve the latter aim, it would be very difficult to simply connect all components, but rather the design of their coordination is crucial (Barsalou et al., 2007).

Interim conclusion. Novelty of grounded cognitive modeling and cognitive robotics

Ernest: We have discussed many important ingredients of embodied computational models, but I can easily imagine that some of them are already used in computational modeling and robotics. In your opinion, what are the most novel elements?

Mary: There are a few points that circulate to some extent in the literature, but to which embodied computational modeling should give extra emphasis: (1) Representations (grounded modal symbols) and cognitive abilities are not "given" but learned through sensorimotor interaction and on the top of sensorimotor skills and genetically specified abilities. Take as an example spatial abilities. A natural way to implement them using early connectionist (PDP) models is to encode spatial relations in the input nodes, and them let the agent learn navigation on top of them by capturing relevant statistics in the input. Rather, in this methodology even spatial relations should be autonomously acquired. (2) Higher-level cognitive abilities (in the individual and social domains) develop on top of the architecture for sensorimotor control. The re-enactment of modal representations rather than the re-coding in amodal format determines them, and they typically reuse existing sensorimotor competences in novel, more cognitive domains (e.g., visuomotor strategies for the temporal domain; counting with fingers) rather than using novel components. (3) Embodied cognitive modeling should go beyond isolated models, for example, attention models, memory models, and navigation models, to focus on complete architectures that develop their skills over time (see, e.g., Anderson, 2010). (4) Embodied cognitive modeling should emphasize the fact that robots and agents are naturally oriented to action. Other abilities (e.g., representation ability, memory ability, categorization ability, attention ability) could be in the service of action themselves, rather than having disconnected functions (e.g., vision as a re-coding of the external world). This latter point would have an impact on the traditional conceptualization of cognition as a stage in the perception–cognition–action pipeline.

Ernest: I see that all these are important points, and I am sure that future research will point out other relevant ingredients as well. Concerning the impact that your research could have, I expect that if a strong case can be made for the efficacy of embodied cognitive models, it would contribute to the success of grounded theories in general (as it was the case for the adoption of modal representations in cognitive science under the influence of early AI systems based on the manipulation of logic rules). So, my question is: what is the equivalent of the grounded perspective in computational modeling?

Mary: Unfortunately, cognitive modeling does not yet have standard, off-the-shelf solutions for implementing grounded phenomena, but there are several lines of research that could lead to convincing solutions. As a reaction to the conceptual and technical limitations of early AI symbolic systems, connectionism emphasized that cognition is based on distributed representations and processing (e.g., statistical processing) rather than on the manipulation of amodal symbols and abstract rules. Similarly, Bayesian systems showed the full effect of statistical manipulations on representations and structures (most of the time, however, this has been shown on predefined representations). Although connectionist and Bayesian systems might provide a good starting point for modeling grounded phenomena, they are not complete answers *per se*. In most connectionist and Bayesian architectures, processing occurs in modular systems separated from the modalities (similar conceptually to a transduction from modal to amodal representations), and the processing of cognitive tasks is specialized rather than shared with perception and action (similar conceptually to the manipulation of abstract rules, except that they are not explicitly represented but are implicitly encoded in the weights of the networks). This means that sensory and motor modalities do not affect cognition during processing, even though they can do so during development. This would be a weak demonstration of grounding and embodiment, showing that sensorimotor and bodily processes are affected by cognition, but not vice versa. Another extremely interesting research approach is *dynamic systems*, which emphasize situated action and the importance of a tight coupling with the external environment for the realization of all perceptual, action, and cognitive phenomena, as well as for their development (see, e.g., Thelen and Smith, 1994).

Within dynamic systems and dynamic fields thinking, cognition is mediated by the dynamics of sensorimotor coordination, and is sensitive to its parameters (e.g., activation level of dynamic fields, and their timing), rather than being separated from the sensory and motor surfaces. For this reason, dynamic systems could be an ideal starting point for modeling grounded cognition (Schöner, 2008). In addition, dynamic systems could potentially offer explanations that span multiple levels, including

brain dynamics, sensorimotor interactions, and social interaction, all using the same language and theoretical constructs. However, the full potential of this approach has not been shown yet. First, we need increasingly more dynamic systems models of the higher-level cognitive phenomena that interest psychologists, and which provide novel accounts of existing data. An interesting example is Thelen et al.'s (2001) model, which offers a novel explanation of children's behavior in the A-not-B paradigm. However, much remains to be done in this direction if dynamic systems want to become a paradigm for implementing complex operations on modal systems. Second, these systems should be increasingly *embodied*, instantiated for instance in robotic architectures, and tested in increasingly realistic *situated* (individual and social) scenarios, in order to tell a more complete story about the passage from realistic sensorimotor processing to realistic higher-level cognitive and social tasks. Third, dynamic systems tend to de-emphasize (or reformulate) internal representation and related notions, which are common currency in psychological and neuroscientific explanations, in favor of a novel ontology that includes conceptual terms such as "stability," "synchronization," "attractor," and "bifurcation." Besides the adequacy of these ideas, there is clearly a more sociological issue, and a new theoretical synthesis is required. It is logically possible that the new ontology replaces the old one (but then it is necessary that it provides higher explanatory power, and this is clearly acknowledged by psychologists and neuroscientists), that it re-explains the old one, offering novel and potentially more interesting theories of traditional concepts such as "representation," or that the two ontologies can be harmonized to some extent, but clearly the foundational aspects of a "dynamicist" cognitive science should be clarified before it can really offer itself as a novel candidate paradigm (Spivey, 2007). Although I have emphasized dynamic systems research, different research traditions, including for instance Bayesian approaches and connectionist networks, offer a good starting point for developing embodied cognitive models as well, providing that they successfully face the same challenges that I outlined before. However, I believe that a necessary complement to all these methodologies is to increasingly adopt cognitive robotics as their experimental platform, rather than designing models of isolated phenomena, or relaxing too many constraints about sensorimotor processing and embodiment. Indeed, it seems to me that cognitive robotics offer a key advantage to the aforementioned methodologies, because it emphasizes almost all of the components of grounded models: the importance of embodiment, the loop among perceptual, motor and cognitive skills, and the mutual dependence of cognition and sensorimotor processes.

TOPIC 2. WHAT CAN EMBODIED COGNITION LEARN FROM COMPUTATIONAL MODELING AND THE SYNTHETIC METHODOLOGY?

We have discussed what aspects of grounded cognition theories should inform computational modeling and the realization of robots informed by embodied cognitive abilities. Apart from the obvious scientific and technological achievements that these robots could provide, we have argued that computational models could help to answer open questions in the grounded cognition literature, and we have offered a few examples of this potential cross-fertilization. Here we focus on methodological issues: the role that computational modeling could play in developing grounded theories of cognition, and how it can complement theoretical and empirical research.

Ernest: It seems to me that computational modeling, as a methodology, is highly complementary to empirical research. It can help shed light on some aspects of grounded theories that are difficult to assess with empirical means only, and in doing so it can help to formulate better theories and specific predictions that can be tested empirically, and to even falsify current views (or at least lower our confidence level) by showing that they are computationally untenable. However, prima facie it is different to imagine how designing efficacious robots could be a convincing argument for psychologists and neuroscientists for or against a certain theory.

Mary: The primary role of embodied cognition models is not necessarily that of designing physical robots, such as the iCub, that are capable of reproducing human embodiment phenomena, although this is also a crucial benefit, as demonstrated by the advantages of biologically inspired systems. Rather, for cognitive scientists the robotic and computer simulation models are a way to verify and extend their hypotheses and theories. A simulation model can be viewed as the implementation of a theory in a computer or robot platform. A theory is a set of formal definitions and general assertions that describe and explain a class of phenomena. Examples of general cognitive theories are the ones on embodied cognition (as discussed in this article) but also other general, and hard to test theories such as in language evolution research that hypothesize a specific ability as the major factor explaining the origins of language (e.g., gestural communication for Armstrong et al., 1995; and tool making for Corballis, 2003). Theories expressed as simulations possess three characteristics that may be crucial for progress in the study of cognition (Cangelosi and Parisi, 2002). First, if one expresses one's theory as a computer program, the theory cannot help but be explicit, detailed, consistent, and complete because, if it lacks these properties, the theory/program would not run in the computer and would not generate results. Second, a theory expressed as a computer program helps generate detailed predictions because, as we have said, when the program runs in the computer, the simulation results are the predictions (even predictions not thought of by the researcher) for human behavior derived from the theory. And finally, computer simulations are not only theories but also virtual experimental laboratories. As in a real experimental laboratory, a simulation, once constructed, allows the researcher to observe phenomena under controlled conditions, to manipulate the conditions and variables that control the phenomena, and to determine the consequences of these manipulations. Indeed, this is one of the main advantages of computational modeling over other techniques. Computer simulations answer questions that cannot be addressed directly by empirical research; for instance, because the addressed phenomena cannot be observed directly or replicated (e.g., evolutionary phenomena) or are simply too difficult, risky, unethical, or expensive to test in the real world (e.g., the consequences of different learning episodes or model architectures for development). In addition, certain empirical phenomena depend on systemic and computational constraints on the behaving (or cognitive) system, irrespective of whether these constraints are posed by the system itself (e.g., bounded resources) or by external factors (e.g., situatedness). For instance, in a computational

study reviewed below, Pezzulo and Calvi (in press) investigated *what simulators could emerge* in a perceptual symbol system due to limited resources and other computational constraints.

Ernest: This is indeed interesting, and I see how computational modeling could contribute to the development and refinement of embodied theories of cognition. For instance, one issue that is widely debated is how to interpret the activation of the motor system during "cognitive" tasks, such as language understanding; or, in other terms, assessing if embodied phenomena are causal or epiphenomenal. So far, the most common methodology consists of measuring the time course of activation of the brain areas; for instance, of motor areas during language perception (Pulvermüller, 2005). In brief, early activations are more compatible with the view that embodied cognition plays a causal role. A more direct approach to the understanding of causality consists of interfering with the cognitive process, such as in TMS studies, but also with behavioral paradigms that create interference between tasks (e.g., a motor and a higher-level cognitive task).

Mary: Computational modeling can in principle help to resolve the aforementioned debate by providing principled ways to assess causality in cognitive processes, or at least provide a "sufficiency proof" that certain cognitive tasks, whose functioning is still unclear, can be explained on the basis of embodied phenomena. For instance, it is possible to compare how competing (epiphenomenal versus causal) computational models explain motor involvement during perception of language (Pulvermüller, 2005; Garagnani et al., 2007) or affordances (Tucker and Ellis, 2001, see later). In "epiphenomenal" models, representations (e.g., linguistic representations) are amodal and their processing is modular (i.e., separated from the sensory and action cortices), and when the "central" processing affects the "peripheries," this is an epiphenomenon without causal influence. On the contrary, in "causal" models all processing involves simulations and manipulation of modal representations. Implementing causal and epiphenomenal theories in computational terms, and embodying them into robot or agent architectures, can help to disambiguate their explanatory power, and to compare their empirical predictions.

Ernest: Another important issue for which computational models can provide insight is the apparently contradictory evidence on facilitatory or inhibitory roles of embodied processes. For instance, observation of actions performed by others can either facilitate or inhibit one's own motor actions, depending not only on the degree of convergence between the observed and executed actions, but also on the time course of the interference, and in some cases on the localization of the processing in the brain. The conflicting facilitatory versus inhibitory effects in the literature could, for instance, reflect the hierarchical nature of perceptual and motor systems, with different kinds of effects reflecting different levels in these systems, or alternatively, they could depend on the time course of the interference. What is lacking is a specific model of when and how various processes produce facilitation or inhibition, which could serve to test different hypotheses.

Mary: Relative to this issue, computational modeling can implement competing theories that aim at explaining interference, such as theories in which timing or competition for shared resources is viewed as the key element for modulating the interaction. Note that in different tasks and domains, the mechanisms and the effect could be different, so computational models should be endowed with precise details and contextual constraints. At the same time, it is desirable that common principles emerge, and thus the promise of computational modeling lies in providing a comprehensive framework for explaining interference within cognitive processing and reconciling the puzzling findings.

Ernest: You have argued convincingly that cognitive robotics could be a good starting point for modeling embodied phenomena, but I see a potential drawback in its method: when modeling cognitive functions, we should not forget that the way they are realized depends on the way they develop. Indeed, issues associated with the *architecture's development and plasticity* are important, too, including morphological, genetic, experiential, and social contributions, and how epigenesis is realized (Elman et al., 1996). For instance, as sensorimotor skills mature over time, cognitive abilities also develop in coordination with the acquisition of action skills (Rosenbaum et al., 2001). Also, social factors contribute substantially to development. Unfortunately, these dynamics are quite problematic to study empirically (but see Thelen and Smith, 1994). Again, however, here computational modeling can really help, addressing, for instance, the following questions: To what extent is a developmental trajectory necessary to "build" a grounded system? If it is necessary, what sorts of learning regimens are critical, and why?

Mary: This also seems like a nice place for modeling, at least in terms of the development or learning of knowledge in some specific domains. So, one could investigate whether you need to begin with some number of association areas already in place, for example. Or possibly this might be an interesting situation for using constructive algorithms like cascade correlation (Fahlman and Lebiere, 1990) or evolutionary techniques (Nolfi and Floreano, 2000) to see how "hidden units," or convergence zones, develop with experience. Overall, I would definitely agree that the dynamics of situated experience play an essential role in the shaping of cognitive abilities, and this is why I see developmental and epigenetic robotics approaches as extremely promising approaches for the construction of grounded systems (Weng et al., 2001).

Ernest: Overall, from this discussion I see significant potential for collaboration and cross-fertilization between the theoretical, empirical, and synthetic methodologies. However, I have recently participated in many good conferences, such as SAB, EpiRob, and ALife, in which I have seen many computational systems at work. Although most scientists in these conferences have similar motivations as you, that is, designing computational models that can tell us something about cognition (and in particular higher-level cognition), and using similar methods as you describe, including cognitive robotics, I am still unsure of what the results are. I don't deny that most of the things that I have seen are inspiring; still I fail to see how they can really influence my work. My guess is that most current computational models are mere "proofs of concept" and lack the adequate level of detail to start deriving precise predictions, or to simply be considered as useful tools by psychologists and neuroscientists. To realize the full impact of cognitive modeling on current (and future) theories, one not only has to develop computational systems that are generically informed by embodied cognition principles, but systems that target specific functionalities and experimental data.

Mary: This is indeed a necessary step. It is worth noting that although we still need a solid methodology for comparing empirical and synthetic data, various approaches have been proposed that compare modeling and empirical data both qualitatively and quantitatively. For example, within the literature on computational and robotic models of embodiment, there are studies where reaction times collected in psychology experiments have been directly compared with other time-related measurements in computational agents. Caligiore et al. (2010) directly compare the reaction times of Tucker and Ellis's (2001) stimulus response compatibility experiments with the time steps required by the simulated iCub robot's neural controller to reach a threshold that initiates the motor response. In Macura et al. (2009), a more indirect comparison for reaction time is used, based on the neural controller's error measurements in the production of the response. In the context of neuroscience, specific quantitative methods have been developed to compare brain-imaging data from human participants (e.g., from fMRI and PET methodologies), with "synthetic brain-imaging" data from computational models ranging from computational neuroscience models (Horwitz and Tagamets, 1999; Arbib et al., 2000) to connectionist models (Cangelosi and Parisi, 2004). The direct comparison between empirical and modeling data remains a major challenge for multi-agent models of cognition. This is the case for evolutionary models (e.g., language evolution models, Cangelosi and Parisi, 2002) where only general qualitative comparisons are possible due to the lack of data, or due to less realistic implementation of the sensorimotor and behavioral systems (e.g., multi-agent models of foraging). There are, however, other important reasons why our ideas fail to permeate cognitive science as a whole. On the one hand, there is a clear "sociological" problem of different languages and different conferences. On the other hand, there are more serious methodological differences that have to be bridged to some extent. It is often the case that modelers and empiricists have different research questions in mind, or use different lexicons. In addition, modelers in the communities that you mentioned tend to emphasize complete architectures and the fact that many processes interact, whereas empirical research often adopts a divide-and-conquer strategy and tends to study brain and behavior as if there were specialized processes, such as memory, attention, and language, with specialized neural representations.

Ernest: I would agree that the methodological differences make collaboration harder, and then that "empiricists" have to change as much as "modelers." The last point you mentioned is the most important one to me. Although you would rarely meet a cognitive scientist who claims to be a modularist, still some modularism (and localizationism) leaks into experimental paradigms in practice. Indeed, there is a tendency to study cognitive processes in isolation, as if they had separable neural substrates and encapsulated representations, a clear "objective" target that can be readily disconnected from the organism's behavior and goals (e.g., memory is for storage and retrieval the maximum number of elements, attention is for selecting stimuli), and as if they had specialized resources, inputs, and outputs that are clearly separable from those implied in other processes that take place at the same time. Overall, an added value of collaboration with embodied computational modelers could be a push toward integrated theories of cognition rather than theories of isolated functions.

TOPIC 3. CURRENT EMBODIED COMPUTATIONAL MODELS: SUCCESSES AND LIMITS

As we have briefly mentioned, there have been many attempts to model embodied and situated phenomena, especially in the connectionist and dynamic systems traditions. In these areas, it is widely recognized that the body, environment, and internal neural dynamics of agents are highly interconnected and shape one another (Clark, 1998; Pfeifer and Scheier, 1999). However, although these models fit the general framework of grounded cognition generically, most of them do not incorporate its specific predictions. In addition, only a few of them explicitly address complex (or even moderately complex) cognitive abilities, which are a true "benchmark" for grounded theories.

This section discusses a few robotic and agentive systems that exemplify current efforts toward the realization of embodied models of higher-level cognitive abilities, and specifically concepts and language understanding. This short review, which is undoubtedly biased by the authors' knowledge, is by no means representative of the most successful systems in technological terms, but is intended to illustrate current directions toward embodied cognitive models, together with their powers and limitations.

Ernest: Two intertwined areas in which embodied cognition currently is central are concepts and the expression and comprehension of those concepts via language. For example, it seems pretty clear at the moment that multiple perceptual modalities are intrinsic to object concepts. Lots of experiments and imaging studies suggest this. One clear demonstration of embodied cognition for the link between visual and motor processes is the stimulus response compatibility effect studied by Ellis and colleagues (Tucker and Ellis, 2001; Ellis et al., 2007). They have consistently demonstrated that when we perform visual categorization tasks (e.g., identifying artifact versus natural objects), the micro-affordances linked to the objects (e.g., power grasp for a large apple or precision grip for a small cherry) affect the visual categorization task.

Mary: These embodied phenomena have been recently implemented in iCub. Macura et al.'s (2009) model also implements the stimulus compatibility effects demonstrated by Tucker and Ellis (2001) in a simulation model of iCub (Tikhanoff et al., 2008). The experiments focus on training the robot to grasp objects using different responses, such as precision versus power grips for small and large objects, respectively. They also replicate the psychological experiments in which the objects can be categorized using different grips (e.g., precision grip for artifacts and power grip for natural objects). Specifically, in the simulation experiments there are four objects: two larger objects ("big-ball" and "big-cube") for power grips; and two smaller objects ("small-ball" and "small-cube") for precision grips. In this simulation, the round objects (big and small balls) are viewed as natural objects, whereas the cubes are viewed as artifacts. The training data consist of a set of grasping sequences for each object. A connectionist network is used to learn and guide the robot's behavior and to acquire embodied representations of objects and actions. The neural architecture, based on the Jordan recurrent architecture, has recurrent connections to permit information integration and the execution of actions such as grasping. The robots successfully learn to handle and categorize the objects as per the two tasks.

One important test of this model of object grasping and microaffordances is the comparison of the congruent condition (where the categorization grip is in agreement with the natural grip) and the incongruent ones (where there is mismatch between the categorization grip and the natural grip). The trained neural networks were presented each object in turn, where the desired target depended on the task being performed. The network test error was used as an equivalent of the participant's reaction time performance. Test results are highly consistent with psychological experiments where categorization latencies are shorter in congruent than in incongruent trials. In addition, the reaction times for larger objects were faster than for smaller objects, as was also the case in psychological experiments. This indicates that the robot was able to generalize a grasping sequence for each task and object from the four grasping sequences used in training, hence learning to appropriately grasp and categorize objects based on their shapes and sizes.

This computational model of Tucker and Ellis's (2001) compatibility effect demonstrates that it is possible to build robots capable of performing object manipulation tasks using the same constraints and mechanisms observed in human embodied cognition. Moreover, related models of microaffordance effects have been developed using a neurally plausible organization of the robot's neural architecture (Caligiore et al., 2010), with an extension of this model to simulate other compatibility effects, such as those studied by Borghi et al. (2004), which are also language-related.

Ernest: This is indeed very interesting, and I am curious to see how these studies evolve toward a comprehensive design methodology. However, most scientists, even those not convinced by embodied theories might admit that certain concepts, and especially concepts for manipulable objects, are partially represented in motor terms, and might recruit the motor system (even if they would not admit that motor processes are *necessary* for their understanding). What is definitely less clear is how you might model abstract concepts, such as objects that have no clear reference to observable or manipulable entities, or concepts that seem to be essentially "linguistic." How might we do this?

Mary: Currently, there are few embodied models of cognition that address the issue of how to develop concepts that depart from the most immediate sensations and actions, and grounded processes for their manipulation. One important line of research touches on the issue of how language and the conceptual system interact. Cangelosi and Riga (2006) present a simulated robotic agent model of the combination of sensorimotor categories, paired with language learning, to autonomously generate new action concepts. This is achieved through a connectionist implementation of the mental simulations in Barsalou's perceptual symbol system. The model is based on two humanoid robotic agents: the demonstrator and the imitator. The demonstrator (teacher) shows the correct performance of basic motor primitives (e.g., close the left arm, go forward, etc.) and also names the actions being demonstrated. The researcher programs this robot to perform these basic actions. The second agent, the imitator (learner), learns the actions by imitating the demonstrator's behavior. This agent is equipped with an artificial neural network that can learn to perform the basic actions by predicting the demonstrator's movement trajectories. The robot's neural controller also learns the words associated with the actions, so that when the imitator "hears" a word, it can perform the corresponding action. This training phase based on the visual demonstration and simultaneous naming of actions is called the "direct grounding" phase and resembles the way in which children acquire new concepts while an adult comments on their actions (Pulvermüller, 2005). Subsequently the demonstrator teaches new composite, higher-order actions solely through language instructions. The demonstrator utters sentences such as "grab is close_left_arm and close_right_arm" (no visual demonstration of the grab action is given). The learner's neural network uses a new learning algorithm that allows it to transfer the sensorimotor grounding of the basic words ("close_left_arm") to the new linguistic concept of "grab." This is achieved by first (internally) simulating the individual actions and by later reusing its own predicted output motor states as teaching inputs for the new linguistic concepts. This is an operational neural network implementation of a mental simulation mechanism in perceptual symbol system theory. Through this internal (mental) simulation, the imitator agent learns to perform and demonstrate the higher-order motor concept of grabbing. This training phased is called "grounding transfer." This model is an example of a higher-level (i.e., language-related) cognitive model of embodiment theory. A related robotic model, implemented by Madden et al. (2009), uses situated simulation as a "middle layer" for connecting propositional representations of sentences and the robot's sensorimotor experience. This system permits the temporal unfolding of propositions under the guidance of situated simulations and, at the same time, successfully demonstrates grammatical control of aspects of the simulation, beginning to tackle the broad issue of language comprehension and its neural bases. This approach could shed light on the relations between simulations and language, and how the linguistic system can be used to control simulations.

Other robotics models of language learning, which have a direct impact on the embodied literature, are Steels and Kaplan's (2000, 2002) models of the cultural evolution of language. In these computational studies, the focus has been on the social interaction between robotics agents that leads to the self-organization of shared lexicons.

Another relevant line of research touches on the issue of how concepts can be grounded in anticipated action or interaction with objects. For example, Moller and Schenck (2008) have studied how navigation-related concepts such as "far" or "closed path" could derive from the internal simulation of robot navigation. Interestingly, these concepts are grounded in the robot's anticipated perception, but go far beyond mere perception and include action possibilities, suggesting a route toward the development of more abstract knowledge. In a similar vein, Roy (2005a) has proposed a schema-based robot architecture in which the meaning of words and sentences in natural language are grounded in expectations relative to the robot's sensorimotor flow. For example, simple words such as "red" that refer to perceptual features have their grounding in expected sensory information in the robot's sensors. Concepts and words that refer to reachable and graspable objects are grounded in perceptual and motor schemas and in the expectations they produce during action planning and performance. For instance, the meaning of "sponge" is a set of expected action outcomes (e.g., an anticipated softness).

Finally, the emergence of grounded categories for motivation-related concepts such as "prey" and "predator" has been studied by adopting a simulated robotic agent methodology (Pezzulo and Calvi, in press). The computational architecture first learns multiple perceptual symbols in the form of action schemas that couple perception of the entities' features and action patterns that are more useful in the presence of the entities, such as escape or approach. Successively, as the agent interacts with the entities in its environment and learns to adaptively find food and escape from predators, associations are created within the perceptual symbols and between the perceptual symbols and the agent's internal motivational states, namely hunger and fear. In this way, entire simulators develop that cluster (or categorize) external entities and events in terms of the integration of possible associated perceptual events, actions, and motivational value for the agent, and which form reusable "situated conceptualizations." Pezzulo and Calvi (in press) observed that many simulators emerge and become encoded within the architecture's associative links. On the one hand, simulators cluster entities that have similar perceptual appearance or behavior; on the other hand, two more simulators emerge that correspond to the two families of "preys" and "predators," and determine highly coherent motivational dynamics (related to hunger and fear, respectively). Importantly, the simulators are not (only) memory structures, but support simulations and the dynamic reactivation of perceptual symbols (Barsalou, 1999). Simulators, with all their associated perceptual symbols, are acquired (partially or in their entirety) even from partial observations of salient events, or from changes in the agent's motivational state. In other words, when formed, simulators become tuned to types and not only tokens (and the exemplars which they were originally developed), simultaneously providing abstraction abilities and graded effects. This occurs because, depending on the circumstances, perceptual symbols in the simulators can be re-enacted to different degrees, or for different periods of time. For this reason, they can be considered categories that are grounded in the agent's sensorimotor and motivational repertoire. In addition, the study shows that the development of simulators produces an increase in the agent's adaptivity and performance rate.

All these are examples of recent efforts in the modeling of embodied phenomena. However, many of them can be considered preliminary investigations that do not derive precise predictions, but instead explore possible, novel ways to conceptualize and model cognitive phenomena, which are broadly inspired by embodied cognition research, but have not yet reached the level of detail that is required for a fruitful dialog with the empirical sciences. In addition, up to now most cognitive abilities, including for instance reasoning and memory abilities, and the realization of a full-fledged symbolic system, have not been addressed. In the rest of the article we discuss these and other related challenges for embodied computational modeling, and suggest promising directions of research.

EMBODIED COMPUTATIONAL MODELING: CHALLENGES FOR FUTURE RESEARCH

We have argued that embodied cognitive theories and computational modeling are complementary, and we have briefly described a few implemented systems that begin to show embodied cognitive abilities. In the remainder of the article, we discuss the most important challenges for future research in embodied computational modeling, and offer a few views on how to tackle them.

CHALLENGE 1. TAKING A DEVELOPMENTAL VIEWPOINT TO EXPLORE WHY AND HOW EMBODIED COGNITION COULD HAVE ORIGINATED

From an engineering standpoint, the smart way to build an intelligent organism would be to use a modular linear-systems approach to learn sensorimotor regularities to the greatest possible extent, and then to perform exhaustive computations on that input until a single optimal motor action could be selected and executed. The "motor babbling" of infants could be seen as supporting such a mechanism. Under such a rubric, one should restrict the motor system to generating and executing movement plans only for actions that have been confidently chosen as the appropriate effector output given the array of sensor inputs. In such a scenario, the motor-movement module is essentially a patient tele-operations system enslaved to the finalized commands of the cognition module, an approach that dominates the cognitive study of human motor control today.

However, the primate nervous system was not engineered by designers with such a linear-systems bias. Instead it evolved over millions of years, from quite different ancestors, through varying environmental niches, with substantial non-linear co-evolution among its many subsystems. The result is that the functional neuroanatomy of the human brain shares none of the feed-forward reasoning that makes a computer circuit-board understandable to an engineer. The human cerebral cortex is rife with top-down feedback projections and lateral connections that quickly scuttle a purely linear-systems analysis (see Carandini et al., 2005, for review). For example the orbito-frontal cortex, which is one of many sensory-integration regions in frontal cortex, has a direct functional projection back to visual cortex (Kveraga et al., 2007). This may allow expectations from multiple sensory sources to prepare visual cortex to process its afferent input in a richly contextualized manner (Mumford, 1992; Rao and Ballard, 1999).

Not only does feedback cause problems for a modular account of cognition, but the continuous flow of information among brain areas tends to blur the boundary that one might wish to daw between cognition and action. In other words, *embodied cognition may be unavoidable given the dynamics of the brain*, which was not designed as a whole, but underwent a complex evolutionary history in which successive adaptive solutions (to the same evolutionary problems) were implemented by adding on to earlier ones.

Evidence that the motor system is not in fact a "patient and obedient slave" that blindly follows the cognition system's finalized commands comes from a variety of studies that measure continuous motor output and multi-cell recording in premotor cortex. For example, Cisek and Kalaska (2005) report partial activation of two non-overlapping population codes in premotor cortex when the monkey is considering two possible reaching destinations. When Gold and Shadlen (2000) presented motion stimuli to a monkey and induced an eye movement via micro-stimulation of the frontal eye fields, the response-based eye movement (with which the monkey indicated the direction of perceived motion) and the micro-stimulated eye movement tended to average together into a single saccade that (with varying stimulus exposure times)

revealed the gradual accrual of sensory information apparent in neural activity in the frontal eye fields. That is, neurons in the frontal eye fields (an oculomotor region in frontal cortex) were accumulating partial information about the perceptual process before it had been allowed to reach completion. Similarly, eye movements in humans often "jump the gun" and fixate objects that correspond to partially active representations that in the end play no role in the person's planned action (Tanenhaus et al., 1995; Allopenna et al., 1998). Likewise, reaching movements will systematically curve toward multiple potential targets during the course of the reach (Tipper et al., 1997; Spivey et al., 2005; Song and Nakayama, 2009). Even repetitive bimanual rhythmic coordination can show effects, in its relative phase dynamics, of changes in a cognitive process "leaking" into the motor system (Shockley and Turvey, 2005). Furthermore, when two people are conversing, their motor systems can become entrained with one another, such that their postural sway becomes coordinated (Shockley et al., 2003), and their eye movements become coordinated (Richardson and Dale, 2005). These rich interactions among multiple sensory and motor subsystems in the brain are especially robust in children (Karmiloff-Smith, 1992), and may even lead to a child's early formation of concepts being undergirded by sensorimotor representations (Mandler, 1992; Lakoff and Johnson, 1999).

The architectural and developmental perspective points to a possible roadmap for studying embodied cognition through epigenetic, evolutionary, or developmental robotics. One could investigate whether organisms provided with simple sensorimotor circuits (Braitenberg, 1984) develop behavioral strategies for facing their adaptive problems (e.g., discriminating good from poisonous food) that can be considered precursors of cognitive abilities (e.g., categorization abilities). Importantly, one could also examine whether these cognitive abilities maintain vestigial (embodied) aspects of the earlier sensorimotor skills. The literature on evolutionary robotics offers examples of how flexible abilities (e.g., categorization abilities) can be developed, which have always been studied under the rubric of cognitive processing and without reference to behavior, and that rely on behavioral strategies only, without rich internal representations. For instance, Nolfi (2005; see also Beer, 2003) has studied how robots equipped with simple sensors can learn to discriminate circles from squares without any memory mechanism, by simply moving toward the most informative points so as to produce a different sensory flow for each shape. Such studies offer an "intuition pump" and a fresh view of how behavioral strategies could actually implement basic forms of cognition, suggesting that they could be reused (at least partially) even in more sophisticated ones – thus making the case that evolution constrained our higher-level cognitive abilities to be embodied. Less studied in this literature is, in general, an analysis of how increasingly complex abilities (e.g., human-level) could have developed on the basis of their putative evolutionary precursors.

A related line of research is the attempt to find basic (computational and neural) mechanisms that could have facilitated the development of cognitive abilities. For instance, some researchers have made the case that prediction abilities, originally developed for the sake of action control, could have bootstrapped higher-level cognitive and social abilities and prospection (Pezzulo and Castelfranchi, 2007, 2009; Moller and Schenck, 2008).

Overall, a first significant challenge for cognitive robotics research consists in studying the development of embodied cognition and how sensorimotor information interfaces with cognition. In other words, a possible lesson for cognitive robotics research is to pay attention to how an ability could have been developed, and not only to its end state. Developmental studies have played an important role in forming our understanding or a continuity of sensorimotor action and cognition (e.g., the dynamical system perspective of Thelen and Smith, 1994; see also von Hofsten, 2004). Cognitive robotics can contribute by systematically manipulating a robot's knowledge and skills, in order to understand what are the necessary prerequisites for the development of a particular cognitive ability, and by studying the environmental conditions that facilitate or prevent cognitive development.

CHALLENGE 2. EXPLORING THE (CAUSAL) INFLUENCE OF EMBODIED PHENOMENA FOR COGNITIVE PROCESSES

As suggested in the previous section, cognitive processes appear to "leak" into the motor system. Therefore, it appears that the very reason we can measure cognitive processes via nearly continuous dense-sampling recordings of motor movement (e.g., eye movements, reaching movements, bimanual rhythmic movements, postural sway) originates from the fact that the neural subsystems implementing those cognitive processes cannot help but "leak" their patterns of neural activation continuously into the various motor subsystems. That is, the very fact that we can learn about cognition by recording continuous motor output strongly implies that cognition is embodied. When those neural subsystems are dynamically coupled, a signal arising in one of them routinely is detectable as a signal in the other. Importantly for computational implementations of embodied cognition, this should be expected to happen in both directions – otherwise embodied phenomena would have no causal impact on cognitive processing.

Bidirectional synaptic pathways are the rule in cortex (Churchland and Sejnowski, 1992), and unidirectional projections between cortical areas are quite rare. Therefore, from a neurophysiological standpoint, it would seem unlikely that cognitive subsystems (in frontal cortex) could "leak" their patterns of neural activation out to sensory subsystems (in occipital and temporal cortices) and motor subsystems (in frontoparietal areas), but not the inverse. However, in formulating a defense for the amodal symbolic cognition framework, Mahon and Caramazza (2008) have proposed exactly that. To their credit, they acknowledge the preponderance of existing laboratory evidence supporting the spreading of activation from cognitive processes to sensory and motor processes. They propose an account in which cognition itself may be "disembodied" in the traditional sense that it conducts its business via amodal rules and abstract symbols (unlike perception and action), and the activation of those symbols then spreads to sensorimotor areas to produce the kind of results found in much of embodied cognition research. If the directionality of that spreading activation is such that the symbolic processes modulate sensorimotor processes, but not vice versa, then computations within the cognition module would not be influenced by whether or not those connections to sensorimotor processes existed (see also Pylyshyn, 1974). Adhering to that unidirectional influence is crucial in Mahon and Caramazza's (2008) proposal because if

sensorimotor processes can directly influence the algorithms being used in the symbolic processes, then those symbolic processes would not be purely amodal and abstract.

Mahon and Caramazza's (2008) argument is a powerful one because most of the evidence for embodied cognition can fit nicely into their spreading activation account, which preserves a role for purely abstract symbolic processing at some level. However, evidence for the other direction of influence, which compromises the purity of abstract symbolic processing, is beginning to accumulate (reviewed in Barsalou et al., 2003).

More embodiment experiments need to explore this directionality of effect: sensory and motor perturbations influencing central cognitive processes. Studies that show an early influence of perceptual and motor processes on cognition (see later on timing issues), as well as studies that suggest causality (e.g., TMS studies) are particularly relevant because these are difficult to explain as a by-product of a spreading activation and reverberation phenomenon. For example, Pulvermüller (1999) reported a collection of neuroimaging findings demonstrating that comprehension of action-based language triggers activation not only of language areas of the brain but also of limb-appropriate areas of motor cortex. Those findings epitomize the typical directionality of effect in embodiment studies. To show the reverse, Pulvermüller et al. (2005) conducted a transcranial magnetic stimulation study showing that mild TMS potentiation of the leg region of motor cortex improved reaction times to leg-action words (compared to arm-action words) in a lexical decision task (see also D'Ausilio et al., 2009). Behavioral studies have demonstrated this type of phenomena as well. After Richardson et al. (2003) showed that the image-schematic orientation of certain verbs influences visual attention in an object discrimination task, Toskos et al. (2004) showed that a controlled regime of repeated horizontal or vertical eye movements influenced memory for the (vertical or horizontal) verbs that were heard during those eye movements. In an elegant pair of studies, Meteyard et al. (2007) first showed that hearing directional motion verbs influenced *d*-prime in a motion detection task, and then Meteyard et al. (2008) showed that watching subtle visual motion signals influenced reaction times to directional verbs in a lexical decision task. Finally, one of the few examples of embodiment influencing high-level cognitive reasoning comes from a problem-solving experiment where the burgeoning onset of insight into the solution (the Aha! moment) for a difficult diagrammatic problem was correlated with a particular pattern of spontaneous eye movements which seemed to "participate" in the generation of the solution (Grant and Spivey, 2003). Thomas and Lleras (2007) then used the same problem and diagram, but enforced that particular pattern of eye movements as a secondary task, and participants were suddenly able to discover the solution with a significantly higher frequency. Both Smith (2005) and Ross et al. (2007) showed that repeated pairing of objects with actions influences their cognitive representation, as measured with a stronger congruency effect in object classification after training compared to before training. This reverse directionality of motor processes influencing cognitive processes is not merely another instance of Mahon and Caramazza's (2008) spreading activation idea. It shows that those cognitive processes do not "go about their business" the same way they always would have irrespective of the motor constraints. It shows that the cognitive algorithms for word reading, speech recognition, object recognition, and even problem-solving, all incorporate information from motor algorithms when producing their results.

Because these directionality effects, and their time courses, add further nuance and structure to the embodied cognition literature, it becomes especially important for theories of embodied cognition to be computationally implemented in order to make quantitative predictions of laboratory results explicit, and to study rigorously what is the role of perceptual and motor processes in cognitive tasks by comparing systems that include or exclude them (for instance, simulating lesions or "virtual lesions" such as TMS). If we are to understand embodied cognition as a natural consequence of rich and continuous recurrent interactions among neural subsystems, then building interactivity into models of cognition should have embodiment fall out of the simulation naturally. A number of neural network models (Howell et al., 2005; Mayberry et al., 2009; Anderson et al., 2010), computational simulations (Joyce et al., 2003; Scheutz et al., 2004; Cangelosi and Riga, 2006; Pezzulo and Calvi, in press) and robots (Brooks and Stein, 1994; Roy, 2005b) have begun to implement such simulations of embodied cognition, and further research along these lines is required.

CHALLENGE 3. SPECIFYING THE TIME COURSE OF ACTIVATION FOR EMBODIED CONCEPTS

We have discussed that processing models of embodied cognitive phenomena are needed at this stage. A key issue for assessing the suitability of such models is their ability to explain the unfolding in time of cognitive processes, and how they relate to the timing of motor and perceptual processes in the brain. Indeed, assessing the time course of activation of motor processes during cognitive tasks is key to arguments regarding their causal role. Take as an example the processing of language. Embodied language processing (both comprehension and production) has now been studied in a number of ways, covering neuroscientific as well as behavioral approaches. Several empirical findings constrain the temporal dynamics of embodiment processes and should thus also advise their computational modeling. One well-known finding is the rapid and somatotopic activation of motor- and premotor cortical areas while passively viewing action verbs (Pulvermüller, 2005). This finding suggests that the meaning of words is available as early as 150 ms after visual stimulus onset, a result that has been corroborated by electrophysiological evidence (Sereno et al., 1998; Sereno and Rayner, 2003). More importantly, it means that a body-specific representation of word meaning has been created within such a short time that it is unlikely that strategic factors will have contributed to the effect. How does a computational architecture for language comprehension include access to bodily representations? Moreover, how does the comprehension process in turn affect the use of motor structures, as is suggested by the kinematic recordings of Boulenger et al. (2006)? These authors found that action verb meaning selectively interferes with action execution within 200 ms of action onset, but only when the action is initiated prior to lexical processing. When lexical processing precedes motor activity, the effect is facilitatory. In a similar vein, the incremental reading work by Zwaan and Taylor (2006) suggests that there is an immediate but time-limited activation of motor congruency effects that can be expressed in faster knob turning when reading about a directionally

corresponding action. Interestingly, this congruency effect can then be re-instantiated when referring back to the described action with an adjective (Taylor and Zwaan, 2008).

The time course of embodiment effects is also of interest as an internal validation of the embodied cognition view of conceptual representation. One would expect that concrete concepts that have had a more direct grounding in sensorimotor experiences should show more rapid embodiment signatures when compared to more abstract concepts which require metaphorical mapping and indirect grounding. But even when controlling for word frequency, this comparison may be flawed, due to the differential age of acquisition which favors concrete words. In addition, as we have already remarked, timing issues could be critical for explaining apparently contradictory findings on the interference (facilitation or inhibition) of embodied processes.

In summary, there is now a detailed body of work on the time course of activation of concepts. What is needed now are integrative models that make specific predictions regarding the time course of embodied cognitive processes. One example is Chersi et al. (2010) who studied the precise dynamics underling the relation between language and action. Their model predicts interference or facilitation effects across linguistic and action tasks as depending on the time course of activation of associated representations.

CHALLENGE 4. DEVELOPING EMBODIED COMPUTATIONAL MODELS OF SYMBOLIC AND LINGUISTIC OPERATIONS

It is a common view that a true "benchmark" for embodied theories of cognition is explaining symbolic operations, which have been the province of amodal theories since the beginning of cognitive science (with few exceptions). This is certainly a difficult challenge for computational models as well; AI systems have tackled symbolic operations from the very beginning, but with little success. The wide domain of symbolic manipulations, which loosely includes reasoning and abstract thinking, predication, conceptual combination, language and communication, and which is typical of humans and possibly few other species, is both a challenge and a huge opportunity. Could embodied cognition be the right route for explaining computational symbolic processing? And, at the same time, could computational modeling provide strong evidence in favor of embodied cognitive theories? Our introductory examples of embodied arithmetic suggest a positive answer.

However, developing embodied approaches to symbolic operations requires a rethinking of most of the basic assumptions of traditional symbolic processing, in which symbols were taken as input, represented and outputted as symbols (with the notable exception of a few connectionist models). Indeed, it is still unclear how, from an embodied perspective, basic symbolic operations should be implemented. Barsalou's (1999, 2003, 2005) articles make the case that perceptual symbols can implement symbolic (or symbolic-like) operations, and provide initial thoughts on how this could be possible, but there are still many open questions, as the following examples attest. What specific computational mechanisms underlie the type-token propositions that result from categorization? What specific computational mechanisms produce the basic inferences that follow from categorization and produce anticipation? What specific computational mechanisms integrate and combine concepts into larger conceptual structures as needed to comprehend the world and achieve goals? What specific computational mechanisms underlie top-down construals of the world that are mapped into online bottom-up sensorimotor experience, producing the fusions that characterize experience? To what extent are these operations implemented explicitly in the system versus emerging implicitly? Preliminary empirical evidence for these accounts of symbolic operations are reviewed in Barsalou (2008b).

As already remarked, symbolic processing is often considered to be highly related to abstraction on the one hand, and to language on the other hand. This leads to five additional important questions. How are abstract concepts represented and processed in a grounded system? What roles do they play in a grounded system? What is the role of language in learning and using them? To what extent should language or other communicative mechanisms be included in a grounded system? Should robotics researchers currently try to build a human (with language) or a non-human (with simple communication)? Recent research on language development and communication in robotics (reviewed above) has begun to elucidate these topics, but there clearly exist numerous avenues for future research.

CHALLENGE 5. REALIZING SITUATED AND COMPLETE ARCHITECTURES WITHOUT LOSING CONTACT WITH DATA

One could ask what specific scenarios should be studied to cause embodied cognitive modeling research to advance most rapidly. On the one hand, if one aims to replicate specific experimental results, the scenario studied is constrained by the original experimental set-up. On the other hand, building many micro-simulations, one for each task to be modeled, could lead to a proliferation of disconnected models. Therefore, it would be equally valuable to create "unified" scenarios, or scenarios that could support the modeling and testing of many embodied phenomena. To do so, it is necessary to review critically what are the most important (and general) characteristics of environments, embodiment, and tasks that could be included.

Because (goal-directed) situated action in the environment is fundamental for all organisms, implementing embodied cognition that supports intelligent activity in a few critical situations (e.g., related to the organism's life and death) may be a good place to start (Robbins and Aydede, 2008). This viewpoint is not novel, as most recent research in artificial cognitive systems has focused on the realization of situated agents, or agents that dwell in complex environments (however, simplified with respect to the real environment), and must "close the loop" from (real) perception to (real) action so as to satisfy their internal needs and motivations (which, in most cases, are quite simplified as well). This set-up is beneficial for embodied cognitive modeling for many reasons. First, it forces modelers to build complete embodied architectures for achieving goals in specific situations rather than implementing specific capabilities, such as goal setting, planning, perception, action, cognition, affect, reward, and learning. This permits us to study how perceptual, motor, affective, and cognitive abilities interact effectively, and how advanced abilities can emerge from the coordination of simpler ones (Barsalou et al., 2007). Indeed, because a central point of embodiment concerns the interactions among, and integration of, these systems, building larger scale models appears to be necessary. Second, in their attempt to build situated agents, modelers have

widely recognized that the ways in which agents interact with their environments profoundly modify their representations and cognitive abilities, and that indeed agents cannot be studied in complete isolation from the environments in which they acquire their skills, and without including realistic details of their embodiment. Thus, this approach points quite directly toward grounded and embodied approaches to cognition.

However, if, on the one hand, generality of scenarios is desirable, this procedure comes at the risk of losing contact with human and animal data. Indeed most models developed under the hat of "situated cognition" (or artificial life, or AI) are indeed only loosely related to what is currently known about animal cognition. In addition, the scenarios that are currently employed in artificial cognitive systems research are more related to the basic survival of the organisms, but make it difficult to tackle higher-level cognitive abilities. One challenge for future research in embodied cognitive modeling is the realization of design principles (for architectures, scenarios, and embodiment) that are general enough to study many phenomena, but at the same time are specific enough to avoid losing contact with data and animal or human experiments.

CHALLENGE 6. REALIZING REALISTIC SOCIAL SCENARIOS FOR STUDYING COLLABORATIVE, COMPETITIVE, COMMUNICATION, AND CULTURAL ABILITIES

In the previous section, we focused on the realization of (possibly complete) goal-directed agent architectures. On the other hand, the realization of social scenarios, which involve human–robot interactions or coordinated interaction of multi-agent teams, is important as well. Although most theories of embodied cognition tend to more strongly emphasize the individual than the social aspects of cognition, they are not at odds with acknowledging the essentially social nature of learning and life of most animal species (including, of course, humans), and on the cultural origin of their representations and behaviors.

A popular research field in robotics and human–robot interaction concerns imitation, mindreading, intersubjectivity, and tool use, with an emphasis on their reliance on subpersonal processes such as prediction and mental simulation, and their sensorimotor roots (Demiris and Khadhouri, 2005; Oztop et al., 2005; Arbib et al., 2009). Other studies have focused on the affective dimension of human–robot interaction (Breazeal, 2003), and could provide interesting insights for embodied cognition research, in which this topic is seldom studied.

Unfortunately, although social and cultural robotics are being increasingly studied (Breazeal, 2004), it is difficult to combine social, cultural, and embodied aspects in the same endeavor. However, this is potentially very interesting for embodied cognition research, since it enables testing the relative importance of embodied and social (or cultural) phenomena in shaping learning and behavior. For instance, one of the promises of the research program of Steels and Kaplan (2000, 2002) is providing insights on how embodiment and situatedness could have constrained language acquisition. Collective robotics studies on the combined evolutionary learning of (collective) behavior and language (Marocco et al., 2003) could be informative as well, as these dynamics would be difficult to test experimentally. These two examples, together with the other models of symbol grounding and language learning discussed above (e.g., Cangelosi and Riga, 2006), provide a computational framework for investigating the embodiment and situated cognition phenomena of language and communication. The double function of language, as a social/communicative means, and as an individual/cognitive capability, derives from its fundamental property that allows us to internally re-represent the world we live in. This is possible through the mechanism of symbol grounding, that is, the ability to associate entities and states in the external and internal world with internal categorical representations. The symbol grounding mechanism, as our language, has both an individual and a social component (Cangelosi, 2006). The individual component, called "Physical Symbol Grounding," refers to the ability of each individual to create an intrinsic link between world entities and internal categorical representations. The social component, called "Social Symbol Grounding," refers to collective negotiation for the selection of shared symbols (words) and their grounded meanings. The extensive evidence on the mirror neuron system in both individual and social cognition provides support for the hypothesis of a link between the social (e.g., imitation) components of action production/recognition and language and communication (Rizzolatti and Arbib, 1998). This hypothesis can be computationally investigated through social and cognitive robotics experiments, as in Tani et al.'s (2004) model of mirror neuron system in language learning. (The modeling of mirror neurons has attracted a lot of attention in the last years; we refer the reader to Oztop et al., 2006 for a detailed review).

Overall, robotic studies of interactive, social, linguistic, and cultural dynamics, along with their interrelations, are particularly important for the extension of current embodied theories, which have not sufficiently incorporated these aspects, instead mostly focusing on individual cognition. One reason for this lack of attention is that all these dynamics are extremely difficult and expensive to study experimentally. Here the synthetic methodology offers a significant contribution because it involves many fewer constraints than empirical studies with living organisms. To make this cross-fertilization possible, however, it is desirable to design social robotic studies that incorporate increasingly more complex social dynamics (imitation, cooperation, joint action, and possibly the dynamics of whole societies of agents) and aim to reproduce social grounding and symbolic learning phenomena. A competing constraint is that it is necessary for these studies to make explicit and testable predictions, which is rare at this time.

CONCLUSION

Embodied effects have been consistently found in many cognitive tasks, including, for instance, action and object observation, memory, and language processing. Comprehensive theories that advocate the importance of grounding, embodiment, and situatedness have been proposed to explain these findings, which are corroborated by empirical data, but in most cases lack a precise computational or mathematical formalization. Computational modeling of embodied phenomena could contribute to the development of embodied theories of cognition by having the same kind of impact that early AI concepts (such as symbols, plans, or chunking) had on early theories in cognitive science. In addition, compared to early AI studies, embodied computational models have the potential to

derive more precise predictions because they typically involve more realistic agent–environment interactions. Furthermore, embodied theories of cognition could provide insights for the realization of robots with sophisticated cognitive abilities that other design methodologies have not been able to realize.

Throughout this article, we have proposed that cognitive robotics is an ideal platform for studying embodied phenomena (including developmental and social aspects), and vice versa that embodied theories can provide insights for the realization of novel, more efficacious, and reliable artificial cognitive systems. To help the realization of embodied computational models, this article has clarified the most important components and processes that such models should include, highlighted how the synthetic methodology could help research in embodied cognition, and proposed six challenges for modelers. Our hope is that, in 5 or 10 years, we will see another special issue on "Embodied and grounded cognition" that describes success stories in tackling these challenges, and that in turn this progress will inspire novel and more complete cognitive theories, and ultimately a novel paradigm for (individual and social) cognition that has grounding, embodiment, and situatedness at its heart.

ACKNOWLEDGMENTS

We thank the reviewers for their precious suggestions. The effort of Giovanni Pezzulo was supported by the European Union FP7 project HUMANOBS. The effort of Angelo Cangelosi was supported by the European Union FP7 projects ITALK and ROBOTDOC. Angelo Cangelosi and Martin H. Fischer also acknowledge the support of the UK Engineering and Physical Sciences Research Council, through the project VALUE.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439.

Altmann, G., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264.

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266.

Anderson, S. E., Huette, S., Matlock, T., and Spivey, M. J. (2010). On the temporal dynamics of negated perceptual simulations. In F. Parrill, M. Turner and V. Tobin (Eds.), *Meaning, Form, and Body.* Stanford: CSLI Publications.

Arbib, M., Bonaiuto, J., Jacobs, S., and Frey, S. (2009). Tool use and the distalization of the end-effector. *Psychol. Res.* 73, 441–462.

Arbib, M. A., Billard, A., Iacoboni, M., and Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Netw.* 13, 975–997.

Armstrong, D., Stokoe, W., and Wilcox, S. (1995). Gesture and the nature of language. Cambridge: Cambridge University Press.

Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1235–1243.

Barsalou, L. W. (1999) Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–600.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 1177–1187.

Barsalou, L. W. (2005). "Abstraction as dynamic interpretation in perceptual symbol systems," in *Building Object Categories. Carnegie Symposium Series*, eds L. Gershkoff-Stowe and D. Rakison (Mahwah, NJ: Erlbaum), 389–431.

Barsalou, L. W. (2008a). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W. (2008b). "Grounding symbolic operations in the brain's modal systems," *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches*, eds G. R. Semin and E. R. Smith (New York: Cambridge University Press), 9–42.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364, 1281–1289.

Barsalou, L. W., Breazeal, C., and Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cogn. Process.* 8, 79–91.

Barsalou, L. W., Niedenthal, P. M., Barbey, A., and Ruppert, J. (2003). "Social embodiment," in *The Psychology of Learning and Motivation*, Vol. 43, ed. B. Ross (San Diego: Academic Press), 43–92.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thought*, eds D. Pecher and R. Zwaan (New York: Cambridge University Press), 129–63.

Beer, R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adapt. Behav.* 11, 209–243.

Bekkering, H., Neggers, S. F. W. (2002). Visual search is modulated by action intentions. *Psychol. Sci.* 13, 370–374.

Borghi, A. M., Glenberg, A., and Kaschak, M. (2004). Putting words in perspective. *Mem. Cogn.* 32, 863–873.

Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., and Nazir, T. A. (2006). Cross-talk between language processes and overt motor behavior in the first 200 ms of processing. *J. Cogn. Neurosci.* 18, 1607–1615.

Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology.* Cambridge, MA: MIT Press.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *Int. J. Hum. Comput. Stud.* 59, 119–155.

Breazeal, C. (2004). *Designing Sociable robots.* Cambridge, MA: The MIT Press.

Brooks, R., and Stein, L. (1994). Building brains for bodies. *Auton. Robots* 1, 7–25.

Caligiore, D., Borghi, A. M.., Parisi, D., and Baldassarre, G. (2010). TRoPICALS: a computational embodied neuroscience model of experiments on compatibility effects. *Psychol. Rev.* 117, 1188–1228.

Calvert, G. A., Spence, C., and Stein, B. E. (2004). *The Handbook of Multisensory Processes.* Cambridge, MA: MIT Press.

Cangelosi, A. (2001). Evolution of communication and language using signals, symbols and words. IEEE Transactions on Evolutionary Computation. 5, 93–101.

Cangelosi, A. (2006). The grounding and sharing of symbols. *Pragmatics Cogn.* 14, 275–285.

Cangelosi, A. (2010). Grounding language in action and perception: from cognitive agents to humanoid robots. *Phys. Life Rev.* 7, 139–151.

Cangelosi, A., and Parisi, D. (eds). (2002). *Simulating the Evolution of Language.* London: Springer.

Cangelosi, A., and Parisi, D. (2004). The processing of verbs and nouns in neural networks: insights from synthetic brain imaging. *Brain Lang.* 89, 401–408.

Cangelosi, A., and Riga, T. (2006). An embodied model for sensorimotor grounding and grounding transfer: experiments with epigenetic robots. *Cogn. Sci.* 30, 673–689.

Carandini, M., Demb, J., Mante, V., Tolhurst, D., Dan, Y., Olshausen, B., Gallant, J., and Rust, N. (2005). Do we know what the early visual system does? *J. Neurosci.* 25, 10577–10597.

Chersi, F., Thill, S., Ziemke, T., and Borghi, A. M. (2010). Sentence processing: linking language to motor chains. *Front. Neurorobot.* 4:4. doi: 10.3389/fnbot.2010.00004

Churchland, P. S., and Sejnowski, T. (1992). *The Computational Brain.* Cambridge, MA: MIT Press.

Cisek, P., and Kalaska, J. F. (2004). Neural correlates of mental rehearsal in dorsal premotor cortex. *Nature* 431, 993–996.

Cisek, P., and Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* 45, 801–814.

Clark, A. (1998). *Being There. Putting Brain, Body, and World Together.* Cambridge, MA: MIT Press.

Corballis, M. C. (2003). From hand to mouth: the origins of language. New York, NY: Princeton University Press.

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666.

Craighero, L., Nascimben, M., and Fadiga, L. (2004). Eye position affects orienting of visuospatial attention. *Curr. Biol.* 14, 331–333.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33, 25–62.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). Motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385.

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J Experimental Psychology: General*, 122, 371–396.

Demiris, Y., and Khadhouri, B. (2005). Hierarchical attentive multiple models for execution and recognition (hammer). *Robot. Auton. Syst. J.* 54, 361–369.

Dijkstra, K., Kaschak, M. P., and Zwaan, R. A. (2008). Body posture facilitates retrieval of autobiographical memories. *Cognition* 102, 139–149.

Ellis, R., Tucker, M., Symes, E., and Vainio, L. (2007). Does selecting one visual object from several require inhibition of the actions associated with non-selected objects? *J. Exp. Psychol. Hum. Percept. Perform.* 33, 670–691.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Eskenazi, T., Grosjean, M., Humphreys, G. W., and Knoblich, G. (2009). The role of motor simulation in action perception: a neuropsychological case study. *Psychol. Res.* 73, 477–485.

Fahlman, S. E., and Lebiere, C. (1990). "The cascade-correlation learning architecture," in *Advances in Neural Information Processing Systems*, Vol. 2, ed. D. S. Touretzky (Los Altos, CA: Morgan Kaufmann Publishers Inc.), 524–532.

Fellous, J.-M., and Arbib, M. (eds). (2006). *Who Needs Emotions? The Brain Meets the Robot,* Vol. 24. New York: Oxford University Press, 271–272.

Fias, W., Brysbaert, M., Geypens, F., and d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition,* 2, 95–110.

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex*, 44, 386–392.

Fischer, M. H., and Hoellen, N. (2004). Space-based and object-based attention depend on motor intention. *J. Gen. Psychol.* 131, 365–377.

Fischer, M. H., Mills, R. A., and Shaki, S. (2010). How to cook a SNARC: number placement in text rapidly changes spatial–numerical associations. *Brain Cogn.* 72, 333–336.

Fischer, M. H., Prinz, J., and Lotz, K. (2008). Obligatory attention to action goals. *Quarterly Journal of Experimental Psychology*, 61, 860–868.

Fischer, M. H., Shaki, S., and Cruise, A. (2009). It takes only one word to quash the SNARC. *Exp. Psychol.* 56, 361–366.

Fischer, M. H., and Shaki, S. (2011). Predilection or preconception? A reply to Treccani and Umilta. *Brain Cogn.* (in press).

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694–723.

Garagnani, M., Wennekers, T., and Pulvermuller, F. (2007). A neural model of the language cortex. *Neurocomputing* 70, 1914–1919.

Gibson, J. (1979). The ecological approach to visual perception. Lawrence Erlbaum Associates, Inc.

Glenberg, A. (1997). What memory is for. *Behav. Brain Sci.* 20, 1–55.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Gold, J., and Shadlen, M. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature* 404, 390–394.

Grant, E. R., and Spivey, M. J. (2003). Eye movements and problem solving: guiding attention guides thought. *Psychol. Sci.* 14, 462–466.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychol. Res.* 66, 174–179.

Hommel, B., Musseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878.

Horwitz, B., and Tagamets, M.-A. (1999). Predicting human functional maps with neural net modeling. *Hum. Brain Mapp.* 8, 137–142.

Howell, S., Jankowicz, D., and Becker, S. (2005). A model of grounded language acquisition: sensorimotor features improve lexical and grammatical learning. *J. Mem. Lang.* 53, 258–276.

Ito, Y., and Hatta, T. (2004). Spatial structure of quantitative representation of numbers: evidence from the SNARC effect. *Mem. Cogn.* 32, 662–673.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia* 33, 1419–1432.

Joyce, D., Richards, L., Cangelosi, A., and Coventry, K. R. (2003). "On the foundations of perceptual symbol systems: specifying embodied representations via connectionism," in *The Logic of Cognitive Systems. Proceedings of the Fifth International Conference on Cognitive Modeling,* eds F. Detje, D. Dörner, and H. Schaub (Bamberg: Universitätsverlag), 147–152.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Knops, A., Thirion, B., Hubbards, E. M., Michel, V., and Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science* 324, 1583–1585.

Kohonen, T. (2001). *Self-Organizing Maps.* Berlin: Springer-Verlag.

Körding, K. P., and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247.

Kosslyn, S. M. (1994). *Image and Brain.* Cambridge, MA: MIT Press.

Kveraga, K., Ghuman, A. S., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168.

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh.* New York: Basic Books.

Lindemann, O., Alipour, A., and Fischer, M. H. (in press). Finger counting habits in Middle-Eastern and Western individuals: an online survey. *J. Cross Cult. Psychol.*

Loetscher, T., Schwarz, U., Schubiger, M., and Brugger, P. (2008). Head turns bias the brain's random number generator. *Curr. Biol.* 18, R60–R62.

Macura, Z., Cangelosi, A., Ellis, R., Bugmann, D., Fischer, M. H., and Myachykov, A. (2009). "A cognitive robotic model of grasping," in *Proceedings of the Ninth International Conference on Epigenetic Robotics*, Venice, November 12–14, 2009.

Madden, C. J., Hoen, M., and Dominey P. F. (2009). A cognitive neuroscience perspective on embodied language for human–robot cooperation. *Brain Lang.* 112, 180–188.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Mandler, J. (1992). How to build a baby: II. Conceptual primitives. *Psychol. Rev.* 99, 587–604.

Marocco, D., Cangelosi, A., and Nolfi, S. (2003). The emergence of communication in evolutionary robots. *Philos. Trans. A Math. Phys. Eng. Sci.* 361, 2397–2421.

Mayberry, M., Knoeferle, P., and Crocker, M. (2009). Learning to attend: a connectionist model of situated language comprehension. *Cogn. Sci.* 33, 449–496.

Meteyard, L., Bahrami, B., and Vigliocco, G. (2007). Motion detection and motion words: language affects low-level visual perception. *Psychol. Sci.* 18, 1007–1013.

Meteyard, L., Zokaei, N., Bahrami, B., and Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Curr. Biol.* 18, R732–R733.

Metta, G., and Cangelosi, A. (in press). "Cognitive robotics," in *Encyclopedia of the Sciences of Learning*, ed. N. M. Seel (Springer).

Metta, G., Sandini, G., Vernon, D., Natale, L., and Nori, F. (2008). "The iCub humanoid robot: an open platform for research in embodied cognition," in *PerMIS: Performance Metrics for Intelligent Systems Workshop*, August 19–21, 2008, Washington, DC.

Moller, R., and Schenck, W. (2008). Bootstrapping cognition from behavior – a computerized thought experiment. *Cogn. Sci.* 32, 504–542.

Mumford, D. (1992). On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251.

Myachykov, A., Platenburg, W., and Fischer, M. H. (2009). Non-abstractness as mental simulation in the representation of number. *Behav. Brain Sci.* 32, 343–344.

Newell, A., and Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nolfi, S. (2005). "Categories formation in self-organizing embodied agents," in *Handbook of Categorization in Cognitive Science*, eds H. Cohen and C. Lefebvre (Elsevier), 869–889.

Nolfi, S., and Floreano, D. (2000). *Evolutionary Robotics*. Cambridge, MA: The MIT Press.

Oztop, E., Kawato, M., and Arbib, M. (2006). Mirror neurons and imitation: a computationally guided review. *Neural Netw.* 19, 254–271.

Oztop, E., Wolpert, D., and Kawato, M. (2005). Mental state inference using visual control parameters. *Cogn. Brain Res.* 22, 129–151.

Patterson K., Nestor P. J., and Rogers T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 12, 976–987.

Pezzulo, G. (2008). Coordinating with the future: the anticipatory nature of representation. *Minds Mach.* 18, 179–225.

Pezzulo, G., and Calvi, G. (in press). Computational explorations of perceptual symbol systems theory. *N. Ideas Psychol*. doi: 10.1016/j.newideapsych.2009.07.004

Pezzulo, G., and Castelfranchi, C. (2007). The symbol detachment problem. *Cogn. Process.* 8, 115–131.

Pezzulo, G., and Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychol. Res.* 73, 559–577.

Pfeifer, R., and Bongard, J. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence.* Cambridge, MA: MIT Press.

Pfeifer, R., and Scheier, C. (1999). *Understanding Intelligence.* Cambridge, MA: MIT Press.

Pinhas, M., and Fischer, M. H. (2008). Mental movements without magnitude? A study of spatial biases in

symbolic arithmetic. *Cognition* 109, 408–415.
Prinz, W. (1997). Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154.
Pulvermüller, F. (1999). Words in the brain's language. *Behav. Brain Sci.* 22, 253–279.
Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.
Pulvermüller, F., Hauk, O., Ilmoniemi, R., and Nikulin, V. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797.
Pylyshyn, Z. (1974). Minds, machines and phenomenology: some reflections on Dreyfus' "What computers can't do." *Cognition* 3, 57–77.
Radulescu, P. V., Adam, J. J., Fischer, M. H., and Pratt, J. (2010). Fitts's law violation and motor imagery: are imagined movements truthful or lawful? *Exp. Brain Res.* 201, 607–611.
Rao, R., and Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.
Richardson, D., and Dale, R. (2005). Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cogn. Sci.* 29, 1045–1060.
Richardson, D., Spivey, M., Barsalou, L., and McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cogn. Sci.* 27, 767–780.
Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194.
Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
Robbins, P., and Aydede, M. (eds). (2008). *Cambridge Handbook of Situated Cognition*. New York: Cambridge University Press.
Rosenbaum, D. A., Carlson, R. A., and Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annu. Rev. Psychol.* 52, 453–470.
Ross, B. H., Wang, R. F., Kramer, A. F., Simons, D. J., and Crowell, J. A. (2007). Action information from classification learning. *Psychon. Bull. Rev.* 14, 500–504.
Roy, D. (2005a). Semiotic schemas: a framework for grounding language in action and perception. *Artif. Intell.* 167, 170–205.
Roy, D. (2005b). Grounding words in perception and action: computational insights. *Trends Cogn. Sci.* 9, 389–386.
Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* 8, 657–661.
Scheutz, M., Eberhard, K., and Andronache, V. (2004). A real-time robotic model of human reference resolution using visual constraints. *Connect. Sci.* 16, 145–167.
Schöner, G. (2008). "Dynamical systems approaches to cognition," in *Toward a Unified Theory of Development: Connectionism and Dynamic Systems Theory Re-Considered*, eds J. P. Spencer, M. S. Thomas, and J. L. McClelland (New York: Oxford University Press).
Schwarz, W., and Keus, I. (2004). Moving the eyes along the mental number line: comparing SNARC effects with manual and saccadic responses. *Percept. Psychophys.* 66, 651–664.
Searle, J. (1980). Minds, brains and programs. *Behav. Brain Sci.* 3, 417–457.
Sereno, S. C., and Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends Cogn. Sci.* 7, 489–493.
Sereno, S. C., Rayner, K., and Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport* 9, 2195–2200.
Shockley, K., Santana, M., and Fowler, C. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326–332.
Shockley, K., and Turvey, M. T. (2005). Encoding and retrieval during bimanual rhythmic coordination. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 980–990.
Simmons, W. K., and Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol.* 20, 451–486.
Smith, L. B. (2005). Action alters shape categories. *Cogn. Sci.* 29, 665–679.
Song, J., and Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends Cogn. Sci.* 13, 360–366.
Spivey, M. J. (2007). *The Continuity of Mind*. New York: Oxford University Press.
Spivey, M. J., Grosjean, M., and Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10393–10398.
Steels, L., and Kaplan, F. (2000). AIBO's first words, the social learning of language and meaning. *Evol. Commun.* 4, 3–32.
Steels, L., and Kaplan, F. (2002). "Bootstrapping grounded word semantics," in *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*, ed. T. Briscoe (Cambridge: Cambridge University Press), 53–73.
Symes, E., Tucker, M., Ellis, R., Vainio, L., and Ottoboni, G. (2008). Grasp preparation improves change detection for congruent objects. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 854–871.
Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science* 268, 1632–1634.
Tani, J., Ito, M., and Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Netw.* 17, 1273–1289.
Taylor, L. J., and Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Q. J. Exp. Psychol.* 61, 869–904.
Thelen, E., Schöner, G., Scheier, C., and Smith, L. (2001). The dynamics of embodiment: a field theory of infant perseverative reaching. *Brain Behav. Sci.* 24, 1–33.
Thelen, E., and Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.
Thomas, L., and Lleras, A. (2007). Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychon. Bull. Rev.* 14, 663–668.
Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008). "An open-source simulator for cognitive robotics research: the prototype of the iCub humanoid robot simulator," in *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS'08)*, eds R. Madhavan and E. R. Messina (Washington, DC: IEEE), 57–61.
Tipper, S., Howard, L., and Jackson, S. (1997). Selective reaching to grasp: evidence for distractor interference effects. *Vis. Cogn.* 4, 1–38.
Toskos, A., Hanania, R., and Hockema, S. (2004). "Eye scanpaths influence memory for spoken verbs," in *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, eds K. Forbus, D. Gentner, and T. Regier (Mahwah, NJ: Erlbaum), 1643.
Tucker, M., and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Vis. Cogn.* 8, 769–800.
Verschure, P., and Coolen, A. (1991). Adaptive fields: distributed representations of classically conditioned associations. *Network* 2, 189–206.
von Hofsten, C. (2004). An action perspective on motor development. *Trends Cogn. Sci.* 8, 266–272.
Wallace, R. J. (1971). Stimulus–response compatibility and the idea of a response code. *J. Exp. Psychol.* 88, 354–360.
Weiller, D., Märtin, R., Dähne, S., Engel, A. K., and König, P. (2010). Involving motor capabilities in the formation of sensory space representations. *PLoS ONE* 5, e10377. doi: 10.1371/journal.pone.0010377
Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science* 291, 599–600.
Wood, G., Nuerk, H.-C., Willmes, K., and Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly*, 50, 489–525.
Ziemke, T. (2003). On the role of robot simulations in embodied cognitive science. *AISB J.* 1, 389–399.
Zwaan, R. A., and Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *J. Exp. Psychol. Gen.* 135, 1–11.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 June 2010; accepted: 04 January 2011; published online: 31 January 2011.

Citation: Pezzulo G, Barsalou LW, Cangelosi A, Fischer MH, McRae K and Spivey MJ (2011) The mechanics of embodiment: a dialog on embodiment and computational modeling. Front. Psychology ***2****:5. doi: 10.3389/fpsyg.2011.00005*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2011 Pezzulo, Barsalou, Cangelosi, Fischer, McRae and Spivey. This is an open-access article subject to an exclusive license agreement between the authors and Frontiers Media SA, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 22 November 2010
doi: 10.3389/fpsyg.2010.00173

Credit assignment in multiple goal embodied visuomotor behavior

Constantin A. Rothkopf[1]* and Dana H. Ballard[2]

[1] *Frankfurt Institute for Advanced Studies, Goethe University, Frankfurt am Main, Germany*
[2] *Department of Computer Science, University of Texas at Austin, Austin, TX, USA*

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Sascha Topolinski, Universität Würzburg, Germany
Stephen J. Flusberg, Stanford University, USA

****Correspondence:***
Constantin A. Rothkopf, Frankfurt Institute for Advanced Studies, Ruth-Moufang-Strasse 1, 60438 Frankfurt am Main, Germany.
e-mail: rothkopf@fias.uni-frankfurt.de

The intrinsic complexity of the brain can lead one to set aside issues related to its relationships with the body, but the field of embodied cognition emphasizes that understanding brain function at the system level requires one to address the role of the brain-body interface. It has only recently been appreciated that this interface performs huge amounts of computation that does not have to be repeated by the brain, and thus affords the brain great simplifications in its representations. In effect the brain's abstract states can refer to coded representations of the world created by the body. But even if the brain can communicate with the world through abstractions, the severe speed limitations in its neural circuitry mean that vast amounts of indexing must be performed during development so that appropriate behavioral responses can be rapidly accessed. One way this could happen would be if the brain used a decomposition whereby behavioral primitives could be quickly accessed and combined. This realization motivates our study of independent sensorimotor task solvers, which we call modules, in directing behavior. The issue we focus on herein is how an embodied agent can learn to calibrate such individual visuomotor modules while pursuing multiple goals. The biologically plausible standard for module programming is that of reinforcement given during exploration of the environment. However this formulation contains a substantial issue when sensorimotor modules are used in combination: The credit for their overall performance must be divided amongst them. We show that this problem can be solved and that diverse task combinations are beneficial in learning and not a complication, as usually assumed. Our simulations show that fast algorithms are available that allot credit correctly and are insensitive to measurement noise.

Keywords: modules, reinforcement, learning, credit assignment, reward

INTRODUCTION

Very early on it was recognized that to realize the sophisticated decisions that humans routinely make, their brains must have some kind of internal model (Tolman, 1948). One of the key figures in the modern day was Neisser (1967) who refined the idea of an internal cognitive architecture. Current systems codify experts' knowledge, e.g. (Anderson, 1983; Laird et al., 1987; Langley and Choi, 2006; Sun, 2006). The principal feature of these systems is their use fine-grained rules with variables and bind them by pattern matching. Their broad intent is to search for a sequence of rules that will solve a problem. Despite the challenging difficulties involved, expert systems have achieved notable successes, particularly in intellectual problems where the symbol bindings can be intuited, such as in algebraic problem solving (Ritter et al., 1998). However a crucial area that these systems have tackled more secondarily is that of perception and action[1].

In contrast, diverse communities in robotics and psychology have been working on cognitive architectures that take a more integrated approach to vision and action, and both have recognized that the ultimate model architecture will have a hierarchical structure (e.g., Brooks, 1986; Newell, 1990; Firby et al., 1995; Arkin, 1998; Bryson and Stein, 2001). Robotics researchers in particular have gravitated to a modular three-tiered structure that models strategic, tactical and detail levels in complex behavior (Bonasso et al., 1997). Embodied cognition integrates elements from all these advances but in addition places a special stress on the body's role in computation. It emphasizes that the brain cannot be understood in isolation as so much of its structure is dictated by the body it finds itself in and the world that the body has to survive in (Ballard et al., 1997b; Roy and Pentland, 2002; Barsalou, 2009; Glenberg, 2010). This has important implications for cognitive architectures, because the brain can be dramatically simpler than it could ever be without its encasing milieu. The reason is that the brain does not have to replicate the natural structure of the world or the special ways of interacting with it taken by the body but instead can have an internal structure that implicitly and explicitly anticipates these commitments. Research in this area has shown that using simulated figures in realistic virtual environments can make delicate manipulation problems of limited clearance more readily solvable (Badler et al., 1993, 1999) and revealed the economies of state needed to interact in dynamic environments (Terzopoulos et al., 1994;

[1]For example in Anderson's ACT-R, vision is appended as a subsystem, with the ability to search for parts of the image by coordinates or feature, its rules being based on Treisman (1980) and Trick and Pylyshyn (1994).

Terzopoulos, 1999; Sprague et al., 2007). Moreover, these compact state descriptions of sensorimotor interactions lend themselves to being modeled with reinforcement learning (RL).

Considerable empirical evidence has demonstrated that activity in human and animal brains can be related to variables in models of RL. The data comprises single cell activity in reward related visuomotor behavior in monkeys (Schultz et al., 1997b; Schultz, 2000) and BOLD activity using fMRI in humans during reward related and cognitive control tasks (Gottfried et al., 2003; Haruno and Kawato, 2006; Pessiglione et al., 2006). Although typical RL models can handle such small problems, they have the drawback that they do not scale up to large problems since the state spaces grow exponentially in the number of state variables. Furthermore, RL is mostly applied to individual tasks and not to tasks with multiple goals and task combinations. These problems have made it difficult to apply RL to realistic settings, with the result that the state spaces considered are generally small.

The scaling issue can be addressed by exploiting the structure present in a complex task through some form of factorization. While some previous work has developed techniques to learn such structure within a complex task (Guestrin et al., 2003), another approach is to start from independent tasks (Singh and Cohn, 1998; Sprague and Ballard, 2003) and consider their combinations. Imagine that you are late getting out of bed in the morning and have to quickly get ready and try to catch the bus. You have to get dressed, gather your things, run down the street toward the bus stop while avoiding other pedestrians, etc. That is, you have to pursue multiple goals at once. Our premise is that each of these goals has some intrinsic value to the overall enterprise, and that the brain has to know what these values are in order to juggle contingencies. Another important reason for knowing the value of the component modules is that this knowledge allows different combinations of multiple active modules to be used in many different tasks.

The computational difficulty arising from the modular approach is that the obtained reward needs to be attributed correctly in order for the modules to learn their respective contribution to the momentary reward. In many settings it only is reasonable to assume that a global signal of reward is available. Thus different active reinforcement learning modules have the problem of dividing the global reward up between them. Our focus is this problem. By solving this credit assignment problem correctly, individual modules can learn their respective contribution to achieving the current task combination.

Our robust solution to credit assignment succeeds by assuming that each module has access to the estimated sum of the reward estimates of other active modules. We derive formulas for estimates of reward that, assuming properties of the duration of episodes during which the concurrent goals are not changing, converge rapidly to their true values. We demonstrate that the algorithm can solve the credit assignment problem in a variant of a classical animal foraging problem in the literature (Singh and Cohn, 1998) as well as a more complex case of a human avatar learning multi-tasking in a virtual environment (Sprague et al., 2007). Thus, we show how a well-established reward-dependent learning algorithm that has been successful in modeling animal and human visuomotor and cognitive learning can be extended to learn solutions to multiple goal visuomotor behavior.

MATERIALS AND METHODS

REINFORCEMENT LEARNING BACKGROUND

A standard formalism for describing the brain's programs is that of Markov Decision Processes (MDPs). An individual MDP consists of a 4-tuple (S,A,T,R) with S being the set of possible states, A the set of possible actions, T the transition model describing the probabilities $P(s_{t+1}|s_t, a_t)$ of reaching a state s_{t+1} when being in state s_t at time t and executing action a_t, and R is a reward model that describes the expected value of the reward r_t, which is distributed according to $P(r_t|s_t, a_t)$ and is associated with the transition from state s_t to some state s_{t+1} when executing action a_t.

The goal of RL is to find a policy π that maps from the set of states S to actions A so as to maximize the expected total discounted future reward through some form of learning. The dynamics of the environment T and the reward function R are not known in advance and an explicit reward function R is learned from experience. RL algorithms effectively assign a value $V^\pi(s)$ to each state, which represents this expected total discounted reward obtainable when starting from the particular state s and following the policy π thereafter. Where γ is a scalar factor that discounts future rewards, $V^\pi(s)$ can be described by:

$$V^\pi(s) = E^\pi\left(\sum_{t=0}^{\infty}\gamma^t r_t\right) \tag{1}$$

Alternatively, the values can be parameterized by state and action pairs, denoted by $Q^\pi(s, a)$. Where Q^* denotes the value associated with the optimal policy π^*, the optimal achievable reward from a state s can be expressed as $V^*(s) = \max_a Q^*(s, a)$ and the Bellman optimality equations for the quality values can be formulated as:

$$Q^*(s,a) = \sum_r rP(r\,|\,s,a) + \gamma\sum_{s'\in S} P(s'\,|\,s,a)\max_{a'} \quad Q^*(s',a') \tag{2}$$

Temporal difference learning (Sutton and Barto, 1998), uses the error between the current estimated values of states and the observed reward to drive learning. In its related Q-learning form, the estimate of the value of a state-action pair is adjusted by this error δ_Q using a learning rate α:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\delta_Q \tag{3}$$

Two important expressions for δ_Q are (1) the original Q-learning rule (Watkins, 1989) and (2) SARSA (Rummery and Niranjan, 1994). The first is an off-policy rule, i.e., it uses errors between current observations and estimates of the values for following an optimal policy, while actually following a potentially suboptimal policy during learning. SARSA[2] is an on-policy learning rule, i.e., the updates of the state and action values reflect the current policy derived from these value estimates. As SARSA allows one to follow a suboptimal policy in the course of learning, it is well-matched for use with modules, which cannot always depend on following their own policy recommendations. Its learning rule is given by:

$$\delta_Q = r_t + \gamma Q(s_{t+1},a_{t+1}) - Q(s_t,a_t). \tag{4}$$

[2]SARSA is an acronym for the quintuple s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1} denoting the actual trajectory followed.

Evidence that both the Q-learning and the SARSA error signals are represented in the brain of animals and humans have been provided in numerous experiments (Schultz et al., 1997b; Schultz, 2000; Morris et al., 2006).

INDIVIDUAL TASK SOLUTIONS: MODULES

The essential architectural commitment is that the required behaviors can be realized with separate MDP modules. The primary assumption is that, to a first approximation, such modules are activated in subsets whose members either do not interfere with each other (Guestrin et al., 2003; Russell and Zimdars, 2003; Sprague et al., 2007), or, if they do, then the interference can be handled in a way that approximates the result one would obtain from the complete state space that included all the active module state values[3]. We first describe the equations that govern the situation wherein the modules are completely independent, then show the modifications for embodiment wherein modules have to agree on the action selected, and finally show the notation used to describe the situation where the instantaneous reward is only known for the total subset of active modules and not for individuals.

Embodied module definitions

An independent RL module with its own actions can be defined as an MDP, i.e., the ith module is given by

$$\mathcal{M}_i = \{S_i, A_i, T_i, R_i\} \tag{5}$$

where the subscripts denote that the information is from the ith MDP. The states of the different modules are assumed all non-overlapping. In such a case, the optimal value function is readily expressible in terms of the component value functions and the states and actions are fully factored so that there is no overlap between states and additionally the following two conditions hold. Where $s = \{s^{(1)}, \ldots, s^{(M)}\}$ is the combined state of the M modules and similar notation is used for **a** and **r**,

$$P\left(s_{t+1} \mid s_t, a_t\right) = \prod_{i=1}^{M} P\left(s_{t+1}^{(i)} \mid s_t^{(i)}, a_t^{(i)}\right) \tag{6}$$

$$P\left(r_{t+1} \mid s_t, a_t\right) = \prod_{i=1}^{M} P\left(r_{t+1}^{(i)} \mid s_t^{(i)}, a_t^{(i)}\right) \tag{7}$$

These two conditions can be used together with Eq. 2 in order to arrive at the result:

$$Q\left(s_t, a_t\right) = \sum_{i=1}^{N} Q\left(s_t^{(i)}, a_t^{(i)}\right) \tag{8}$$

If Eqs. 6 and 7 hold and all the rewards are known, the action maximizing Eq. 8 can be selected and is guaranteed to be optimal. In this decomposed formulation, each module can follow its own policy π^i, mapping from the local states s^i to the local actions a^i. This case is appropriate for a multi-agent setting when each module can be identified with a separate agent that may be expected to act independently.

However, our focus is the embodied cognition setting, where single agent pursues multiple goals that are divided up between multiple independent modules that a single agent can activate concurrently (Humphrys, 1996; Karlsson, 1997; Singh and Cohn, 1998; Sprague and Ballard, 2003). The consequence is that the action space is shared, so that all active modules must choose a single action. Thus the embodiment requires some form of action selection in order to mediate the competition between the possibly rivalrous actions proposed by individual modules. We use the probabilistic softmax action selection:

$$P\left(a_t^{(j)} \mid Q\left(s_t^{(1)}, a_t\right), \ldots, Q\left(s_t^{(N)}, a_t\right)\right) = \frac{e^{Q\left(s_t^{(j)}, a_t^{(j)}\right)/\tau}}{\sum_{i=1}^{M} e^{Q\left(s_t^{(i)}, a_t^{(i)}\right)}/\tau} \tag{9}$$

to choose the action, and once it has been selected, it is used for all modules. This type of action selection has been shown to model human behavior well in a variety of single goal decision making tasks (Daw et al., 2006; Rangel and Hare, 2010). The parameter τ controls the balance between exploration and exploitation during learning and usually decreases over time to reflect the shift toward less exploratory decisions over the course of learning. Note that in this formulation we propose to select a decision based on the combined value that the modules predict. This is different from Doya et al. (2002) where all modules contribute weighted by how well each module is predicting the dynamics of the world, irrespective of value or overall outcome. This is also different from Daw et al. (2005) where a single controller selects the next action alone based solely on the uncertainty of the current value estimates of a state so that each module needs to represent the same, full set of state variables.

This model has been very effective in representing human performance in the case where the multiple tasks are to walk down a sidewalk while simultaneously staying on the sidewalk, picking up litter objects and avoiding obstacles. **Figure 1**, a replication of Sprague et al. (2007), shows the results of the learning via RL of separate modules for each of these three tasks by an avatar model embedded in the same environment. The top panels in the figure show the discounted reward values as a function of the state space in front of the agent. The bottom panels show the respective policies. Note that for each of the modules the state estimate is different, as a consequence of the disposition of the agent in the environment and the relative positions of surrounding objects. **Figure 1** illustrates the action selection issue that crops up with the use of modules: actions recommended by individual modules may be different, requiring resolution by the use of Eq. 9.

Finally we can address the new constraint we are after and that is that the individual rewards due to each module are not known, but only the global reward is supplied to the agent at each time step. Using only this information, the agent needs to compute the share of the credit for each module. To describe this situation formally, we can write

$$M_i = \{S_i, A_i, T_i, G_{\mathcal{M}(t)}\} \tag{10}$$

[3]One always has to worry about whether the state definition does indeed capture all the relevant information. Formally one tackles this by appealing to additional structure of the *partially-observable* MDP that contains probabilistic machinery to represent the fact that being in any particular state is uncertain and has only an associated probability. However in the embodied cognition setting this extra machinery may not always be required, as extensive sensori-motor feedback, can render the uncertainties in the state estimate manageable with standard estimation techniques, such as Kalman filters as is done here. Nonetheless, the presented solution based on MDPs could be extended to consider belief states.

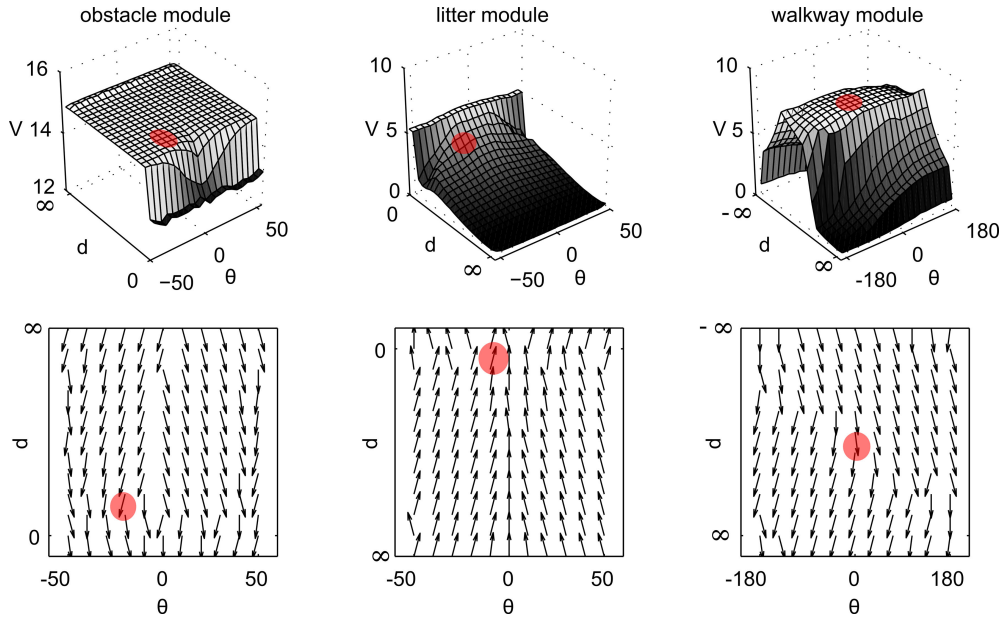


FIGURE 1 | Value functions (Top row) and their associated policies (Bottom row) for each of three modules. These functions for obstacle avoidance, litter collection, and sidewalk preference in left to right order have been learned by a virtual avatar walking along a sidewalk strewn with litter and obstacles. The red disk marks the state estimate uncertainty for each of them for a particular moment in the traverse.

where the subscript $\mathcal{M}(t)$ in $G_{\mathcal{M}(t)}$ denotes the modules that are *active* at time step t. In later formulae we abbreviate this as G_t for economy.

Evidence for modules

Direct measurements of brain activity provide a plethora of evidence that the segments in a task take the form of specialized modules. For example the Basal Ganglia circuitry shows specific neural circuits that respond to short components of a larger task (Schultz et al., 1997; Hikosaka et al., 2008). Moreover, embodied cognition studies provide much additional evidence. Studies of dual task performance provide evidence that separate task representations compete for shared resources, such as internal resources (Franco-Watkins et al., 2010) or eye gaze (Shinoda et al., 2001).

One compelling example is that of Rothkopf and Ballard (2009) that measures human gaze fixations during the navigation task that has separate trophic (picking up litter objects) and anti-trophic (avoiding obstacles) components. The overall setting is that of our virtual environment and uses identical litter and obstacle shapes that are only distinguished by object color. When picking up the object as litter, subjects' gaze is allocated to the center of the object, but when avoiding the same object, subjects' gaze fall on the objects' edges. **Figure 2** shows both of these cases. The inference is that when approaching an object, subjects use the expanding flow field to home in on it, but when avoiding an object, subjects use the different strategy of rotating about an edge of it. These two different strategies would be efficiently handled by individual task solutions, i.e., different modules, so that individual solutions can be learned and reused in combinations.

Modules and gaze arbitration

Another justification for independent modules is that they provide an elegant model for the disposition of gaze. Owing to the small visual angle of the human fovea, approximately 1°, gaze is not easily shared in servicing different tasks, and must be allocated amongst them. Arbitrating gaze requires a different approach than arbitrating control of the body. Reinforcement learning algorithms are best suited to handling actions that have direct consequences for a task. Actions such as eye fixations are difficult to put in this framework because they have only indirect consequences: they do not change the physical state of the agent or the environment; they serve only to obtain information.

A much better strategy than the straightforward RL protocol is to choose to use gaze to service the behavior that has *the most to gain* by being updated. The advantage of doing so is that uncertainty in the state information is reduced, leading to better policy choices. As time evolves, the uncertainty of the state of a module grows, introducing the possibility of low rewards. Deploying gaze to estimate that state more accurately reduces this risk, as shown in **Figure 3**.

Estimating the cost of uncertainty is equivalent to estimating the expected cost of incorrect action choices that result from uncertainty. Given that the Q-functions are known, and that Kalman filters can provide the necessary distributions over the state variables, it is straightforward to estimate this factor, loss_b, for each behavior b by sampling, using the following analysis. The loss value can be broken down into the losses associated with the uncertainty for each particular behavior b:

$$\text{loss}_b = E\left[\max_a \left(Q_b(s_b,a) + \sum_{i\in B, i\neq b} Q_i^E(s_i,a)\right)\right] - \sum_i Q_i^E(s_i,a_E). \quad (11)$$

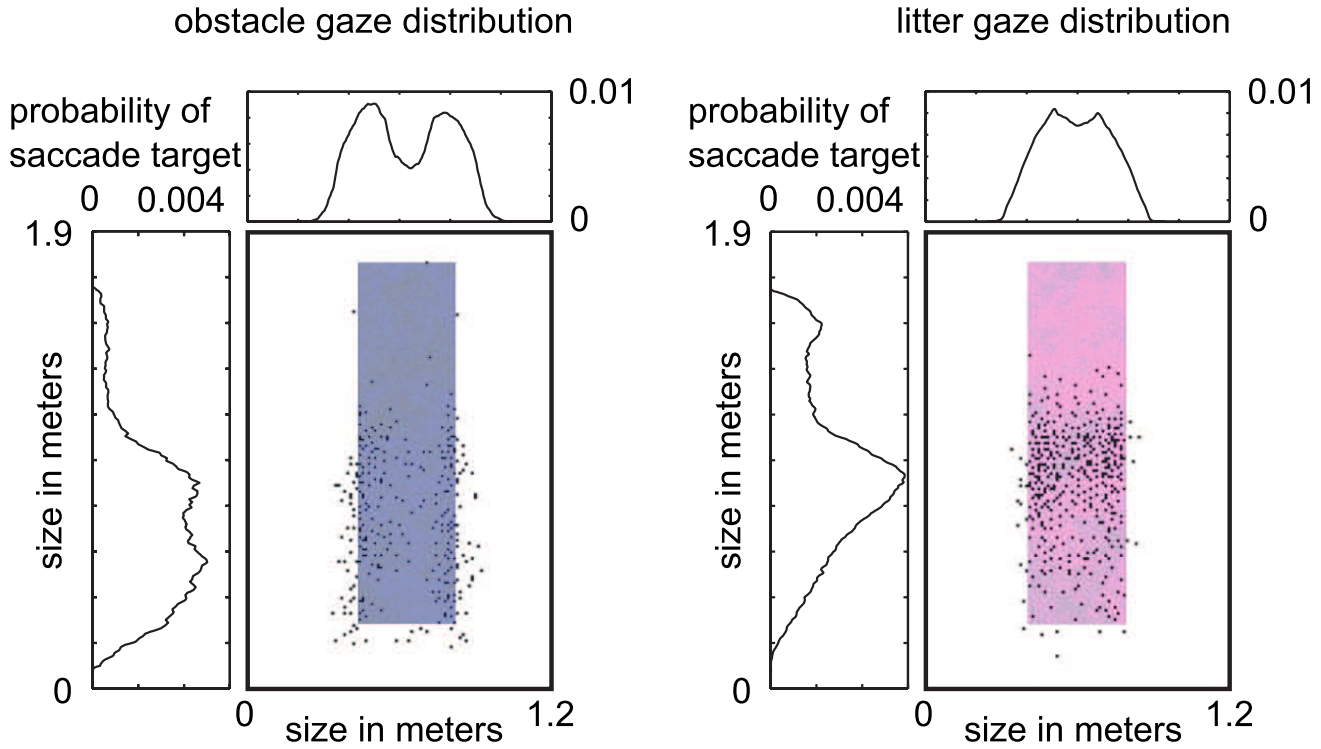


FIGURE 2 | Human gaze data for the same environment showing striking evidence for visual routines. Humans in the same environment as the avatar precisely manipulate gaze location depending on the specific task goal. The small black dots show the location of all fixation points on litter and obstacles. When avoiding obstacles (left) gaze points cluster at the edges of the object. When picking up a similar object (right) gaze points cluster on the center. From Rothkopf and Ballard (2009).

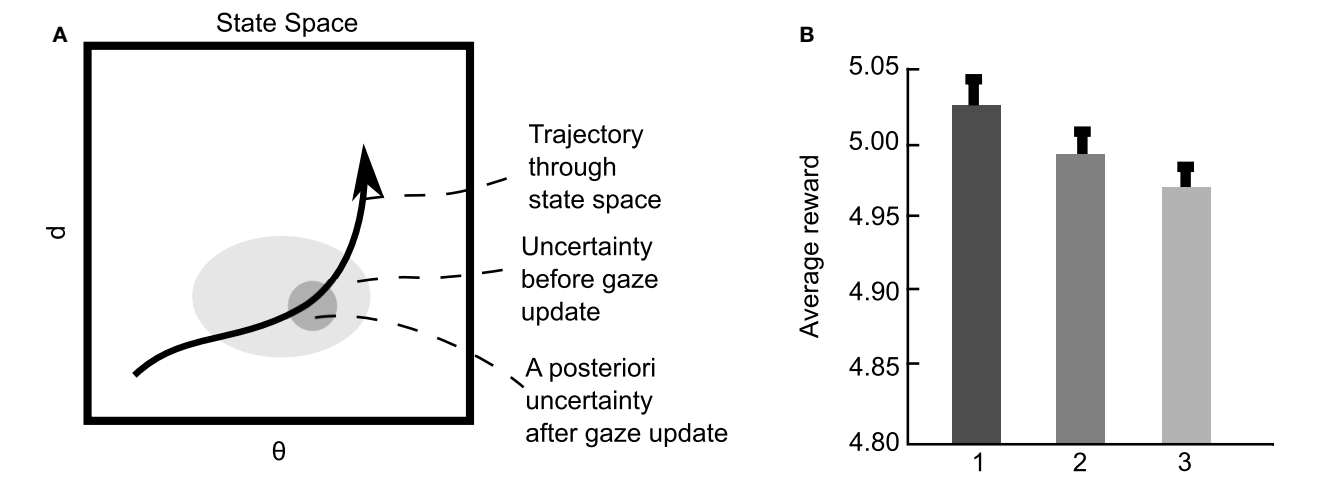


FIGURE 3 | Module-based gaze allocation. Modules compete for gaze in order to update their measurements. **(A)** A caricature of the basic method for a given module. The trajectory through the agent's state space is estimated using Kalman filter that propagates estimates in the absence of measurements and, as a consequence, build up uncertainty (large shaded area). If the behavior succeeds in obtaining a fixation, state space uncertainty is reduced (smaller shaded area). The reinforcement learning model allows the value of reducing uncertainty to be calculated. **(B)** The Sprague model out performs other models. Bars, left to right: Sprague model (1), round-robin (2), random selection (3).

Here, the expectation on the left is computed only over s_b. The value on the left is the expected return if s_b were known but the other state variables were not. The value on the right is the expected return if none of the state variables are known. The difference is interpreted as the cost of the uncertainty associated with s_b. The maximum of these values is then used to select which behavior should be given control of gaze.

The module activation protocol

Our central assumption is that an overall complex problem can be factored into a small set of MDPs, but any given factorization can only be expected to be valid for some transient period. Thus, the set of active modules is expected to change over time as the actions taken direct the agent to different parts of the composite state space. This raises two issues that we finesse: (1) How is a module activated? We assume that the sensory information provides a trigger as to when a module will be helpful. (2) How many modules can be active at a time? Extensive research on the capacity of humans to multi-task suggest that this number might be small, approximately four (Luck and Vogel, 1997). Taking both these constraints into consideration in our simulations, we use trigger features and use the value of four as a bound on the number of simultaneously active modules. Although this module activation protocol will allow the modules to learn as long as they sample their state-action spaces sufficiently often, there is still the question of how often to use it with respect to the SARSA algorithm. If it is used at every time step, the modules chosen will have little time to explore their problem spaces and adjust their Q-values. Thus for the length of module activation, we introduce the notion of an *episode* with an associated length parameter Δ (see **Figure 4**). In general this constraint should be soft as the module composition may have to be changed to deal with important environmental exigencies, but for our simulations we use a constant value. During each episode, only a subset of the total module set is active. The guiding hypothesis is that in the timecourse of behavior, a certain set of goals is pursued and therefore the corresponding modules that are needed to achieve these goals become active and those that correspond to tasks that are not pursued become inactive (Sprague et al., 2007).

In addition to the episode, we need two other assumptions:

1. The sum of the current estimates of the reward across an entire subset is accessible to each individual module in the subset at each moment by assumption;
2. The sampled subsets collectively must span the module space because the reward calculations demand this;

and the consequences of a module being activated are that:

1. It has used an associated procedure, such as a visual routine (Ullman, 1984; Ballard et al., 1997), to compute the initial state the module is in. In our examples we assume or supply a routine that does this;
2. Its Q-values are included in the sum indicated in Eq. 9 used to select an action, and
3. It influences the global reward that is handed out at every time step.

THE CREDIT ASSIGNMENT ALGORITHM

Each active module represents some portion of the composite state space and contributes through the selection of the composite action through Eq. 9, but without some additional constraint they only have access to a global performance measure, defined as the sum of the individual rewards collected by all of the $\mathcal{M}$ active modules at each time step:

$$G_t = \sum_{i \in \mathcal{M}} r_t^{(i)}. \tag{12}$$

The central problem that we tackle is how to learn the composite Q-values $Q^{(i)}(s^{(i)}, a)$ when only global rewards G_t are directly observed, but not the individual values $\{r_t^{(i)}\}$ (see **Figure 4**).

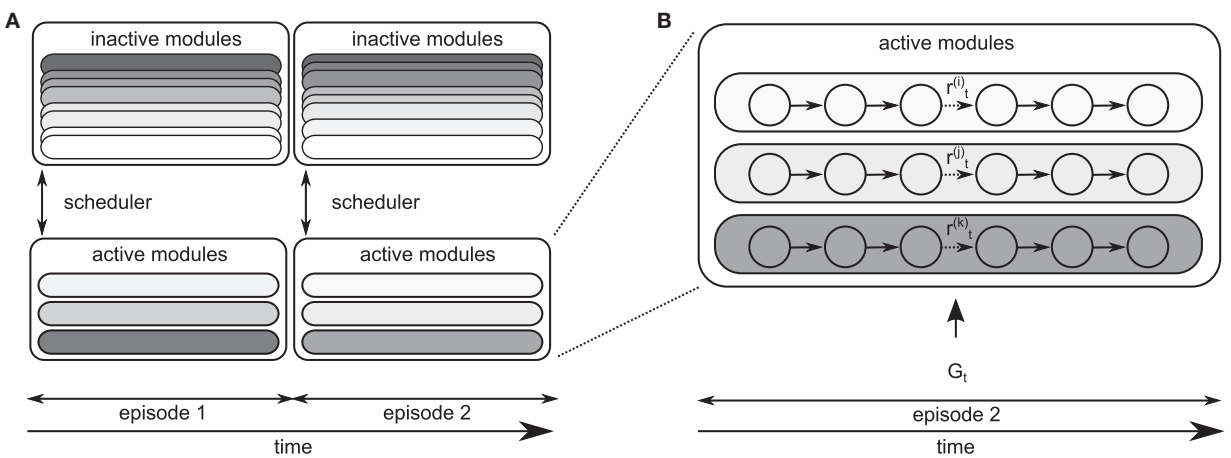


FIGURE 4 | Schematic representation of the modular credit assignment problem. (A) In any period during behavior there is only a subset of the total module set that is active. We term these periods episodes. In the timecourse of behavior, modules that are needed become active and those that are no longer needed become inactive. The vertical depicts two sequential episodes of three modules each, denoted with different shadings. The vertical arrows denote the scheduler's action in activating and deactivating modules. Our formal results only depend on each module being chosen sufficiently often and not on the details of the selection strategy. The same module may be selected in sequential episodes. **(B)** A fundamental problem for a biological agent using a modular architecture. At any given instant, shown with dotted lines, when multiple modules are active and only a global reward signal G is available, the modules each have to be able to calculate how much of the rewards is due to their activation. This is known as the credit assignment problem. Our setting simplifies the problem by assuming that individual reinforcement learning modules are independent and communicate only their estimates of their reward values.

The key additional constraint that we introduce is an assumption that the system can use the sum of current reward estimates from the modules that are co-active at any instant. This knowledge leads to the idea to use the different sets to estimate the difference between the total observed reward G_t and the sum of the current estimates of the individual rewards of the concurrently running behaviors. Credit assignment is achieved by bootstrapping these estimates over multiple task combinations, during which different subsets of behaviors are active. Dropping the temporal subscript for convenience, this reasoning can be formalized as requiring the individual behaviors to learn independent reward models $r^{(i)}(s^{(i)}, a)$. The current reward estimate for one particular behavior i, is obtained as

$$\hat{r}^{(i)} \leftarrow \hat{r}^{(i)} + \beta\delta_{r^{(i)}} \tag{13}$$

where the error on the reward estimates δ_r is calculated as the difference between the global reward and the sum of the component estimates:

$$\delta_{r^{(i)}} = G - \sum_{j\in\mathcal{M}} \hat{r}^{(i)} \tag{14}$$

so that Eq. 13 becomes:

$$\hat{r}^{(i)} \leftarrow \hat{r}^{(i)} + \beta\left(G - \sum_{j\in\mathcal{M}} \hat{r}^{(j)} \right)$$

which can be informatively rewritten as:

$$\hat{r}^{(i)} \leftarrow (1-\beta)\hat{r}^{(i)} + \beta\left(G - \sum_{j\in\mathcal{M}, j\neq i} \hat{r}^{(j)} \right) \tag{15}$$

To interpret this equation: Each module should adjust its reward estimate by a weighted sum of its own reward estimate and the estimate of its reward inferred from that of the other active modules. Together with the module activation protocol and Δ, Eq. 15 represents the core of our solution to the credit assignment problem. When one particular subset of tasks is pursued, each active behavior adjusts the current reward estimates $\hat{r}_i$ in the individual reward functions according to Eq. 15 at each time step. Over time, the set of tasks that have to be solved will change, resulting in a different set of behaviors being active, so that a new adjustment is applied to the reward functions according to Eq. 15. This bootstrapping process therefore relies on the assertion that the subsets of active behaviors visits all component behaviors.

The component Q-values for the state-action pairs of the individual behaviors are learned using the above estimates of the individual reward functions. Given the current reward estimates obtained through repeated application of Eq. 15, the SARSA algorithm is used to learn the component Q-functions:

$$Q_i\left(s_t^{(i)}, a_t^{(i)}\right) \leftarrow Q_i\left(s_t^{(i)}, a_t^{(i)}\right) + \alpha\delta_{Qi} \tag{16}$$

where δ_{Qi} now contains these estimates $\hat{r}_t^{(i)}$ and is given by:

$$\delta_{Q_i} = \hat{r}_t^{(i)} + \gamma Q_i\left(s_{t+1}^{(i)}, a_{t+1}^{(i)}\right) - Q_i\left(s_t^{(i)}, a_t^{(i)}\right) \tag{17}$$

The usage of an on-policy learning rule such as SARSA is necessary as noted in Sprague and Ballard (2003), because the arbitration process specified by Eq. 9 may select actions that are suboptimal for one or more of the modules. A feature of the SARSA algorithm is that it estimates the values of the policy that is actually used for control.

A concern one might have at this point is that since the rewards and the policies based on them are varying in separate algorithms, the net result might be that neither estimate converges. However it can be proved that this is not the case as long as $(k-1)\beta < 1$ where k is the maximum number of modules active at any one time. Furthermore convergence in the reward space is very rapid as shown by the simulations (Rothkopf and Ballard, 2010, (in press)).

Dealing with uncertainty

During the computation, the modules' MDPs are typically in different states of completion and consequently have different levels of uncertainty in their reward estimates. Unfortunately if Eq. 15 is used with a single fixed β value, this means, that on a particular task combination, all component behaviors will weight reward estimates in the same way, independent of how well component behaviors have already estimated their share. Thus a drawback of the fixed β updating scheme is that it is possible for a behavior to unlearn good reward estimates if it is combined with other behaviors whose reward estimates are far from their true values.

The problem of combining different modules' reward estimates that have different states of uncertainty can be fixed by considering the respective uncertainties in the estimates of the respective rewards separately. Thus one can have individual β_i values for each module reflect their corresponding reward estimates of uncertainty values. Assuming that the between-module fluctuations are uncorrelated, one can express the gain for each reward estimate in terms of the individual uncertainties in the respective reward estimates $(\sigma^{(i)})^2$:

$$\begin{aligned}\beta_i &= \frac{\left(\sigma^{(i)}\right)^2}{\sum_{j=1}^{N}\left(\sigma^{(j)}\right)^2} \\ &= \frac{\left(\sigma^{(i)}\right)^2}{\sum_{j\neq i}^{N}\left(\sigma^{(j)}\right)^2 + \left(\sigma^{(i)}\right)^2}\end{aligned} \tag{18}$$

where the last term in the denominator is variance in the observation noise.

Expanding the sum in the denominator in Eq. 18 suggests the second approximation, in which each individual module uses an on-line estimate for the variance $\Sigma_{j\neq i}^{N}(\sigma^{(j)})^2$ by tracking the variance in the difference between the global reward G and the sum of the reward estimates of the other modules $\Sigma_{j\neq i}\hat{r}^{(j)}$. In the following simulations each module tracked this difference using a recursive least squares estimator with exponential forgetting and maintained the uncertainty about the rewards of individual state-action pairs $(\sigma^{(i)})^2$ locally[4].

[4]The exponential forgetting factor of the recursive least squares estimator is chosen so that it reflects the timescale of the switching of the behaviors. The idea is that whenever the composition of tasks is changed the estimates will also change, because the sum of the estimates will change with the set of learners. The equation that is used is: $\lambda = 1 - (1/(\nu^{2/3}))$, where λ is the forgetting factor and ν is the time order of the sequence. For the simulations presented, the time order was established as the expected value of the number of iterations over which an individual module is switched on.

RESULTS

We demonstrate the algorithm on two separate problems. The first is a classic predator and food source problem that uses 15 different food sources and 5 predators. The second is the multi-tasking problem we described earlier of an agent in a simulated three-dimensional world walking on a sidewalk while avoiding obstacles and picking up litter (Sprague and Ballard, 2003). For all these simulations, the RL learning parameter α was 0.1. The first experiment uses both constant β values from the set {0.01, 0.1, 0.2, 0.5} as well as the variance weighted β computed according to Eq. 18. The experiment involving learning to navigate along a path while avoiding obstacles and approaching targets uses a constant β value of 0.1.

MODULE SELECTION IN THE FACE OF MULTIPLE REWARD SOURCES

This problem is described in Singh and Cohn (1998) where the authors explore the use of multiple tasks in a grid-world problem. This single-agent problem comes close to representing a problem that would have to be addressed by a biological agent since the action space is shared by the modules.

In the original formulation, an agent moves on a 5×5 grid. Possible actions move the agent in the eight compass directions. Moves at the edge of the grid-world which would result in the agent leaving the grid result in the agent staying in the current position. The grid is populated by three food items and one predator. The picking up of a food item results in a reward of one unit and the repositioning of the food item to a new and empty location. The world is also populated by a predator, which moves every other time unit toward the current position of the agent. The agent obtains a reward of 0.5 units for every time step during which it does not collide with the predator. Each learner represents the position of the respective food item or predator, i.e., there are 625 states for each of the component learners, and a total of four learners were always active in order to solve the four component tasks (see **Figure 5**).

Previously Singh and Cohn (1998) and Sprague and Ballard (2003) used this task in multi-goal learning but both studies used individual rewards that were delivered for each task as separate reward signals.

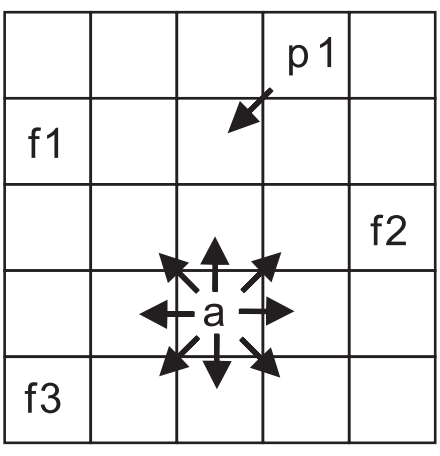


FIGURE 5 | Predator-prey grid-world example following SinghCohn,1998. An agent is located on a 5x5 grid and searches to find three different food sources f1 to f3 and tries to avoid the predator p, which moves every other time step toward the agent.

Here the problem was modified so that the reward each behavior sees is only the global sum of the individual rewards. Furthermore, at the beginning of each episode, three food sources are selected randomly according to a uniform distribution over a total of 15 different food sources. Similarly, at the beginning of each episode, one predator is selected randomly from a pool of five different predators according to a uniform distribution, so that during every episode a total of three food sources and one predator are present, as in the original problem. The Δ for each episode is 50 iterations.

Simulations were run for different values of β and with a variance weighted β for comparison. The rewards for all foods and predators were set to the values of the original Singh and Cohn problem (Singh and Cohn, 1998) in order to be able to compare the present results with the original problem formulation, which allowed a maximum average reward per episode of four units. **Figure 6** shows the average reward earned at each time step and demonstrates the improvement over learning as well as the superiority in the speed of acquiring maximal reward for the variance weighted learner.

Figure 6 furthermore demonstrates that for intermediate learning rates for β between 0.01 and 0.2, the reward estimates approach the true reward values and similarly the error in all computed value functions decreases. This is assessed by computing the RMS error between all reward estimates and the true rewards, as well as the RMS error between the true and learned value functions over time. By contrast, a learner with a learning rate of $\beta = 0.5$ does not converge on the correct reward model over the course of the simulations. Accordingly, this learner does neither approach the correct value function as shown in **Figure 6** nor approach the average reward collected by the other learners. Again, the variance weighted learner is able to learn the reward model faster than the learners using a constant learning rate β so that the error in both the reward estimates as well as in the value function decrease fastest.

LEARNING WALKWAY NAVIGATION IN A VIRTUAL 3D ENVIRONMENT

This problem uses a humanoid agent navigating in a realistic virtual reality environment that has three dimensionality. The agent uses simulated vision to compute features from the environment that define each module's state space. Also the agent's discrete state spaces must guide it successfully through the much more fine-grained environment. The walkway navigation task was first considered by Sprague et al. (2007) where a factorized solution was presented. However, that solution was obtained by delivering each of the individual learners their respective reward; that is, the agent received three separate rewards, one for the walkway following module, one for the obstacle avoidance module, and one for the litter picking up module. This problem was re-coded here but with the additional constraint of only global reward being observed by all modules in each task combination. The global reward was always the linear sum of the rewards obtained by the individual modules according to Eq. 12.

The parameterization of the statespace is shown in **Figure 7**. Each module represents the states with a two-dimensional vector containing a distance and an angle. For the picking up and the avoidance behaviors, these are the distance to the closest litter object and obstacle respectively and the signed angle between the current heading direction and the direction toward the object. The distance is scaled logarithmically similarly to the original setup (Sprague et al., 2007) and the resulting distance d_i is then discretized into 21

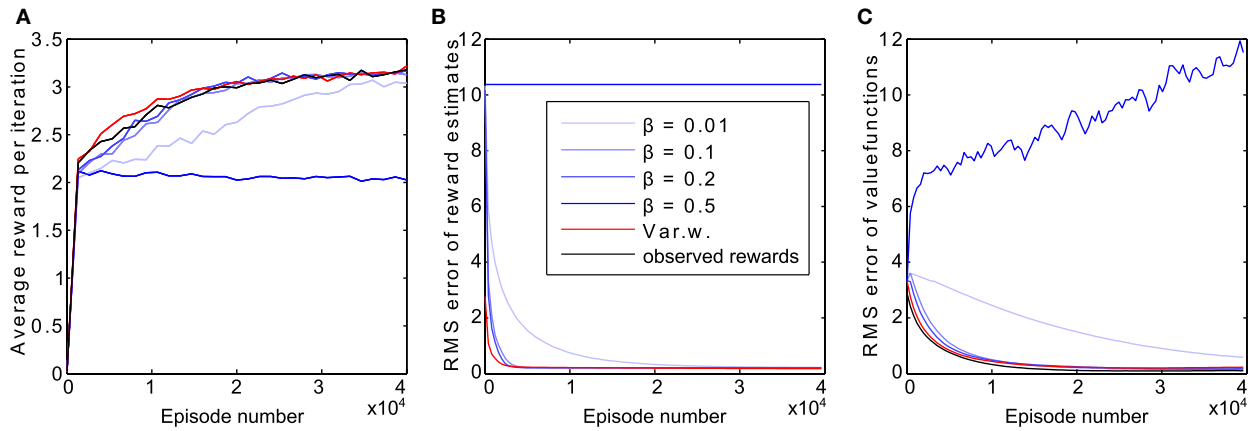


FIGURE 6 | Comparison of learning progress. (A) Effect of different learning rates and the variance weighted learning ("Var.w.") on the accumulated reward over episodes for the simulated foraging agent for the case of only knowing global reward Gt. For comparison, a case where the rewards given to each module were known is shown as the "observed rewards" trace. **(B)** Root mean squared error error between the true rewards and the reward estimates of all behaviors over episodes. The three curves correspond to the different learning rates β in Eq. 15. **(C)** Root mean squared error between the true value function and the learned value functions of all behaviors over trials.

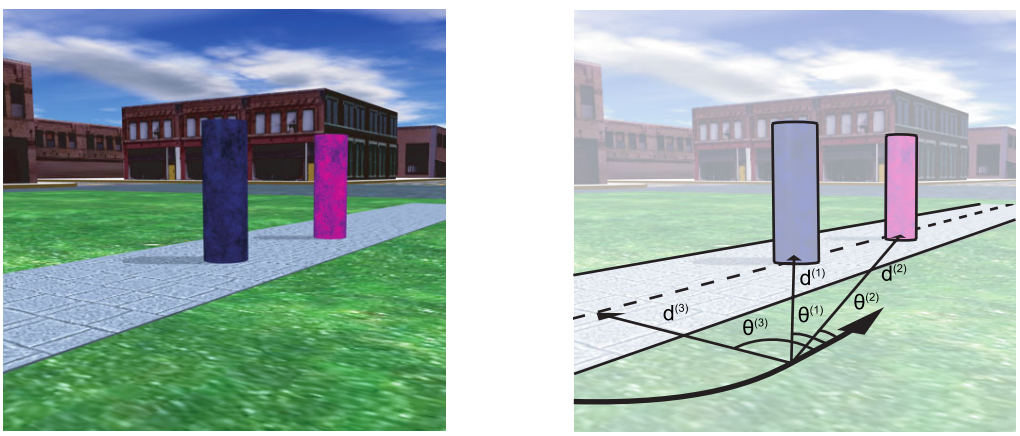


FIGURE 7 | The walkway navigation tasks. Left: typical view from the agent while navigating along the walkway. The three possible tasks are following the walkway, avoiding obstacles, which are the dark cylinders, and picking up litter, corresponding to the light cylinders. Right: Schematic representation of the statespace parameterization for the learners. Each module represents the distance to the closest object in the field of view and the angle between the current heading direction and the object center axis. The module learning the walkway behavior uses the signed distance to the midline of the walkway and the angle between the heading direction and the vector in the direction of the walkway.

possible values between 0 and infinite distance. The angles within the field of view, i.e., with a magnitude smaller than 50° are similarly discretized to 21 values. The walkway statespace is slightly different from Sprague et al. (2007) in that it represents all positions of the agent relative to the walkway for all possible walking directions. Finally, instead of 3 possible actions as in Sprague et al. (2007) the current simulations use 5 actions corresponding to steering at one of the five angles $\{-15, -7.5, 0, 7.5, 15\}$ with additive Gaussian noise of variance $\sigma^2 = 1$. To learn policies and Q-values, different subsets of modules were selected for different episodes and the correct global reward supplied for each individual subset.

The basic time unit of computation was chosen to be 300 ms, which is the average duration of a fixational eye movement. Human subjects took an average duration of 1 min an 48 s to carry out these tasks, which is approximately 325 intervals of 300 ms. Therefore, each episode consists of $\Delta = 325$ discrete time steps. At the beginning of each episode it is determined which tasks have high priority. During each episode, it is equally probable that either two or three tasks are pursued. For each episode between 35 and 40 obstacles are used, together with a similar number of litter objects.

The reward values displayed as a function of the state space locations are shown in **Figure 8A**. Starting from random values and receiving only global reward at each step, the agent's modules are able to arrive at good estimates of the true reward. The accuracy of these estimates is shown in **Figure 8B**.

The value functions and policies of these simulations are shown in **Figure 9**, at both the first iteration with random initial values and after learning, when the agent has walked the walkway for 1000 episodes.

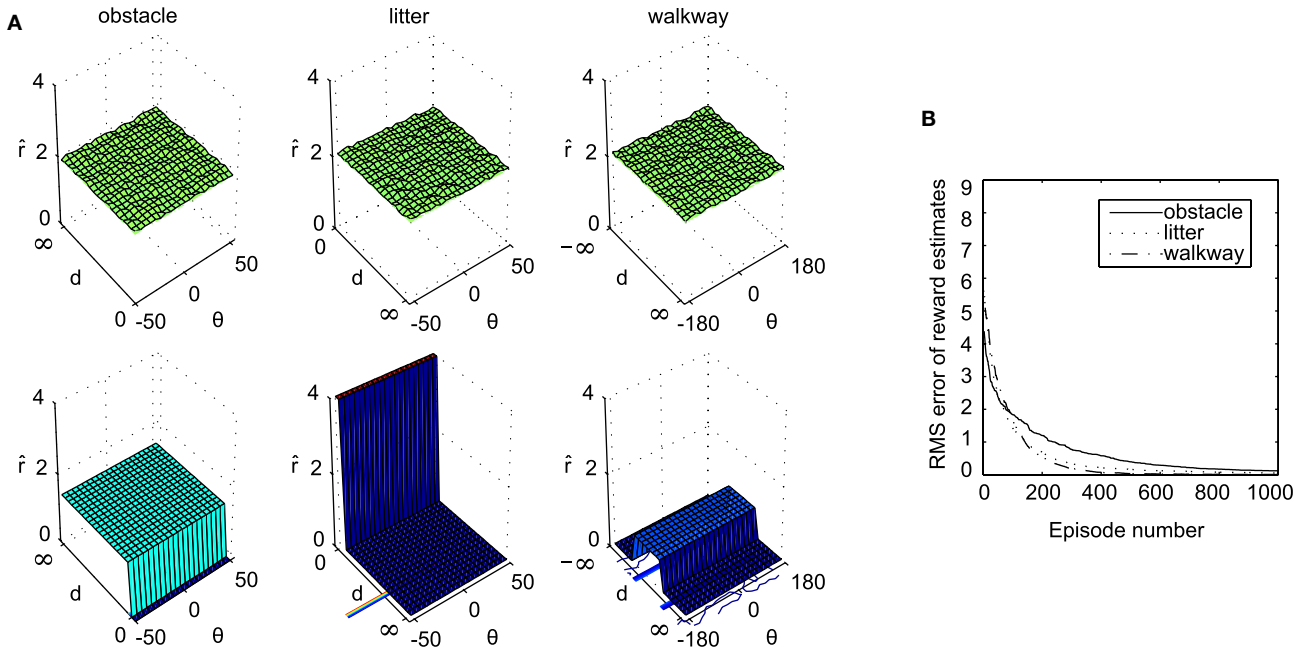


FIGURE 8 | Reward calculations for the walkway navigation task for the three component behaviors using the credit assignment algorithm. (A) Top row: Initial estimates of the reward functions. Bottom row: Final reward estimates. **(B)** Time course of learning reward functions for each of the three component behaviors. RMS error between true and calculated reward as a function of iteration number.

As can be seen from the representation of the reward estimates, the individual behaviors have learned the true rewards of their respective tasks, where not intersecting with an obstacle results in a reward of one unit, intersecting a litter object gives four units of reward, and staying on the walkway results in a reward of 0.8 units. The figures of the reward estimates also demonstrate that a function approximation scheme should be better at capturing structure in the reward space such as smooth reward landscapes, reward functions with only one state being rewarded, or separate areas with discrete rewards.

DISCUSSION

The primary contribution of this paper is to describe a way that individual task solutions can be learned by individual modules with independent state variables while pursuing multiple goals and observing only the global reward. The proposed method relies on the agent carrying out multiple task combinations over time, which enables the correct learning of individual rewards for the component tasks. Accordingly, carrying out multiple concurrent task combinations is not a complication but enables learning about the rewards associated with individual tasks. The key constraints, motivated by the need for a system that would potentially scale to a large library of behaviors, are (1) the overall system must be structured such that the system could achieve its goals by using only a subset of its behavioral repertoire at any instant, (2) the reward gained by this subset is the total of that earned by its component behaviors, and (3) the modules must be used in linearly independent combinations. The use of modules allows the rewards obtained by reinforcement to be estimated on-line. In addition this formulation lends itself to use the uncertainties in current reward estimates for combining them amongst modules, which speeds convergence of the estimating process.

The linear independence constraint (the second additional assumption in the module activation protocol) is important as without it, the Q-values and their corresponding value functions V cannot be correctly computed. To see this, note that adding a constant to the reward function does not change the policy (Ng et al., 1999) but changes the value function. When γ is near unity, a small additive constant c into the reward results in a large difference between the corresponding value function V' and the original V as shown by the following:

$$V'^{\pi}(s) = \frac{c}{1-\gamma} + V^{\pi}(s)$$

Thus, although it may be possible to learn some of the policies for component modules for one particular task combination, the value functions will be corrupted by a large bias, which will be especially problematic when new task combinations are to be solved. The reward estimates will be biased such that they have to be relearned, but will again be biased.

In our venue small numbers of behaviors that are appropriate for the current situation are selected in an on-line faction. In this situation it is essential to get the Q-values right. An algorithm that models other modules' contribution as pure noise will compute the correct policy when *all* the behaviors/agents are active but this result will not extend to active *subsets* of modules and behaviors because incorrect Q-values, when used in subsets, will cause chaotic behavior.

Considering the demonstrations, the closest work to our walkway simulation would be that of Warren and Fajen (2004) who was the first to quantitatively address the question of human dynamic trajectories in such a venue. The trajectories generated

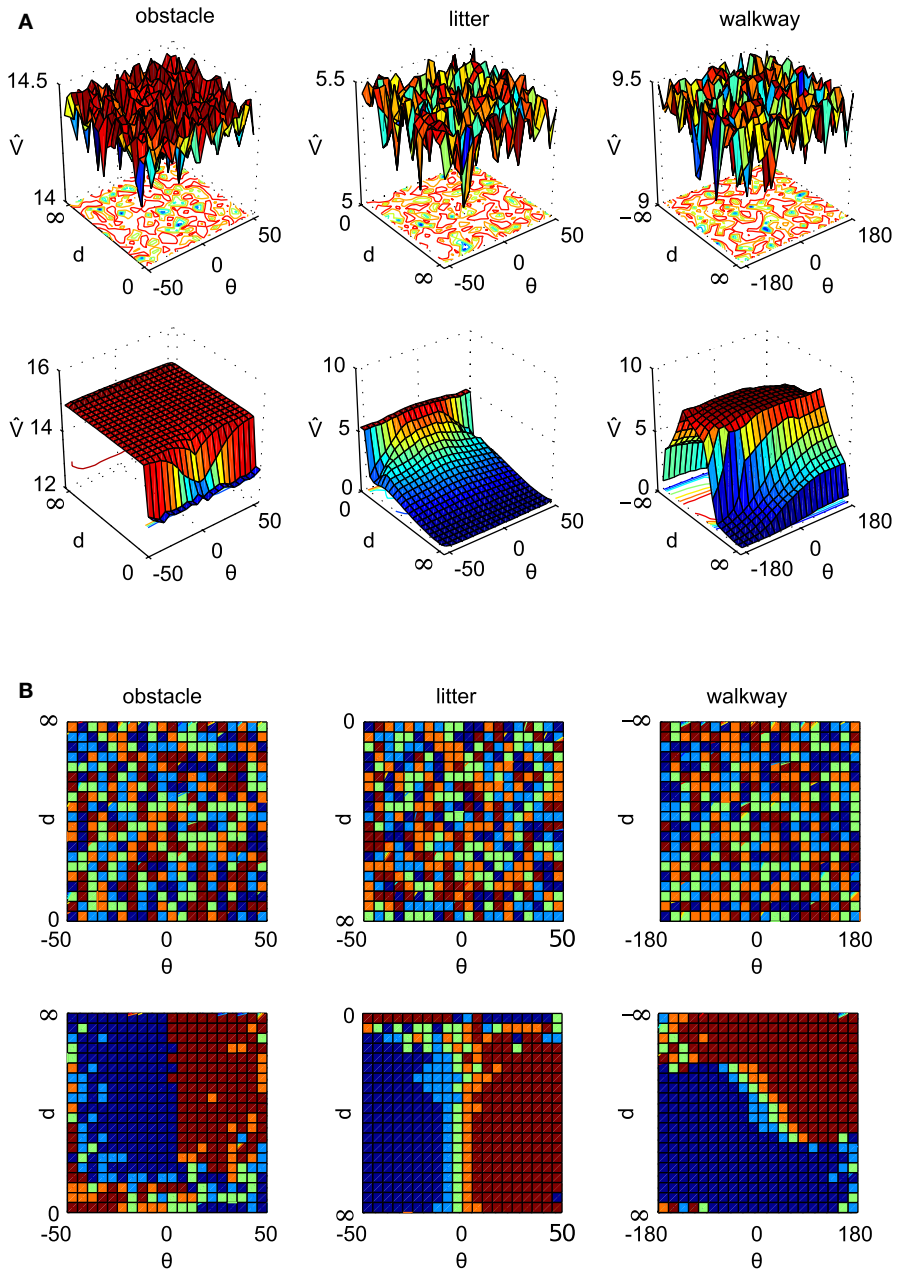


FIGURE 9 | Representations of value functions and policies in the walkway navigation task for the three component behaviors. (A) Top row: initial value function estimates $\hat{V}(s)$. Bottom row: final value estimates. **(B)** Representations of policies. Top row: initial policy estimates $\hat{\pi}(s)$. Bottom row: final policy estimates. The navigation actions are coded as follows: left turns are red, straight ahead is light green, right turns are blue.

by both models are qualitatively similar, but Warren's are curve fit to underlying differential equations and so far are not connected with concepts of reward. In principal it is easy to show differences by having objects to be picked up with different reward values. Warren's formalism has no way of expressing this so all the data for attractor and repulser objects would have to be refit.

In its formalism, the present work is related to earlier approaches that start out with compositional solutions to individual problems and then devise methods in order to combine a large number of such elemental solutions (e.g., Meuleau et al., 1998; Singh and Cohn, 1998). Both approaches are concerned with learning solutions to large MDPs by utilizing solutions or partial solutions to smaller component MDPs. In Meuleau et al. (1998) the solutions to such components are heuristically combined to find an approximate solution to the composite MDP by exploiting assumptions on the structure of the joint action space. A way of learning a composite MDP from individual component MDPs by merging has been described in Singh and Cohn (1998). However, the composite problem is solved in a more ad hoc way using bounds on the state values derived from the state values of the individual component MDPs.

Attempts to overcome the scaling problem in more elegant ways than *ab initio* factoring try to exploit inherent structure in the problem (Dayan and Hinton, 1992; Parr and Russell, 1997; Sutton et al., 1999; Barto and Mahadevan, 2003; Vigorito and Barto, 2010). Factoring can use graphical models that express conditional independencies can reduce the size of the variables necessary for a full description of the problem at hand (Boutilier and Goldszmidt, 2000; Guestrin et al., 2003). The approach by Sallans and Hinton (2004) can also be conceptualized as exploiting the statistical structures of the state and action spaces. Doya et al. (2002) and Samejima et al. (2003) use a number of actor-critic modules and learn a linear combination of the controllers for the local approximation of the policy. All these methods constitute advances and our method is extensible to them to the extent that they can be encapsulated into modules that are made explicit and the issues related to module activation are addressed.

The credit assignment problem is an important problem in embodied cognition as a behaving animal in a complex environment has to solve this problem. While common laboratory experiments consider mostly single tasks with reward being unambiguously associated with the single task component, in the natural environment multiple concurrent goals have to be solved and the contributions of the individual actions to the total observed reward have to be learned. In this setting, the ability to assign credit correctly may confer an additional advantage. When the brain encodes new behaviors, it needs to know their values. Since the brain uses its own internally generated secondary reward signals such as dopamine to estimate these values, there is the delicate issue of how to keep the overall system in calibration. The credit assignment algorithm suggests a partial solution: by using global reward and concurrent subsets of active modular behaviors, the rewards can at least be held to a global consistency.

ACKNOWLEDGMENTS

The research reported herein was supported by NIH Grants RR009283, EY019174 and MH060624. Constantin A. Rothkopf was additionally supported by EC MEXT-project PLICON and by the German Federal Ministry of Education and Research within the Bernstein Focus: Neurotechnology research grant.

REFERENCES

Anderson, J. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Arkin, R. (1998). *Behavior Based Robotics*. Cambridge, MA: MIT Press.

Badler, N., Palmer, M., and Bindiganavale, R. (1999). Animation control for real-time virtual humans. *Commun. ACM* 42, 64–73.

Badler, N. I., Phillips, C. B., and Webber, B. L. (1993). *Simulating Humans: Computer Graphics Animation and Control*. New York, NY: Oxford University Press.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., and Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behav. Brain Sci.* 20, 723–767.

Barsalou, L. (2009). Simulation, situated conceptualization, and prediction. *Phil. Trans. R. Soc. B* 364, 1281–1289.

Barto, A. G., and Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dyn. Syst.* 13, 41–77.

Bonasso, R. P., Firby, J. R., Gat, E., Kortenkamp, D., Miller, D. D. P., and Slack, M. G. (1997). Experiences with an architecture for intelligent, reactive agents. *J. Exp. Theor. Artif. Intell.* 9, 237–256.

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE J. Robot. Autom.* 2, 14–23.

Bryson, J. J., and Stein, L. A. (2001). "Modularity and design in reactive intelligence," in *International Joint Conference on Artificial Intelligence*, Seattle, WA.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879.

Dayan, P., and Hinton, G. E. (1992). "Feudal reinforcement learning," in *Advances in Neural Information Processing Systems,* (San Francisco, CA: Morgan Kaufmann Publishers Inc.) 5, 271–278.

Dearden, R. Boutilier, C., and Goldszmidt, M. (2000). Stochastic dynamic programming with factored representations. *Artif. Intell.* 121, 49–107.

Doya, K., Samejima, K., Katagiri K., and Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Comput.* 14, 1347–1369.

Firby, R. J., Kahn, R. E., Prokopowicz, P. N., and Swain, M. J. (1995). "An architecture for vision and action," in *International Joint Conference on Artificial Intelligence* (Montreal, Canada; Morgan Kaufmann Publishers Inc.), 72–79.

Franco-Watkins, A. M., Rickard, T. C., and Pashler, H. (2010). Taxing executive processes does not necessarily increase impulsive decision making. *Exp. Psychol.* 57, 193–201.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 586–596.

Gottfried, J. A., O'Doherty, J., and Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 301, 1104–1107.

Guestrin, C. E., Koller, D., Parr, R., and Venkataraman, S. (2003). Efficient solution algorithms for factored MDPs. *J. Artif. Intell. Res.* 19, 399–468.

Haruno, M., and Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J. Neurophysiol.* 95, 948–959.

Hikosaka, O., Bromberg-Martin, E., Hong, S., and Matsumoto, M. (2008). New insights on the subcortical representation of reward. *Curr. Opin. Neurobiol.* 18, 203–208.

Humphrys, M. (1996). "Action selection methods using reinforcement learning," in *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, eds P. Maes, M. Mataric, J.-A. Meyer, J. Pollack, and S. W. Wilson (Cambridge, MA: MIT Press, Bradford Books), 135–144.

Karlsson, J. (1997). *Learning to Solve Multiple Goals*. PhD thesis, University of Rochester, Rochester.

Laird, J. E., Newell, A., and Rosenblum, P. S. (1987). Soar: an architecture for general intelligence. *Artif. Intell.* 33, 1–64.

Langley, P., and Choi, D. (2006). Learning recursive control programs from problem solving. *J. Mach. Learn. Res.* 7, 493–518.

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.

Meuleau, N., Hauskrecht, M., Kim, K.-E., Peshkin, L., Kaelbling, L., Dean, T., and Boutilier, C. (1998). "Solving very large weakly coupled Markov decision processes," in *AAAI/IAAI*, 165–172.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., and Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* 9, 1057–1063.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Ng, A. Y., Harada, D., and Russell, S. (1999). "Policy invariance under reward transformations: theory and application to reward shaping," in *Proceedings of the Sixteenth International Conference on Machine Learning*, Bled, Slovenia.

Parr, R., and Russell, S. (1997). "Reinforcement learning with hierarchies of machines," in *Advances in Neural Information Processing Systems*, M. I. Jordan, M. J. Kearns, and S. A. Solla. (Cambridge, MA: MIT Press), 1043–1049.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., and Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442, 1042–1045.

Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* 20, 262–270.

Ritter, S., Anderson, J. R., Cytrynowicz, M., and Medvedeva, O. (1998). Authoring content in the pat algebra tutor. *J. Interact. Media Educ.* 98, 1–30.

Rothkopf, C. A., and Ballard, D. H. (2009). Image statistics at the point of gaze during human navigation. *Vis. Neurosci.* 26, 81–92.

Rothkopf, C. A., and Ballard, D. H. (2010). "Learning and coordinating reper-

toires of behaviors: credit assignment and module activation," in *Intrinsically Motivated Cumulative Learning in Natural and Artificial Systems,* Eds G. Baldassarre and M. Mirolli (in press).

Roy, D. K., and Pentland, A. P. (2002). Learning words from sights and sounds: a computational model. *Cogn. Sci.* 26, 113–146.

Rummery, G. A., and Niranjan, M. (1994). "On-line Q-learning using connectionist systems," Technical Report CUED/F-INFENG/TR 166, Cambridge University Engineering Department, Cambridge.

Russell, S., and Zimdars, A. (2003). "Q-decomposition for reinforcement learning agents," in *Proceedings of the International Conference on Machine Learning*, Washington, DC.

Sallans, B., and Hinton, G. E. (2004). Reinforcement learning with factored states and actions. *J. Mach. Learn. Res.* 5, 1063–1088.

Samejima, K., Doya, K., and Kawato, M. (2003). Inter-module credit assignment in modular reinforcement learning. *Neural Netw.* 16, 985–994.

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.

Shinoda, H., Hayhoe, M. M., and Shrivastava, A. (2001). What controls attention in natural environments? *Vis. Res.* 41, 3535–3546.

Singh, S., and Cohn, D. (1998). How to dynamically merge Markov decision processes. *Neural Inf. Process. Syst.* 10, 1057–1063.

Sprague, N., and Ballard, D. (2003). "Multiple-goal reinforcement learning with modular sarsa(0)," in *International Joint Conference on Artificial Intelligence*, Acapulco.

Sprague, N., Ballard, D., and Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Trans. Appl. Percept.* 4, 1–23.

Sun, R. (2006). *Cognition and Multi-Agent Interaction*, chapter 4. Cambridge: Cambridge University Press, 79–99.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Sutton, R. S., Precup, D., and Singh, S. P. (1999). Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning. *Artif. Intell.* 112, 181–211.

Terzopoulos, D. (1999). Artificial life for computer graphics. *Commun. ACM* 42, 32–42.

Terzopoulos, D., Tu, X., and Grzeszczuk, R. (1994). Artificial fishes: autonomous locomotion, perception, behavior, and learning in a simulated physical world. *Artif. Life*, 1, 327–351.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208.

Treisman, A. M. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.

Trick, L. M., and Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol. Rev.* 101, 80–102.

Ullman, S. (1984). Visual routines. *Cognition* 18, 97–157.

Vigorito, C. M., and Barto, A. G. (2010). Intrinsically motivated hierarchical skill learning in structured environments. *IEEE Trans. Auton. Ment. Dev.* 2, 132–143.

Warren, W. H., and Fajen, B. R. (2004). Behavioral dynamics of human locomotion. *Ecol. Psychol.* 16, 61–66.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards*. PhD thesis, University of Cambridge, Cambridge.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 July 2010; paper pending published: 23 August 2010; accepted: 28 September 2010; published online: 22 November 2010.

Citation: Rothkopf CA and Ballard DH (2010) Credit assignment in multiple goal embodied visuomotor behavior. Front. Psychology ***1****:173. doi: 10.3389/fpsyg.2010.00173*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Rothkopf and Ballard. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 08 December 2010
doi: 10.3389/fpsyg.2010.00221

A computational model of the lexical-semantic system based on a grounded cognition approach

Mauro Ursino*, Cristiano Cuppini and Elisa Magosso

Department of Electronics, Computer Science and Systems, University of Bologna, Bologna, Italy

Edited by:
Anna M. Borghi, University of Bologna and Institute of Cognitive Sciences and Technologies, Italy

Reviewed by:
Ken Mcrae, University of Western Ontario, Canada
Angelo Cangelosi, University of Plymouth, UK

****Correspondence:***
Mauro Ursino, Department of Electronics, Computer Science and Systems, University of Bologna, Viale Risorgimento 2, I40136 Bologna, Italy.
e-mail: mauro.ursino@unibo.it

This work presents a connectionist model of the semantic-lexical system based on grounded cognition. The model assumes that the lexical and semantic aspects of language are memorized in two distinct stores. The semantic properties of objects are represented as a collection of features, whose number may vary among objects. Features are described as activation of neural oscillators in different sensory-motor areas (one area for each feature) topographically organized to implement a similarity principle. Lexical items are represented as activation of neural groups in a different layer. Lexical and semantic aspects are then linked together on the basis of previous experience, using physiological learning mechanisms. After training, features which frequently occurred together, and the corresponding word-forms, become linked via reciprocal excitatory synapses. The model also includes some inhibitory synapses: features in the semantic network tend to inhibit words not associated with them during the previous learning phase. Simulations show that after learning, presentation of a cue can evoke the overall object and the corresponding word in the lexical area. Moreover, different objects and the corresponding words can be simultaneously retrieved and segmented via a time division in the gamma-band. Word presentation, in turn, activates the corresponding features in the sensory-motor areas, recreating the same conditions occurring during learning. The model simulates the formation of categories, assuming that objects belong to the same category if they share some features. Simple exempla are shown to illustrate how words representing a category can be distinguished from words representing individual members. Finally, the model can be used to simulate patients with focalized lesions, assuming an impairment of synaptic strength in specific feature areas.

Keywords: gamma-band oscillations, synchronization, Hebbian rules, word production, word recognition, object recognition, lexical deficit, categorization

INTRODUCTION

Traditional theories on human language and semantic memory assume that cognition consists in the manipulation of abstract symbols, separate from the modal system for perception and action. More specifically, word meaning in these theories is often represented as a vector in a multidimensional space, with the elements of this vector consisting of abstract features. This traditional point of view, however, fails to take into account that the meaning of concrete objects is strongly grounded in daily experience and exploits the perceptual modalities.

Recently, various theories in different domains (including linguistics, psychology, cognitive neuroscience, and robotics) rejected the idea that semantic memory merely relies upon abstract symbols, and emphasized the importance of concrete experience in the formation and retrieval of object meaning. "Grounded cognition" or "embodied cognition" (Gibbs, 2003; Barsalou, 2008; Borghi and Cimatti, 2010) assumes that our concepts consist of "perceptual symbols," and that the retrieval of concept meaning is a form of re-activation of past sensory-motor and introspective experience.

Neuroimaging studies support the idea that simulation of past experience plays a pivotal role in conceptual processing. Dealing with information on food or smell activates gustatory and olfactory areas, respectively (Simmons et al., 2005; Gonzalez et al., 2006). Regions of the posterior temporal cortex involved in object representation become active during conceptual processing of pictures and words, as well as during auditory sentence comprehension (Davis and Johnsrude, 2003; Giraud et al., 2004; Rodd et al., 2005). Answering questions concerning visual, auditory, tactile, or taste properties activates regions involved in each of these modalities (Goldberg et al., 2006). Reading action words activates regions in the premotor cortex that are active when the subject performs the corresponding movement (Hauk et al., 2004). These results support the idea that conceptual processing is largely based on simulation of past experience, and that this experience provides part of the fundamental grounding for the construction of semantic memory and the use of language (the interested reader can find more details in excellent review papers on the subject such as Martin and Chao, 2001; Martin, 2007; Barsalou, 2008).

Several qualitative theories of semantic memory proposed in recent years may be reconciled with the grounded cognition viewpoint (see Hart et al., 2007). Most of them assume that a concept is described in a semantic memory as a collection of sensory-motor features, which spread over different cortical areas (Warrington and Shallice, 1984; Warrington and McCarthy, 1987; Caramazza and Shelton, 1998; Gainotti, 2000; Humphreys and Forde, 2001). These models are motivated by results on

neurologically damaged patients showing selective impairment in category representation (for instance in recognition of animate vs. inanimate items).

Among others, a few conceptual models merit a brief description due to their analogy with some aspects of the model presented herein. Tyler and collaborators (Tyler et al., 2000; Tyler and Moss, 2001) developed a model named "Conceptual Structure Account." This model assumes that semantic memory is organized as a distributed network of features. An important aspect is that objects may share certain features: common features would indicate a category membership, whereas members in the same category can be distinguished by distinctive features. Hence, peculiar features are essential to identify individual objects in the same category.

Barsalou and Simmons (Barsalou et al., 2003; Simmons and Barsalou, 2003) proposed a model (named the "Conceptual Topography Theory") in which features are organized according to a topographical principle: the spatial proximity of neurons reflects similarity in the features they encode. According to this model, groups of features send their information to other convergence neurons, which replicate the concept of "convergence zone" originally proposed by Damasio (1989).

Hart et al. (2002) and Kraut et al. (2004) proposed the "Neural Hybrid Model of Semantic Object Memory" in which different regions encode not only sensorimotor but also high-order cognitive information (language, emotion, …). This information is then bound by synchronization of gamma rhythms modulated by the thalamus.

The previous models, however, were merely qualitative or conceptual. As Barsalou (2008) pointed out, the wealth of well-documented experimental data now urges the development of computational models to formalize the qualitative theories and inspire experimental tests. Unlike classical artificial intelligence methods which manipulate abstract symbols, the grounded cognition assumption should adopt the formalism of dynamic systems and neural network architectures to implement its basic ideas in computational models (Barsalou, 2008).

Several connectionist models have been developed in recent years, using attractor networks. Most of them are aimed at analyzing how the statistical relationships between features and categories can be incorporated in a semantic memory model, to explore the consequences for language processing (for instance, semantic priming) and to simulate semantic deficits by damage to network connection weights. Some models also investigated modal vs. amodal representation, and the role of distinctive vs. shared features.

Among others, Farah and McClelland (1991) developed a model in which differences between living and non-living things were simulated using a different number of functional and perceptual features for each concept. The model was able to explain selective category impairment by removing some features. Hinton and Shallice (1991) and Plaut and Booth (2000) used a back-propagation network with a feed-forward from orthography to semantics, and a feedback loop from semantics to hidden units. The damaged network exhibited behavior emulating phenomena found in deep dyslexia. McGuire and Plaut (1997) trained a network to map a visual or tactile representation onto phonology and action units via a common set of intermediate units. Plaut (2002) further developed this model assuming that connections among hidden units are constrained to favor short connections. An important result is that hidden units develop a graded degree of modality (i.e., while some regions in the hidden units network are mainly unimodal (visual or tactile) other regions are multimodal). The model was then used to simulate optic aphasia.

Rogers et al. (2004) developed a model in which perceptual, linguistic, and motor representations communicate via a heteromodal region (probably located in the anterior temporal cortex), which encodes semantic aspects and recalls the "convergence zone" hypothesized by Damasio (Damasio et al., 1996). The model has been used to differentiate between semantic dementia (which causes a generalized semantic impairment) and other pathologies characterized by category-specific deficits (Lambon Ralph et al., 2007).

Vigliocco et al. (2004) developed a model assuming two levels of semantic representations: conceptual features and lexico-semantic representations. The first is assumed to be organized according to modality. An important aspect of their model is that the lexico-semantic space is trained in an unsupervised manner to develop a self-organizing map. This map allows a similarity to be built between lexical representations.

Other authors focused on the role of distinctive and shared features. Randall et al. (2004) trained a feed-forward three-layer back-propagation model to map words from semantic features, and studied the role of shared and distinctive features. Their results suggest that distinctive features are vulnerable due to weak correlation with other features. Cree et al. (1999) developed an attractor network with three sets of features (word-form, semantic features, and hidden units) trained with the back-propagation through time learning algorithm. Cree et al. (2006) used a variation of this model, in which semantic units were directly connected by reciprocal weights, to investigate the role of distinctive features in the computation of word meaning. Contrary to Randall et al. (2004), their results suggest that distinctive features have a privileged role in the computation of word meaning.

Alternative models used attractor networks trained by the Hebb rule to simulate phenomena such as semantic priming in normal subjects and in schizophrenia patients (Siekmeier and Hoffman, 2002) or to study the type of errors made by dyslexic patients (McRae et al., 1997) and by patients with dementia (Gonnerman et al., 1997).

Miikkulainen (1993, 1997) developed a model consisting of two self-organizing maps (one for lexical symbols and the other for word meaning) and of associative connections between them based on Hebbian learning. This model is perhaps the most similar to the model presented in this work.

The brief overview presented above highlights the increasing impact of computational models on the study of semantic and lexical aspects. This study describes a new model of the lexical-semantic memory, which is coherent with the grounded cognition hypothesis. Characteristics shared with previous models are: (i) a distinction between a conceptual representation (based on features) and a lexical store; (ii) the conceptual store is divided into distinct areas, which may be devoted to different modalities; (iii) the lexical aspects are implemented in a "convergence zone"; (iv) concepts are retrieved using attractor dynamics.

Original aspects of the model, not incorporated in previous studies are the following. (i) The use of oscillatory units in the gamma-band. Hence, attractors are not steady states but synchronized oscillations among neurons participating in the same object representation. This solution allows several objects to be retrieved simultaneously in memory and correctly segmented via temporal phase separation. This may allow the realization of more sophisticated semantic memories in which several concomitant objects concur to form a complex scene. Indeed, in many cognitive problems, several representations in memory may need to be maintained to have a complete understanding of the scene or realize a complex task (let us consider, for instance, sentence comprehension, or working memory tasks). The role of gamma-band synchronization in object recognition is supported by many data in the neurophysiological literature, not only in relation to perceptual problems, but also for high-level cognitive tasks (Pulvermüller et al., 1996; Tallon-Baudry et al., 1997, 1998; Bhattacharya et al., 2001; Salinas and Sejnowski, 2001; Osipova et al., 2006; Melloni et al., 2007). (ii) Features in the semantic space have a topological organization typical of modal representations in the cortex (for instance in motor, tactile and auditory areas, and in many visual areas). With the exception of the model by Miikkulainen (1993, 1997), other models do not use topological maps to describe the conceptual aspects of their objects. A self-organizing map was used by Vigliocco et al. (2004) but was devoted to lexical aspects. (iii) A different and more physiological rule (i.e., a time dependent Hebbian rule) is used to train synapses, employing both potentiation and depression, whereas most models used a back-propagation supervised algorithm to train synapses. In particular, our rules spontaneously lead to a clear distinction between the role of shared and distinctive features in the conceptual network. Although other models analyzed this distinction (Cree et al., 2006), our structure of synapses is new, as shown in the Section "Results" and analyzed in the Section "Discussion."

A simpler version of our current model, limited to the semantic aspects, was outlined in a recent work (Ursino et al., 2009). The previous version, however, had several limitations: (i) each object could be represented by a fixed number of features, whereas, according to Pexman et al. (2002, 2003) the number of features plays an important role in object recognition. In particular, these authors observed that words with a higher number of features respond more quickly than words with fewer features, a result which supports a distributed representation of meaning; (ii) there was no clear difference between distinctive features and shared features, to represent objects in the same class; (iii) there was no link between the semantic and lexical aspects.

Aim of this paper is to present an upgraded version of the model, and test it with a few examples on simulated objects, emphasizing the analogy with the grounded cognition assumption. The main improvements are: (i) the model can manage objects with a different number of features; (ii) it clearly differentiates between the role of distinctive and shared features in object recognition; (iii) it includes a lexical area for the representation of lemmas (or word-forms); (iv) the model learns the relationships between features of the same object (i.e., object semantics), and its word-form from exempla, using physiological learning rules.

Simulations are presented to show how the model can evoke correct words from a partial cue in the semantic area, and how a word can reconstruct the conceptual representation of the object in the semantic network, spreading over different feature areas. Model limits and lines for future investigation are then discussed.

MODEL DESCRIPTION

The model incorporates two neuronal networks, as illustrated in the schematic diagram of **Figure 1**. The first is devoted to representing the conceptual meaning of objects, described as a collection of features (semantic network). The second represents word-forms or lemmas (lexical network). The semantic network is explained first, including mechanisms for learning the object conceptual meaning from exempla. Then the lexical network is introduced, explaining mechanisms to link semantic and lexical aspects. Equations for the new (lexical and training) aspects are given in the Appendix, together with parameter numerical values.

THE SEMANTIC NETWORK

Qualitative description

The first network, named "semantic network," is devoted to a description of objects represented as a collection of sensory-motor features. These features are assumed to spread along different cortical areas (in both the sensory and motor cortex and perhaps also in other areas, for instance emotional) and are organized topologically according to a similarity principle. This means that two similar features activate proximal neural groups in the network.

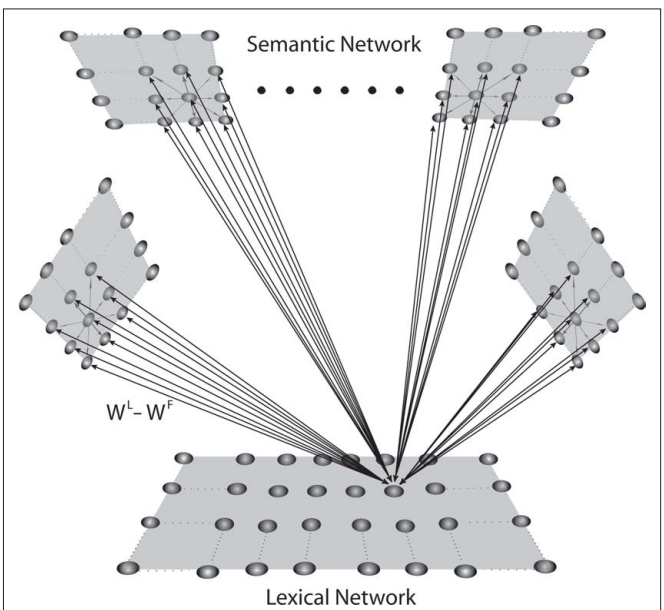


FIGURE 1 | Schematic diagram describing the general structure of the network. The model presents a "semantic network" and a "lexical network." The semantic network consists of nine distinct Feature Areas (upper shadow squares), each composed of 20 × 20 neural oscillators. Each oscillator is connected with other oscillators in the same area via lateral excitatory and inhibitory intra-area synapses, and with oscillators in different areas via excitatory inter-area synapses. The lexical area consists of 20 × 20 elements (lower shadow square), whose activity is described via a sigmoidal relationship. Moreover, elements of the feature and lexical networks are linked via recurrent synapses (W^F, W^L).

The network is composed of *F* distinct cortical areas (see **Figure 1**). Each area, in turn, consists of a lattice of neural oscillators. An oscillator may be silent, if it does not receive enough excitation, or may oscillate in the γ-frequency band (30–70 Hz) if excited by sufficient input. Oscillator dynamics is realized via the local feedback connection of an excitatory and an inhibitory population. This arrangement can be seen as a simple description of a cortical column or of a cortico-thalamic circuit. An oscillatory activity in this network allows different objects to be simultaneously held in memory (i.e., it favors the solution of the binding and segmentation problem) via γ-band synchronization (see Ursino et al., 2009). Oscillators representing the properties of the same object should oscillate in phase, whereas oscillators representing properties of different objects should oscillate with a different phase. Gamma-band synchronization has been proposed as an important mechanism in high-level cognitive tasks, including language recognition and semantic processing (Steriade, 2000; Slotnick et al., 2002; Kraut et al., 2004).

During the simulation, a feature is represented by a single input localized at a specific coordinate of the network, able to trigger the oscillatory activity of the corresponding unit. We assume that this input is the result of an upstream processing stage that extracts the main sensory-motor properties of the objects. In previous works (Ursino et al., 2009), we assumed that each object is described by a fixed number of features. Conversely, we now consider that the number of features can vary from one object to the next. The way these features are extracted and represented in the sensory and motor areas is beyond the aim of the present model. The use of authentic objects with realistic features may represent a further evolution of this model.

The present network has a maximum of nine features: this constraint was imposed merely to reduce the simulation computational cost.

A topological organization in each cortical area is realized assuming that each oscillator is connected with the others in the same area via lateral excitatory and inhibitory synapses (*intra-area synapses*). The synapses have a Mexican hat disposition, i.e., proximal neurons excite reciprocally and inhibit more distal ones. This disposition produces an "activation bubble" in response to a single localized feature input: not only is the neural oscillator representing that individual feature activated, but also the proximal ones linked via sufficient lateral excitation. This has important consequences for object recognition: neural oscillators in proximal positions share a common fate during the learning procedure and are subject to a common synapse reinforcement. Hence, they participate in the representations of the same objects. In this way, an object can be recognized even when some of its features are slightly altered (*similarity principle*).

Throughout the present paper, the lateral intra-area synapses and the topological organization will not be trained, i.e., they are assigned "*a priori.*" This choice is convenient to maintain a clear separation between different processes in our model (i.e., implementation of the similarity principle on the one hand and implementation of object semantics on the other). Of course, topological maps can also be learned through experience (Hertz et al., 1991), but this mechanism probably develops in early life and precedes object semantic learning.

Besides the intra-area synapses, we also assumed the existence of excitatory long-range synapses between different cortical areas in the semantic network (inter-area synapses). These are subject to a training phase (see below) and implement the conceptual (i.e., semantic) knowledge of the object.

Training the semantic network

The inter-area synapses connecting different features are trained in a first phase, in which single objects (described by all their features) are presented to the network one by one. We assume that synapses are reinforced based on the correlation between the activity in the post-synaptic unit, and the activity in the pre-synaptic unit mediated over a previous 20 ms time window (Markram et al., 1997; Abbott and Nelson, 2000). However, reinforcement alone would produce a symmetric pattern of synapses, whereas an asymmetric pattern of synapses may be useful in case of objects with common features. Let us consider an object sharing common features with other objects but with some distinctive features (examples will be considered in the Results). It can be expected that distinctive features are highly important to recognize an object and evoke the remaining features (including all common ones). Conversely, it can be expected that common features (shared with other objects) do not evoke distinctive features. To obtain this behavior, asymmetrical synapses are needed: synapses from common features to distinctive features must be weaker, whereas synapses from distinctive features to common features must be stronger. This asymmetrical pattern of synapses is obtained assuming that a synapse weakens when the pre-synaptic neuron is active and the post-synaptic neuron is inhibited (homosynaptic depression). In this way, synapses from common to distinctive features weaken at any presentation of a new object sharing the same common features. An example of the synapses obtained after presentation of two objects, each with seven features (four common features and three distinctive features) is given in **Figure 2**.

Behavior after training

After training, the semantic network exhibits the typical behavior of an auto-associative memory, i.e., it can reconstruct the overall conceptual information of an object starting from a partial content. However, several aspects differentiate this network from classic auto-associative nets (such as the Hopfield net, see Hertz et al., 1991). First, thanks to the topological implementation, an object can be reconstructed even after moderate changes in a few features (similarity principle). Second, thanks to the forgetting factor included via the homosynaptic depression, distinctive features of an object play a greater role than shared features. Lastly, oscillatory activity allows multiple objects to be simultaneously held in memory.

THE LEXICAL NETWORK

To represent lexical aspects, the model includes a second layer of neurons, denoted "lexical network." Each computational unit in this network codes for a lemma or a word-form and is associated with an individual object representation. In this case too, the input must be considered the result of an upstream processing stage, which recognizes the individual words from phonemes or from orthographic analysis. Description of this processing stream is well beyond the

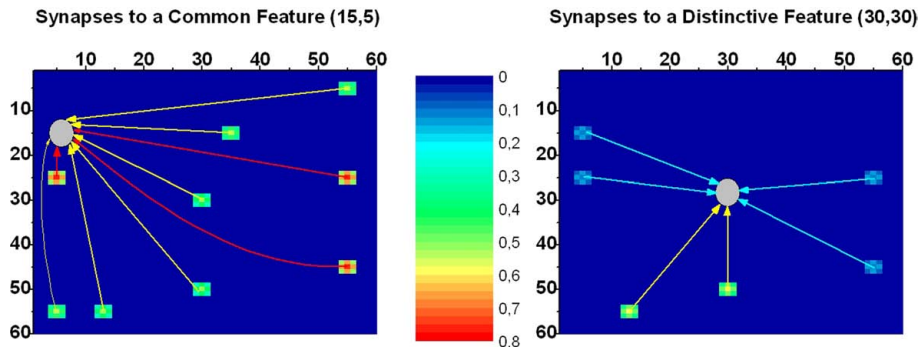


FIGURE 2 | An example of the synapses connecting features in the semantic network after training. The network was trained using obj3 and obj4 described in **Table 1** (other objects are not included to simplify the analysis). The color in each figure represents the strength of the synapses reaching a given post-synaptic neuron, coming from different pre-synaptic neurons. The position of the post-synaptic neuron is marked with a gray circle. Arrows have been included for clarity. Two exempla are reported in the present figure. The left panel describes the strength of the synapses reaching a neuron coding for a common feature (this is the neuron at position 15, 5, whose feature is common for obj3 and obj4). The right panel describes the synapses reaching a neuron coding for a distinctive feature (this is the neuron at position 30, 30, whose feature is distinctive for obj4). The figure content can be explained as follows: (i) a common feature (left panel) receives synapses from all features in obj3 and obj4. The synapses coming from the other three common features are stronger (red color) than those coming from the six distinctive features (green color), since common features are more often encountered during training; (ii) a distinctive feature (right panel) receive synapses only from the other six features of the same object (obj4). Synapses from the distinctive features have the same strength as in the left panel (green color), whereas synapses from common to distinctive features are weaker (cyan color) as a consequence of depression (see text for details). Hence, common features are strongly interconnected with other common features, but do not evoke distinctive features. Distinctive features are moderately interconnected, and can evoke both other common features and other distinctive features. Finally, it is worth noting that a neuron receives synapses not only from neurons coding for the "exact" feature of the same object, but also from proximal neurons, thereby constituting an "activation bubble." This implements a similarity principle.

aim of this model: some exempla can be found in recent works by others (see Hopfield and Brody, 2001) for word recognition from phonemes, and (Farah and McClelland, 1991; Hinton and Shallice, 1991) for orthographic processing models.

Of course, units in this network can also be stimulated through long-range synapses coming from the semantic network: in this regard, the lexical network constitutes an amodal convergence zone, as hypothesized in the anterior temporal lobe (Damasio, 1989; Gainotti, 2005). Long-range synapses between the lexical and the semantic networks are subjected to learning (see below) and may be either excitatory or inhibitory.

For the sake of simplicity, computational units in this network are not described as oscillators. Hence, if stimulated with a constant input, they reach a given steady-state activation value after a transient response (but, of course, they oscillate if stimulated with an oscillating input coming from the semantic network).

Training the lexical network

In order to associate words with their object representations, we performed a second training phase in which the model receives a single input to the lexical network (i.e., a single word is detected) together with the features of a previously learned object. Synapses linking the objects with words in both directions (i.e., from the lexical network to the feature network and vice versa) are then trained.

While synapses from words to features ($W^L_{ij,hk}$ in **Figure 1**) are simply excitatory and are trained on the basis of the pre- and post-synaptic correlation, when computing the synapses from features to words ($W^F_{ij,hk}$ in **Figure 1**) we tried to address a fundamental requirement: a word must not be evoked if spurious features (not originally belonging to the prototypical object) are active. This situation may occur when two or more concomitant objects are not correctly segmented, and some of their features pop up together. Hence, this requirement corresponds to a correct solution of the segmentation problem. This requirement also avoids a member of a category evoking the word representing the whole category (assuming that features in a category are shared by its members).

To address this requirement, we assumed that before training all units in the feature network send strong inhibitory synapses to units in the lexical network. Hence, activation of any feature potentially inhibits lexical units. These synapses are then progressively *withdrawn* during the training phase, and *excitatory* synapses are formed on the basis of the correlation between activity in the feature unit and in the lexical unit. Moreover, we assumed that after sufficient training the *sum* of all excitatory synapses reaching a word must be constant, irrespective of the number of features (i.e., we adopted a normalization of synaptic weights). The consequence of this choice is that after training a word receives inhibition from all features that do not belong to its semantic representation, but receives excitation from its own feature units. When all features are present, without spurious features, the neuron coding for the specific word is excited above a threshold and switches from the off to the on state. Switching of neurons has been mimicked using a sharp sigmoidal relationship.

Behavior after training

After training, a word can reconstruct its conceptual representation in the semantic network by evoking the same cortical activity present during object learning, in agreement with the grounded cognition assumption ("word recognition task"). Similarly, reconstruction of a complete object representation in the semantic network from a partial cue evokes the corresponding word in the

lexical area ("object recognition task"). Finally, the network can distinguish between a category and its individual members (and evokes the corresponding words) on the basis of the shared and distinctive features.

SIMULATIONS

Simulations shown in this work were performed using six distinct objects: two of them (named obj1 and obj2) have five features, the others (obj3, obj4, obj5, and obj6) have seven features. Moreover, obj3 and obj4, are correlated, i.e., they exhibit four shared features plus three distinctive features. We assume that the four shared features of obj3 and obj4 denote a category (say "ctg1") (just to fix our ideas, obj3 can be "dog," obj4 can be "cat," and the four shared features may represent the category "pet"). Another two objects (obj5 and obj6) also exhibit four shared features, which denote a new category (say "ctg2"). Moreover, obj3, obj4, obj5, and obj6 have two shared features that identify a wider category (say "ctg3") incorporating "ctg1" and "ctg2." To fix our ideas, obj5 and obj6 may be "cattle" and "goat," "ctg2" the category "ruminant" and "ctg3" the category "animal." During a first training period, the objects were given to the network several times in a random fashion. The learning rate and the number of iteration steps during the training phase were chosen so that at the end of training the objects could be reconstructed giving about one half of their features. Subsequently, during a second training period the objects (including the categories) were given to the semantic network together with the corresponding word in the lexical network, and the association objects-words was created. The position of the individual features in the semantic network, and the position of the corresponding word-forms in the lexical network are given in **Table 1**.

Other objects and categories, uncorrelated with the previous ones, with a number of features ranging between two and nine, were also tested. They behave substantially like those presented here.

RESULTS

SIMULATION OF OBJECT RECOGNITION AND WORD RECOGNITION TASKS

Figure 3 shows the example of a concomitant "object recognition" and "word recognition" task. In this simulation, a word (obj3) was given to the lexical network, and just four features of a second object (obj1) were given to the semantic network. As the figure clearly shows, after a short transient period the semantic network is able to recover the missing feature of obj1; the overall conceptual meaning of obj1 then evokes the corresponding word in the lexical area but almost completely inhibits activity of the word coding for obj3. After a short period of time, when the conceptual representation of obj1 is turned off, the word representing obj3 evokes its conceptual representation in the semantic network. The semantic representations of obj3 and obj1 then oscillate in time division. It is worth noting that the network recreates the same cortical representations in the semantic network that were originally present during the learning of obj1 and obj3, according to the grounded cognition hypothesis.

Figures 4 and 5 describe the results of two distinct object recognition tasks, in which the subject must recognize a single category (**Figure 4**) or a member from a category (**Figure 5**) starting from partial cues in the semantic network. More particularly, in **Figure 4**

Table 1 | Position of the features in the semantic network for the six simulated objects (second column), and position of the corresponding word-forms in the lexical network (third column).

Object	Feature positions	Word-form position
Obj1	[50 15]	[5 10]
	[50 30]	
	[30 30]	
	[50 45]	
	[30 45]	
Obj2	[10 15]	[5 15]
	[10 30]	
	[30 55]	
	[50 5]	
	[5 45]	
Obj3 (dog)	[15 5]	[15 5]
	[15 35]	
	[55 5]	
	[25 5]	
	[45 55]	
	[5 55]	
	[25 55]	
Obj4 (cat)	[15 5]	[15 10]
	[30 30]	
	[55 13]	
	[25 5]	
	[45 55]	
	[50 30]	
	[25 55]	
Obj5 (cattle)	[15 5]	[3 3]
	[15 45]	
	[25 25]	
	[25 15]	
	[55 55]	
	[45 25]	
	[25 55]	
Obj6 (goat)	[15 5]	[7 7]
	[15 55]	
	[35 35]	
	[25 15]	
	[55 55]	
	[55 35]	
	[25 55]	
Ctg1 (pet)	[15 5]	[15 15]
	[25 5]	
	[45 55]	
	[25 55]	
Ctg2 (ruminant)	[15 5]	[13 13]
	[25 15]	
	[55 55]	
	[25 55]	
Ctg3 (animal)	[15 5]	[11 11]
	[25 55]	

The first two objects have five features each, while the others have seven features each. The last objects denote categories constituted by the shared properties of obj3, obj4, obj5, and obj6.

Table 2 | Model parameters used during the simulations.

LEXICAL AREA	
ϑ^L	50
p^L	1
τ^L	1 ms
HEBBIAN RULE FOR SYNAPSES WITHIN THE SEMANTIC NETWORK (PHASE I)	
W_{max}	1
β_0	0.25
λ	0.07
HEBBIAN RULE FOR SYNAPSES BETWEEN FEATURES NETWORK AND LEXICAL NETWORK (PHASE II)	
β_0^L	0.125/5
W_{max}^L	5
β_0^{Fex}	0.125/2
W_{summax}^{Fex}	2
W_{max}^{Fin}	0.06
β^{Fin}	0.125/2

three features shared by obj3 (dog) and obj4 (cat) are given to the semantic network. The model reconstructs the fourth shared feature, and the word denoting the category "pet" is finally evoked in the lexical area, without evoking the words denoting individual members. In **Figure 5**, the semantic network receives two distinctive features of obj3 ("dog") and one shared feature of "pet." The semantic network now recovers all seven features of "dog," and the word denoting the individual member is correctly evoked in the lexical area. These simulations clearly disclose the different role of shared and distinctive features in object recognition.

Figure 6 shows the case in which all 10 features of obj3 and obj4 (i.e., cat and dog) are given to the semantic network. The two objects are correctly segmented (despite the presence of the four common features) and the two words are evoked in the lexical area. It is worth noting that some isolated features appear at some instants during the simulation, but do not evoke any response in the lexical area.

Finally, **Figure 7** shows the result obtained by giving the lexical network two words of the same category (the word "dog" and the word "cat"). If the two words are given simultaneously, the behavior of the semantic network is remarkable. The four common features oscillate in synchronism, and the word "pet" appears in the lexical network. Hence, the model is able to generalize from two words to a category. The three distinctive features of cat, and the three distinctive features of dog oscillate independently, without a synchronization with the remaining four features.

INCREASING THE NUMBER OF OBJECTS AND CATEGORIES

In the previous exempla we used just a small number of objects and categories (three objects and just one category). An important problem is whether the network can manage a larger number of objects representing many categories, and can manage a taxonomy of categories. As further commented in the Section "Discussion," we can store additional objects in the network without altering its behavior, provided the new objects are uncorrelated to the existing ones. We tested this aspect by storing new objects (not shown in **Table 1** for brevity) with the following characteristics: (i) new objects with distinct features (from two to nine) but no correlation with the others. They can be retrieved correctly; (ii) pairs of objects with shared features, but no correlation with previous objects; they generate new categories and behave like obj3 and obj4.

A more complex condition may occur if we consider the case of three or more objects within the same category, or objects which involve a taxonomy of categories. To test this condition, we incorporated two new objects in the network (obj5 and obj6, see **Table 1**). Objects now generated a simple taxonomy: a category (say "animal") now includes four objects (dog, cat, cattle, and goat) with two shared features; another two categories (say "pet" and "ruminant"), included within "animal," have two objects each with four shared features. All objects were first trained separately in the semantic network, then objects and categories were associated with word-forms in the lexical network (see **Table 1**).

Simulations show that the model can correctly discriminate between these categories and individual objects, and correctly evoke the corresponding words. In particular, a word in the lexical network evokes the correct conceptual representation in the semantic network, both if the word represents an object and when it represents a category. For instance, the word "goat" evokes all seven features; the word "ruminant" evokes the four features common to cattle and goat only; the word "animal" evokes just the two features shared by dog, cat, cattle, and goat. Similarly, providing a number of features in the semantic net causes the correct object reconstruction (without confounding objects and categories) and evokes the corresponding word. An example is given in **Figure 8**. The upper panel shows a simulation when five features of goat were given to the network (all four shared features in ruminant and one feature distinctive of goat); the remaining two features are evoked and the correct word denoting "goat" appears in the lexical area. The mid panel shows the case when the semantic net receives one feature in animal, together with two features in ruminant (these are distinctive features for this category); the fourth feature is evoked, together with the word-form denoting "ruminant." Finally, the bottom panel shows the case when only two features of animal, shared by four objects, are given. In this case, in the absence of any distinctive feature, the larger category ("animal") is evoked in the lexical net.

SIMULATION OF LEXICAL DEFICITS

A common assumption to explain selective impairment in category representation is that different categories exploit different subsets of features (for instance, sensory features are essential to recognize animate objects, whereas motor features are essential to recognize tools). Hence, a lesion damaging a specific zone of the cortex would cause a selective impairment only for those categories which intensively exploit critical features in that zone.

In order to simulate selective impairment with our model, we considered the first two objects in **Table 1** (obj1 and obj2) which have five features each, without common features. We chose to compare two objects with the same number of features, so that any difference can be ascribed to the position of features rather than to the complexity of the semantic representation. To simulate a semantic deficit, we assumed that a given percentage of synapses *emerging* from neurons in a given region (see **Figure 9**, upper panels) are damaged in a random fashion (i.e., they have been randomly set at zero). This

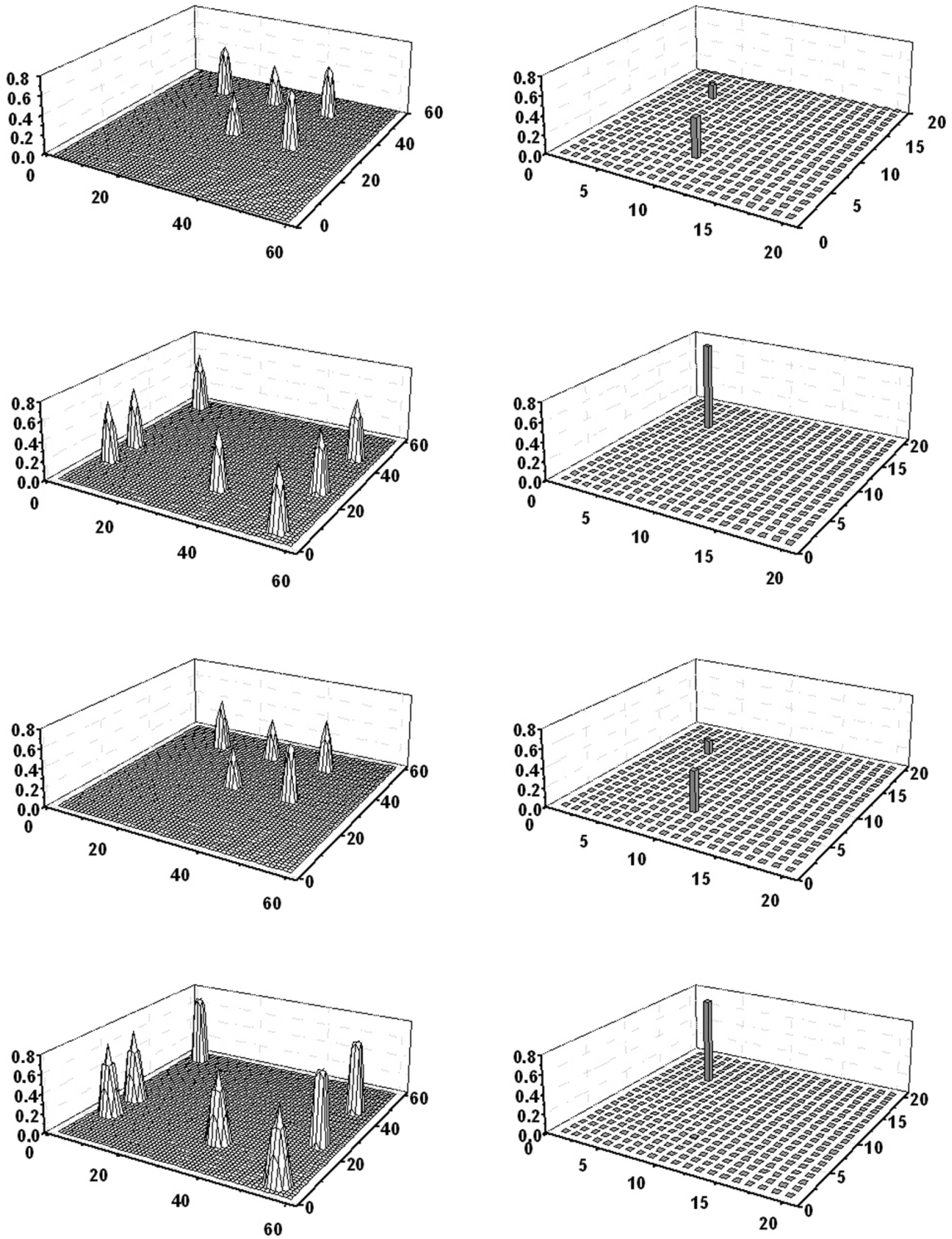


FIGURE 3 | Simulation of a "word recognition task" and an "object recognition task" performed simultaneously. Four features of obj1 were given as input to the semantic network, while the word denoting obj3 was given to the lexical network. The four lines represent activity in the semantic network (left panels) and in the lexical network (right panels) at steps 136, 156, 199, and 217 of the simulation (duration of each step: 0.2 ms). At step 136 (first line) obj1 is completely reconstructed in the semantic network, the word coding for obj1 is activated in the lexical network, while the word coding for obj3 is inhibited. At step 156 (second line), when the conceptual representation of obj1 has been switched off, the word coding for obj3 evokes its conceptual representation in the semantic network. Then, the two representations alternate, oscillating in the gamma range.

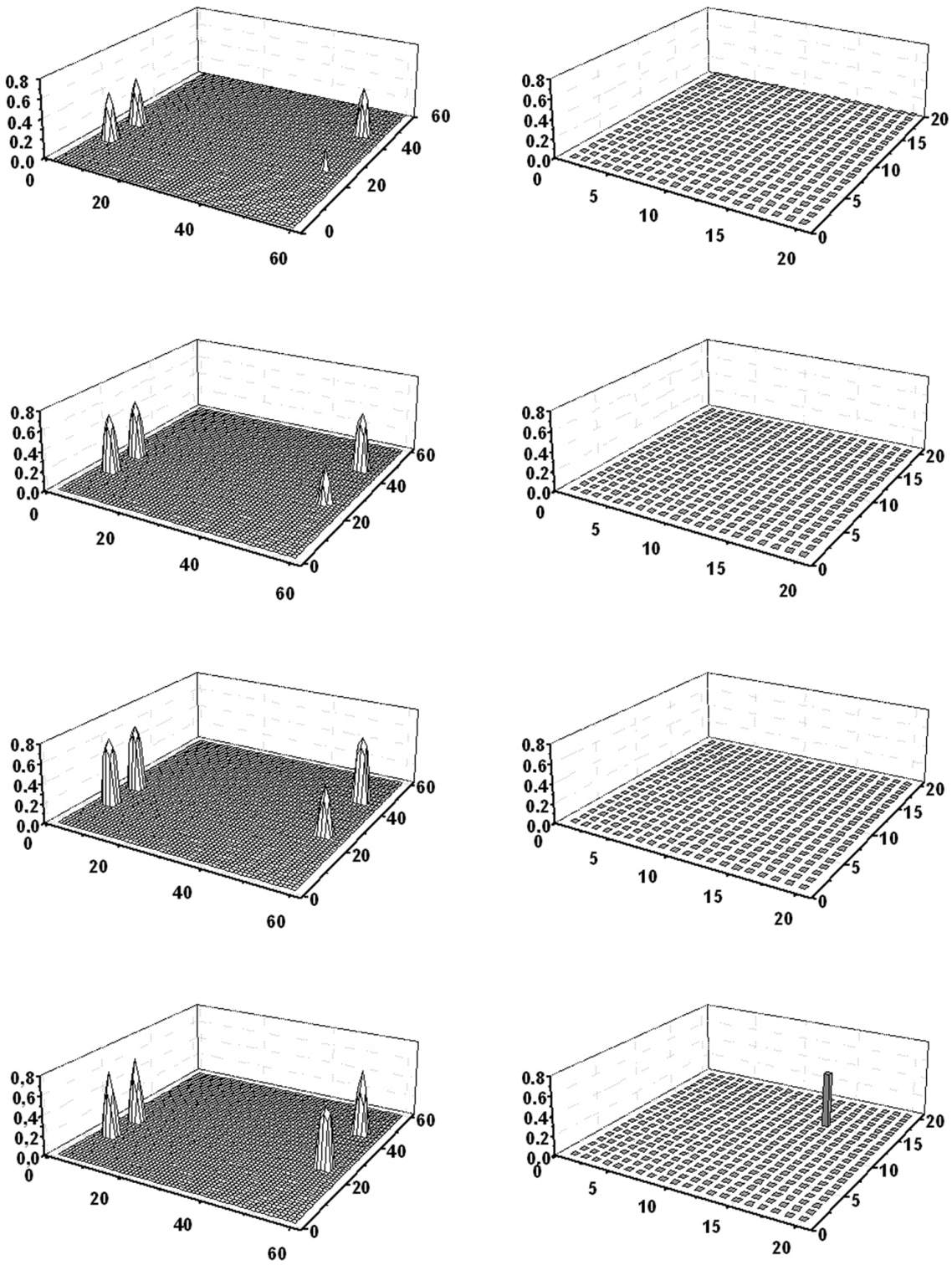


FIGURE 4 | Simulation of an object recognition task in which the subject must recognize a category. Three properties common to the dog and cat are given to the semantic network. The four lines represent activity in the semantic network (left panels) and in the lexical network (right panels) at steps 67, 68, 69, and 71 of the simulation (duration of each step: 0.2 ms). As evident in the left panels, the three features progressively evoke the fourth common feature in the semantic network; when the overall conceptual representation of the category is reconstructed, the word coding for that category ("pet") is activated in the lexical area (fourth line). It is worth noting that the distinctive properties of dog and cat are not evoked, and so individual members of the category do not appear in the lexical network.

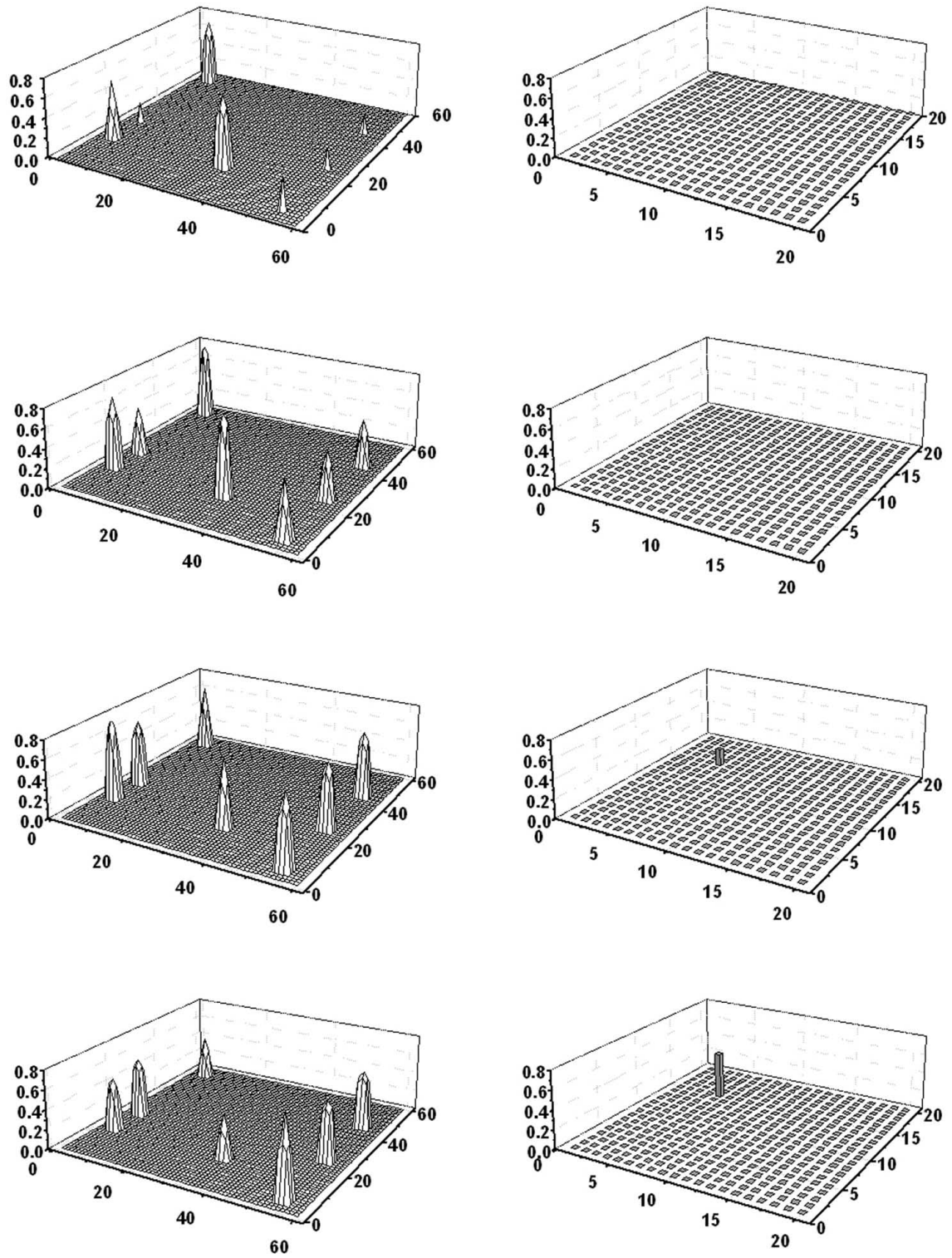


FIGURE 5 | Simulation of an object recognition task in which the subject must recognize a member of a category. Two distinctive properties of dog and one common property shared by dog and cat are given to the semantic network. The four lines represent activity in the semantic network (left panels) and in the lexical network (right panels) at steps 131, 133, 135, and 137 of the simulation (duration of each step: 0.2 ms). As shown in the left panels, the three features progressively evoke the four remaining features of dog in the semantic network; when the overall conceptual representation is reconstructed, the word coding for "dog" is activated in the lexical area (fourth line). It is worth noting that the distinctive properties of cat are not evoked. Neither the word coding for "cat," nor the word naming for the category ("pet") appear in the lexical network.

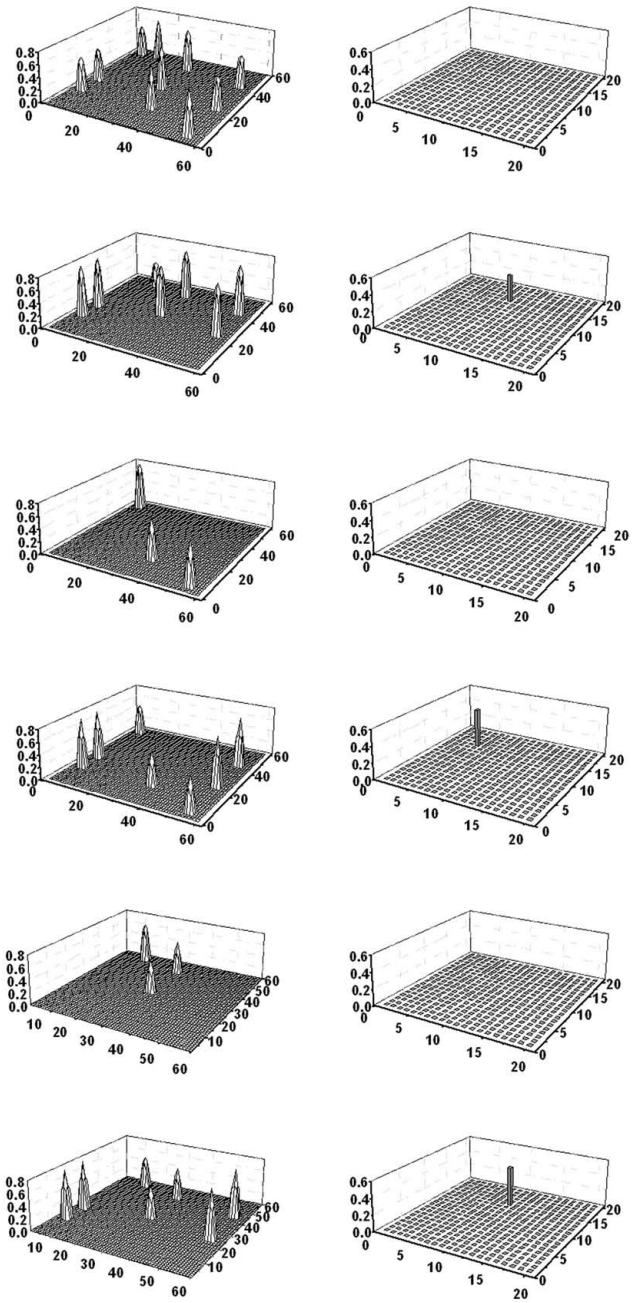


FIGURE 6 | Simulation of an object recognition task in which the subject must recognize two members of the same category. All 10 properties of dog and cat (four shared properties and three distinctive properties for each object) are given to the semantic network. The six lines represent activity in the semantic network (left panels) and in the lexical network (right panels) at steps 11, 133, 176, 239, 285, and 347 of the simulation (duration of each step: 0.2 ms). As the figure shows, the conceptual representations of the two objects are correctly reconstructed in the semantic network, and the corresponding words ("dog" and "cat") evoked in the lexical network, despite the presence of four shared features (see the second, fourth, and sixth lines). The word designating the category ("pet") does not appear in the lexical network. It is worth noting that some isolated features appear at some instants during the simulation (third and fifth lines), but do not evoke any response in the lexical area.

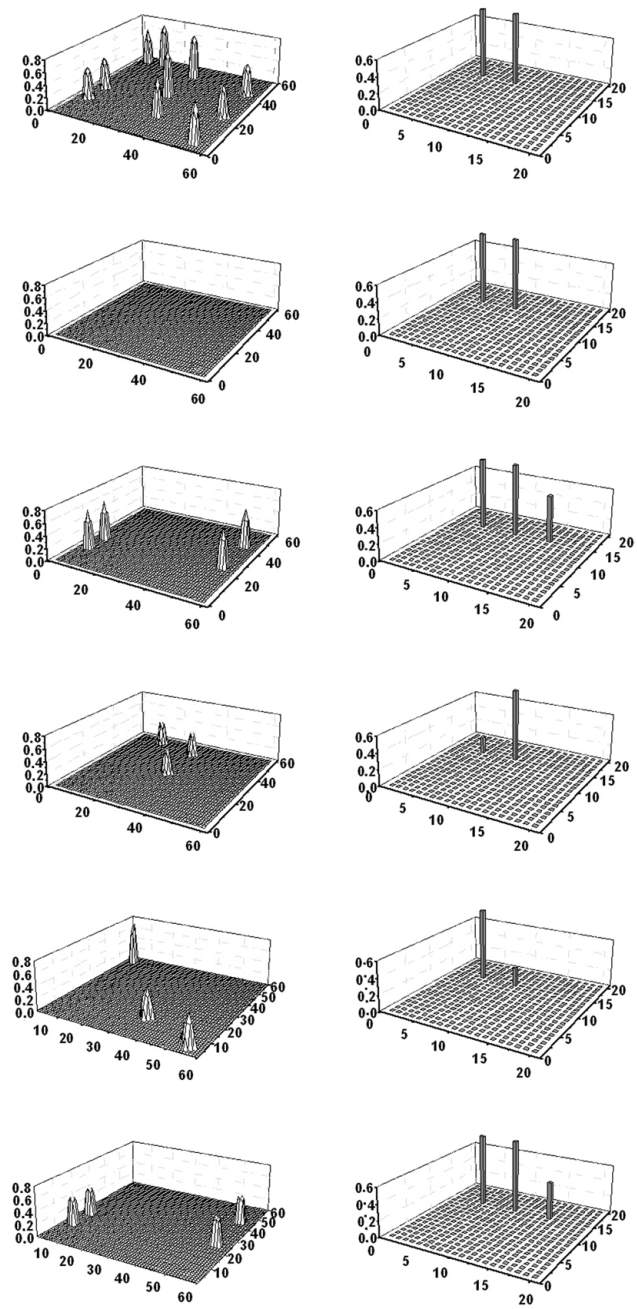


FIGURE 7 | Simulation of a word recognition task in which the subject must recognize two words from the same category. The two words naming "dog" and "cat" are given to the semantic network. The six lines represent activity in the semantic network (left panels) and in the lexical network (right panels) at steps 12, 32, 60, 81, 100, and 120 of the simulation (duration of each step: 0.2 ms). As the figure shows, the two words initially evoke the four shared properties in the semantic network and consequently the word denoting the category ("pet") is activated in the lexical area (third line). Hence, the network can generalize from the two members of the category to the category name. The three distinctive features of the two objects oscillate in time division (fourth and fifth lines) causing the momentarily inhibition of the alternative word. Hence, the three words (two members and their category) oscillate in time division in the gamma range.

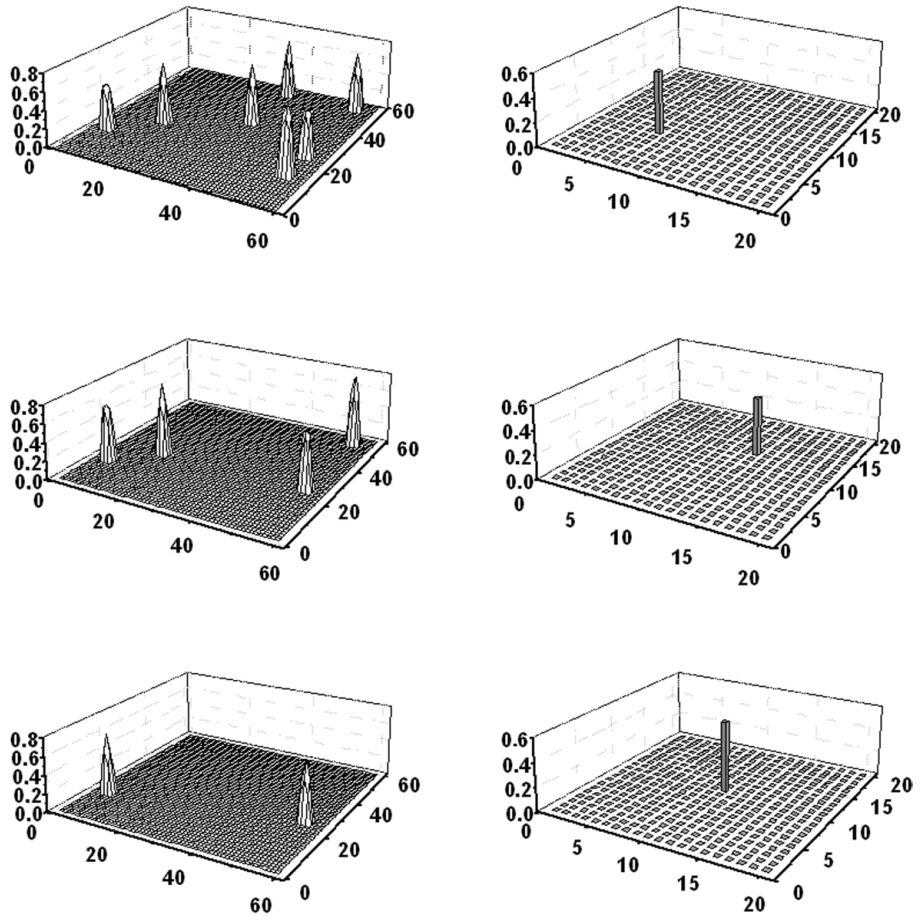


FIGURE 8 | Simulation of three object recognition tasks (performed separately) involving a taxonomy of categories. The upper panel shows a single snapshot obtained when the four properties of ruminant are given to the semantic network, together with one distinctive feature of goat. The network restores the two remaining features of goat, and the correct word ("goat") is evoked in the lexical area. The mid panel shows a single snapshot obtained when the two shared features of animal are given to the semantic network, together with a feature of ruminant. The fourth feature of ruminant is restored and the word "ruminant" evoked in the lexical area. Finally, the bottom panel shows a single snapshot obtained by giving the two features of animal to the semantic network. This information does not spread toward other features and the word "animal" is evoked in the lexical network.

may reflect a loss of neurons caused by a local lesion. As shown in the upper panels of **Figure 9**, the first object has four features in the damaged region. Conversely, the second object has just one feature in this region. For each level of synapse damage (from 0 to 40%) we repeated 10 simulations for each object presentation, and computed the response of the corresponding word in the lexical area. A response is assumed to be correct if presentation of the object is able to evoke a sufficient activity (above a given threshold) in the lexical area. The bottom panels of **Figure 9** show the percentage of correct responses for the two objects, as a function of the percentage of synapse impairment, and using two different thresholds for detection (one very small, 0.1 and the other quite high, 0.4). It is clear that if the percentage of synapse damage increases above 20–30%, the network frequently fails to recognize obj1 until, for a percentage of synapse damage as high as 40%, it almost completely loses the possibility to evoke a correct word. Conversely, recognition of obj2 is almost unaffected by the synapse damage, despite the presence of one feature in the damaged area. The four remaining features of obj2 are able to restore the fifth one, and recover the correct word irrespective of the local damage.

DISCUSSION

Modern theories on grounded cognition assign a pivotal role to "simulation" in the formation of the conceptual meaning of objects. Although the term "simulation" can have different meanings (see the recent paper by Borghi and Cimatti, 2010 for a clear analysis), a typical shared viewpoint considers simulation a form of re-activation of past experience. This may consist in the recruitment of the same neural networks involved in perception and action (Jeannerod, 2007) and in excitation of the same neural groups which coded for sensory and motor experience (Barsalou, 1999).

Our model implements grounded cognition by exploiting a few fundamental assumptions. Most of these find significant support in the recent neurophysiological and cognitive literature.

(i) The semantic and lexical aspects of declarative knowledge are stored in two distinct networks. Moreover, the semantic network spreads over different areas and exploits a distributed representation to describe object meaning. These ideas are frequently accepted in the neurocognitive literature, and are

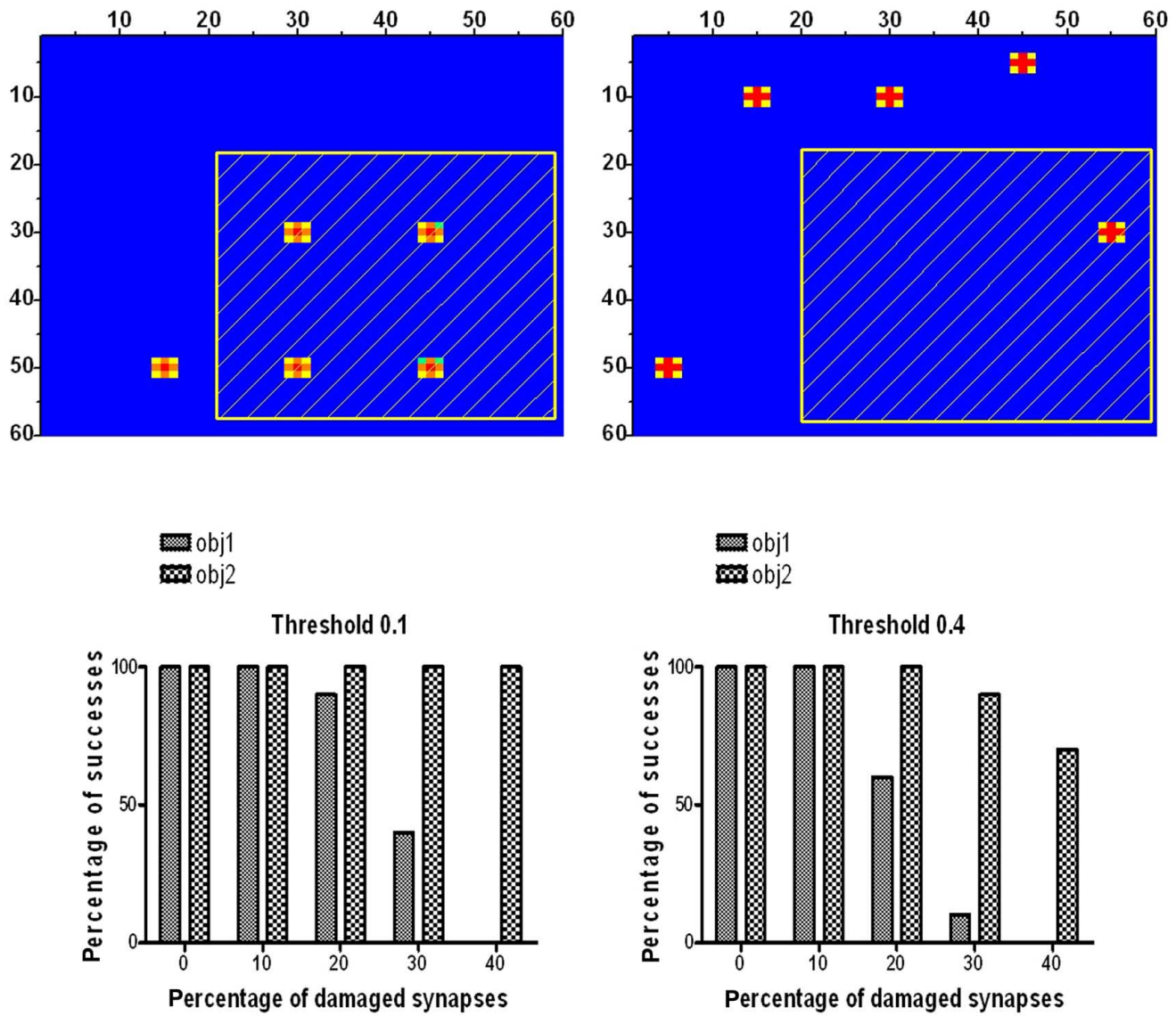


FIGURE 9 | Simulation of lexical deficit. The upper panels show the representations of two distinct objects in the semantic network (obj1 and obj2 in **Table 1**), with five features each, used to simulate selective impairment in category representation. The dashed area denotes the lesioned region, where a given percentage of synapses has been removed randomly. Obj1 has four features in the lesioned region, whereas obj2 has just one feature in that region. The bottom panels represent the percentage of success in recognition of obj1 and obj2, as a function of the percentage of damaged synapses. Two different thresholds for word recognition in the lexical area were used. It is worth noting that obj1 is frequently missed for a percentage of synapse damage greater than 20%, whereas the recognition of obj2 is quite robust despite synapse damage.

supported by analysis of bilingual subjects (Potter et al., 1984; Kroll and Stewart, 1994; McRae et al., 1997) who exhibit a common semantic representation but distinct lexical items.

(ii) The semantic network works like a classic auto-associative circuit, i.e., it can restore the overall past experience starting from a partial cue. Nevertheless, our semantic net differs from classic auto-associative networks in many important instances: it implements a similarity principle (an object can be restored even if it is just similar to a previous one) and can manage multiple objects simultaneously by means of rhythm synchronization in the gamma-band. The first aspect is realized through a topological organization of features, as in Barsalou and Simmons (Barsalou et al., 2003; Simmons and Barsalou, 2003). Of course, a modified version of our model can be imagined in which just some features (like colors, sounds, motion) have a topological organization and exploit a similarity principle, while other features are not topologically organized. This may be the subject of future more realistic model versions. The role of gamma-band synchronization in high-level cognition is well-documented, and is the subject of active research (the interested reader can find more references in Engel and Singer, 2001; Kraut et al., 2004; Fries et al., 2007). A further important

aspect differentiating our model from classic auto-associative nets is the presence of asymmetric synapses. While shared features do not evoke distinctive features, distinctive features are able to evoke shared features thereby leading to a complete object reconstruction. This aspect implements a different role for distinctive and shared features in the recognition of categories or members within a category, on the basis of synapse weights after training. A similar analysis of weights is reported in Cree et al. (2006) but with a noteworthy difference from our approach. Cree et al.'s model differentiates between shared and distinctive features by using the weights between words and features. The authors reported that weights from words to distinctive features are significantly higher than the weights to shared features. Similarly, weights leaving the distinctive features to words are higher than those leaving the shared features. As a consequence, distinctive features are more active than shared features and more strongly involved in word activation. Our results differ since differences between distinctive and shared features are implemented in the synapses within the semantic net (see **Figure 2**): distinctive features send stronger synapses to shared features, and receive smaller synapses from them. Our model first reconstructs an object completely in the semantic net and then associates it with the corresponding word-form. Conversely, the interaction between words and features is essential for object reconstruction in Cree et al.'s (2006) model. Both strategies may be implemented in real semantic memories.

(iii) The lexical network represents a sort of amodal convergence zone (Damasio, 1989). Neurons in this zone become active if and only if the overall semantic representation of the corresponding object is simultaneously active in the semantic net. The presence of a convergence zone supports the idea of "gamma-based computation" (Fries et al., 2007): if all features of the object are oscillating in phase, the convergence zone receives enough input excitation to trigger the word-coding neuron. Conversely, if features oscillate out of phase, neurons in the convergence zone do not receive enough excitation to become active. The lexical zone, however, does not only receive converging inputs (thanks to entering synapses), but also sends information back to cortical areas (thanks to outgoing synapses) thus enabling the re-creation of past experience. When a word is given to the lexical net, its emerging synapses recover the same activity in the semantic network that was present at the moment of object learning. If this activity is assumed to be motor, perceptual or emotional in type, the grounded experience of that object is recreated.

(iv) Features and word-forms in our model are assumed to be the result of upstream processing networks, which extract these features from previous sensory or motor information (for the semantic net) or from phonological or orthographic information (for the lexical net). A description of these processing stages is beyond the aim of the present study, but may be integrated in the model (possibly starting from already existing models) in future versions. One consequence is that neural activity, recovered to represent the conceptual meaning of objects, is not present in the primary motor and/or perceptual cortices, but is a kind of neural information processed in higher cortical areas (for instance, in the temporal visual pathway for visual information). This aspect agrees with the notion of "perceptual symbols" put forward by Barsalou (1999) and may explain the difference between merely perceiving an object and recreating its conceptual meaning. Object perception also involves activation in the primary cortical areas, while recreating the conceptual meaning of objects mainly activates higher areas involved in feature representation. Neuroimaging data in favor of this idea can be found in Martin (2007).

(v) Past experience is stored in model synapses using physiological learning rules. These exploit the correlation (or anti-correlation) of pre-synaptic and post-synaptic activities over a time window (10–20 ms) compatible with gamma-band activity (Markram et al., 1997; Abbott and Nelson, 2000). The model, however, assumes different versions for the learning rule to build synapses within the semantic network and to build lexical-semantic connections. These are essential to reach a correct model behavior (i.e., to have a different role for shared and distinctive features in the semantic net, and to represent objects in the lexical net with a different number of features). These rules may be the subject of *ad hoc* testing on the basis of available experimental data in the literature or new experiments on synaptic plasticity.

After implementing the previous basic ideas, the model can cope with word recognition and object recognition tasks quite well, also involving multiple objects, categories, and words. Furthermore, the model can distinguish the members of a category, and evoke a category from multiple members, exploiting the differences between shared and distinctive features. A particular behavior of the model is the response to two simultaneous words, representing two members of the same category. The model is able to generalize from members to the category by isolating the shared features from the distinctive features of the two objects (see **Figure 7**) and activating the three words (the two denoting members and the third denoting the category) in time division. We do not know if this model response is correct from a cognitive perspective, but it certainly represents an interesting emergent behavior. Finally, the present model implementation represents a simple and straightforward way to look at selective impairment in category representation. As shown in **Figure 9**, selective random damage to synapses in a zone of the semantic network naturally leads to a deficit in recognition of objects whose conceptual meaning exploits many features in that zone.

The present work considered just six different objects and three categories adopting a single hierarchical level between categories (i.e., we have one larger category which contains two smaller ones). Of course, the number of objects and categories can be increased without significant deterioration in model performance provided the new objects and the new categories are uncorrelated to the older ones (i.e., if patterns are orthogonal). This is a well-known property of auto-associative networks (Hertz et al., 1991). Indeed, we trained the network with some additional uncorrelated objects (with a number of features ranging between 2 and 9), and with further objects having some shared features (which implement new categories uncorrelated to previous categories) and the network functions correctly in storing and retrieving these objects.

A more complex problem may occur if the number of correlated objects increases, i.e., trying to store many objects with shared features, since this may cause a deterioration in the performance of auto-associative networks. The present work had a maximum of four correlated objects (i.e., a maximum of four objects within the same category) with a moderate correlation (2/7). Two pairs of objects have a greater correlation (4/7). Studying the capacity of the network to manage a larger number of correlated objects may be the focus of subsequent works. However, storing a large number of correlated objects in the network (to simulate a realistic data set) may require an increase in network size since the capacity of an auto-associative network depends on the ratio between the number of objects and the number of neurons (Hertz et al., 1991). The present study kept the number of neurons and feature areas quite low to contain the size of the synapse matrix and avoid an excessive computational charge.

A possibility offered by the model in future works is to study the potential occurrence of under-generalization or over-generalization during training. The present work used a similar learning rate for the different objects, and all objects were trained using a complete set of their features to obtain a correct behavior during word recognition and object recognition tasks. Over-generalization might occur if one object is stored much more strongly than other objects in the same category (for instance due to longer training or a higher learning rate). Conversely, under-generalization might occur when a specific feature (not really belonging to a category) is erroneously associated with the word representing the category. All these occurrences may be investigated with the model, studying its behavior during the training period as a function of the parameters used (learning rate, duration of the inputs) and the statistics of the input features.

Finally, it is important to stress aspects of the present model which deserve further investigation and may be the subject of future research.

In the introduction, we stressed that our model agrees with the "grounded" or "embodied" cognition viewpoint. Indeed, it may be objected that it is not easy to distinguish between features in our approach and symbols normally used in amodal computation, and so there is no real embodiment in the model. Conversely, we think that the present model contains some significant embodiment aspects, especially in the way features are represented and organized. First, as stressed above, in our model the features exhibit a topological organization, i.e., they are organized in maps resembling those found in many cortical regions within the motor and perceptual areas. Due to computational limits, each map is represented by means of a 20×20 lattice (i.e., we just have 400 variations of the same feature) but, of course, a much finer map could be constructed in which features exhibit minimal nuances (such as in real cortical maps). We claim that this aspect of the model clearly differentiates it from a symbolic representation. The topological representation of features allows implementation of a similarity principle. Indeed, each feature excites an "activation bubble," i.e., a group of neurons which respond to a similar attribute. This aspect again makes the model suitable to simulate perceptual or motor modalities, instead of abstract symbols. We are not aware of previous semantic models (with the exception of Miikkulainen, 1993, 1997) which implement this topological organization in the feature areas, although previous models have implemented a topology in the lexical area to organize words (Vigliocco et al., 2004). Another important aspect embodied in the model is that word recognition occurs by recreating over different cortical areas the same representation present when the object was originally learnt. Of course, we are aware that, at this level of modelization, our features cannot be completely distinguished from amodal symbols and in this regard our features resemble the "perceptual symbols" proposed by Barsalou. Future model application with more realistic data sets will replace the present schematic objects with real ones, using modal features in the different areas. This may allow more model predictions to be formulated and challenged against real data.

A further limitation of the present model is that we used just a localist representation of word-forms, whereas previous models included connections in the lexical area (see Miikkulainen, 1993, 1997; Vigliocco et al., 2004). In our model, possible connections among words occur just indirectly, i.e., are mediated by a correlation in the conceptual representation within the semantic net. It is likely that direct connections among words in the lexical network may be created by experience, especially if words occur frequently together or in close temporal proximity, even in the absence of a clear semantic correlation. Of course, a relationship among words may be of the utmost value to implement syntactic aspects in the model. The use of a localist representation of word-forms has been introduced to lay emphasis only on the semantic aspects. A more sophisticated description of lexical aspects (including a distributed representation of the activity in the net and connections among words) will be the object of future model improvement.

In conclusion, the present model provides a theoretical framework for the formalization of recent theories on the semantic-lexical memory system based on a grounded cognition approach. Original aspects consist in the possibility to manage multiple objects and words, and to distinguish between categories and individual members by learning distinctive and shared features on the basis of past experience. Although the present version only deals with simulated objects, it points out important aspects which may drive future research. These are especially concerned with the organization of the network and with the learning rules included. Subsequent versions of the model should consider the possibility to represent real objects and to simulate results of cognitive tests. This may permit the validation or rejection of hypotheses, a comparison with existing data, and the design of new tests.

REFERENCES

Abbott, L. F., and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.* 3, 1178–1183.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91.

Bhattacharya, J., Petsche, H., and Pereda, E. (2001). Long-range synchrony in the gamma band: role in music perception. *J. Neurosci.* 21, 6329–6337.

Borghi, A. M., and Cimatti, F. (2010). Embodied cognition and beyond: acting and sensing the body. *Neuropsychologia* 48, 763–773.

Caramazza, A., and Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J. Cogn. Neurosci.* 10, 1–34.

Cree, G. S., McNorgan, C., and McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: implications for theories of semantic memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 643–658.

Cree, G. S., McRae, K., and McNorgan, C. (1999). An attractor model of lexical conceptual processing: simulating semantic priming. *Cogn. Sci.* 23, 371–414.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems level proposal for the neural substrates of recall and recognition. *Cognition* 33, 25–62.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., and Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature* 380, 499–505.

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25.

Farah, M. J., and McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *J. Exp. Psychol. Gen.* 120, 339–357.

Fries, P., Nikolic, D., and Singer, W. (2007). The gamma cycle. *Trends Neurosci.* 30, 309–316.

Gainotti, G. (2000). What the locus of brain lesions tells us about the nature of the cognitive defect underlying category-specific disorders: a review. *Cortex* 36, 539–559.

Gainotti, G. (2005). The influence of gender and lesion location on naming disorders for animals, plants and artefacts. *Neuropsychologia* 43, 1633–1644.

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain Lang.* 84, 1–15.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., and Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb. Cortex* 14, 247–255.

Goldberg, R. F., Perfetti, C. A., and Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* 26, 4917–4921.

Gonnerman, L. M., Andersen, E. S., Devlin, J. T., Kempler, D., and Seidenberg, M. S. (1997). Double dissociation of semantic categories in Alzheimer's disease. *Brain Lang.* 57, 254–279.

Gonzalez, J., Barros-Loscertales, A., Pulvermuller, F., Meseguer, V., Sanjuan, A., Belloch, V., and Avila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage* 32, 906–912.

Hart, J., Anand, R., Zoccoli, S., Maguire, M., Gamino, J., Tillman, G., King, R., and Kraut, M. A. (2007). Neural substrates of semantic memory. *J. Int. Neuropsychol. Soc.* 13, 865–880.

Hart, J., Moo, L. R., Segal, J. B., Adkins, E., and Kraut, M. A. (2002). "Neural substrates of semantics," in *Handbook of Language Disorders,* ed. A. Hillis (Philadelphia, PA: Psychology Press), 207–228.

Hauk, O., Johnsrude, I., and Pulverm³ller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation.* Redwood City, CA: Addison-Wesley Publishing Company.

Hinton, G. E., and Shallice, T. (1991). Lesioning an attractor network: investigations of acquired dyslexia. *Psychol. Rev.* 98, 74–95.

Hopfield, J. J., and Brody, C. D. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1282–1287.

Humphreys, G. W., and Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behav. Brain Sci.* 24, 453–476.

Jeannerod, M. (2007). *Motor Cognition: What Actions Tell the Self.* Oxford: Oxford University Press.

Kraut, M. A., Pitcock, J. A., and Hart, J. (2004). Neural mechanisms of semantic memory. *Curr. Neurol. Neurosci. Rep.* 4, 461–465.

Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174.

Lambon Ralph, M. A., Lowe, C., and Rogers, T. T. (2007). Neural basis of category-specific demantic deficits for living things: evidence from semantic dementia, HSVE and a neural network model. *Brain* 130, 1127–1137.

Markram, H., Lubke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201.

McGuire, S., and Plaut, D. C. (1997). *Systematicity and Specialization in Semantics: A Computational Account of Optic Aphasia.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 502–507.

McRae, K., de Sa, V. R., and Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *J. Exp. Psychol. Gen.* 126, 99–130.

Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., and Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *J. Neurosci.* 27, 2858–2865.

Miikkulainen, R. (1993). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory.* Cambridge, MA: MIT Press.

Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the Lexicon. *Brain Lang.* 59, 334–366.

Osipova, D., Takashima, A., Oostenvald, R., Fernández, G., Maris, E., and Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *J. Neurosci.* 26, 7523–7531.

Pexman, P. M., Holyk, G. G., and Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Mem. Cogn.* 31, 842–855.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: a computational account of optic aphasia. *Cogn. Neuropsychol.* 19, 603–639.

Plaut, D. C., and Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychol. Rev.* 107, 786–823.

Potter, M. C., So, K. F., Eckardt, B. V., and Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *J. Verbal Learn. Verbal Behav.* 23, 23–38.

Pulvermüller, F., Preissl, H., Lutzenberger, W., and Birbaumer, N. (1996). Brain rhythms of language: nouns versus verbs. *Eur. J. Neurosci.* 8, 937–941.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., and Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: behavioral and computational studies. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 393–406.

Rodd, J. M., Davis, M. H., and Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. Cortex* 15, 1261–1269.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol. Rev.* 111, 205–235.

Salinas, E., and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539–550.

Siekmeier, P. J., and Hoffman, R. E. (2002). Enhanced semantic priming in schizophrenia: a computer model based on excessive pruning of local connections in association cortex. *Br. J. Psychiatry* 180, 345–350.

Simmons, W. K., and Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol.* 20, 451–486.

Simmons, W. K., Martin, A., and Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cereb. Cortex* 15, 1602–1608.

Slotnick, S. D., Moo, L. R., Kraut, M. A., Lesser, R. P., and Hart, J. (2002). Interactions between thalamic and cortical rhythms during semantic memory recall in human. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6440–6443.

Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience* 101, 243–276.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1997). Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *J. Neurosci.* 17, 722–734.

Tallon-Baudry, C., Bertrand, O., Peronnet, F., and Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *J. Neurosci.* 18, 4244–4254.

Tyler, L. K., and Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends Cogn. Sci.* 5, 244–252.

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., and Levy, J. P. (2000). Conceptual structure and the structure of concepts: a distributed account of category-specific deficits. *Brain Lang.* 75, 195–231.

Ursino, M., Magosso, E., and Cuppini, C. (2009). Recognition of abstract

objects via neural oscillators: interaction among topological organization, associative memory and gamma band synchronization. *IEEE Trans. Neural Netw.* 20, 316–335.

Vigliocco, G., Vinson, D. P., Lewis, W., and Garrett, M. F. (2004). Representing the meaning of objects and action words: the featural and unitary semantic space hypothesis. *Cogn. Psychol.* 48, 422–488.

Warrington, E. K., and McCarthy, R. A. (1987). Categories of knowledge: further fractionations and an attempted integration. *Brain* 110, 1273–1296.

Warrington, E. K., and Shallice, T. (1984). Category specific semantic impairments. *Brain* 107, 829–854.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 July 2010; accepted: 20 November 2010; published online: 08 December 2010.

Citation: Ursino M, Cuppini C and Magosso E (2010) A computational model of the lexical-semantic system based on a grounded cognition approach. Front. Psychology ***1****:221. doi: 10.3389/fpsyg.2010.00221*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Ursino, Cuppini and Magosso. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

APPENDIX

All equations and parameter numerical values describing the semantic network can be found in a previous work (Ursino et al., 2009), hence are not reported here for briefness. The lexical network, and the learning rules used to train both the semantic and lexical aspects are new, hence are described in detail below.

THE LEXICAL NETWORK – MODEL EQUATIONS

In the following each element of the lexical area will be denoted with the subscripts *ij* or *hk* (*i, h* = 1, 2, …, M_1; *j, k* = *1, 2,*…, M_2) and with the superscript *L*. In the present study we adopted $M_1 = M_2 = 20$. Each single element is described via the following differential equation:

$$\tau^L \cdot \frac{d}{dt} x_{ij}^L(t) = -x_{ij}^L(t) + H^L\left(u_{ij}^L(t)\right); \tag{1}$$

τ^L is the time constant, which determines the speed of the answer to the stimulus, and $H^L(u^L(t))$ is a sigmoidal function. The latter is described by the following equation:

$$H^L\left(u^L(t)\right) = \frac{1}{1+e^{-\left(u^L(t)-\vartheta^L\right)p^L}}; \tag{2}$$

where ϑ^L defines the input value at which neuron activity is half the maximum (central point) and p^L sets the slope at the central point. Eq. 2 conventionally sets the maximal neuron activity at 1 (i.e., all neuron activities are normalized to the maximum).

The overall input, $u_{ij}^L(t)$, to a lexical neuron in the *ij*-position can be computed as follows

$$u_{ij}^L(t) = I_{ij}^L(t) + V_{ij}^F \tag{3}$$

$I_{ij}^L(t)$ is the input produced by an external linguistic stimulation. V_{ij}^F represents the intensity of the input due to synaptic connections from the semantic network; this synaptic input is computed as follows:

$$V_{ij}^F = \sum_h \sum_k W_{ij,hk}^F \cdot x_{hk} \tag{4}$$

where x_{hk} represents the activity of the neuron *hk* in the Feature Areas (see Ursino et al., 2009) and $W_{ij,hk}^F$ the strength of synapses. These synapses may have both an excitatory and an inhibitory component (say $W_{ij,hk}^{\mathrm{Fex}}$ and $W_{ij,hk}^{\mathrm{Fin}}$, respectively) which are trained in different ways (see Second Training Phase, below). Hence, we can write

$$W_{ij,hk}^F = W_{ij,hk}^{\mathrm{Fex}} - W_{ij,hk}^{\mathrm{Fin}}$$

FIRST TRAINING PHASE – MODEL EQUATIONS

In order to obtain asymmetric synapses between common and distinctive features, we assume that the Hebb rule depends on the average activity of both the post-synaptic and the pre-synaptic neurons, but the post-synaptic activity is compared with a threshold, to determine whether this neuron is (on the average) in the off or in the on state. Hence, we have

$$W_{ij,hk}\left(t+T_S\right) = \left[W_{ij,hk}(t) + \beta_{ij,hk} \cdot \left[m_{ij}(t) - \lambda\right] \cdot m_{hk}(t)\right]^+ \tag{5}$$

where λ is a threshold for comparing the post-synaptic activity and *m* denotes a moving average signal. The latter reflects the average activity during the previous 10 ms, as follows

$$m_{hk}(t) = \frac{\sum_{m=0}^{N_s-1} x_{hk}\left(t - mT_S\right)}{N_s} \tag{6}$$

The moving average of the post-synaptic activity ($m_{ij}(t)$) is computed with an equation analogous to Eq. 6. T_s is the sampling period and N_s the number of samples contained within 10 ms Symbols $[]^+$ in the right-hand member of Eq. 5 denotes the positive part, which is used to avoid that synapses among features become negative.

In order to assign a value for the learning factor, $\beta_{ij,hk}$, we assumed that inter-area synapses cannot overcome a maximum saturation value. This is realized assuming that the learning factor is progressively reduced to zero when the synapse approaches its maximum saturation. Furthermore, neurons belonging to the same area cannot be linked by a long-range synapse. We have

$$\beta_{ij,hk} = \begin{cases} \beta_0\left(W_{\max} - W_{ij,hk}\right) & \text{if } ij \text{ and } hk \text{ belong to different areas} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where W_{max} is the maximum value allowed for any synapse, and $\beta_0 W_{\max}$ is the maximum learning factor (i.e., the learning factor when the synapse is zero).

SECOND TRAINING PHASE – MODEL EQUATIONS

Long-range synapses among the lexical and the semantic networks are trained during a second phase, in which an object is presented to the network together with its corresponding word.

Synapses from the lexical network to the semantic network (say $W_{ij,hk}^L$) are learned using an Hebbian rule similar to that used in Eqs. 6 and 7. We can write

$$W_{ij,hk}^L(t+T_S) = W_{ij,hk}^L(t) + \beta_{ij,hk}^L \cdot x_{ij}(t) \cdot m_{hk}^L(t) \tag{8}$$

where $\beta_{ij,hk}^L$ represents the learning factor and $m_{hk}^L(t)$ is the averaged signal:

$$m_{hk}^L(t) = \frac{\sum_{m=0}^{N_s-1} x_{hk}^L\left(t - mT_s\right)}{N_s} \tag{9}$$

$$\beta_{ij,hk}^L = \beta_0^L\left(W_{\max}^L - W_{ij,hk}^L\right) \tag{10}$$

Conversely, synapses from the semantic network to the lexical network (i.e., parameters $W_{ij,hk}^F$ in Eq. 4) include both excitatory and inhibitory contributions:

$$W_{ij,hk}^F(t) = W_{ij,hk}^{\mathrm{Fex}}(t) - W_{ij,hk}^{\mathrm{Fin}}(t) \tag{11}$$

The excitatory portion is trained (starting from initially null values) using equations similar to Eqs. 5 and 7, but without long term depression and assuming that the *sum* of synapses entering a word must not overcome a saturation value (say $W_{\mathrm{sumMax}}^{\mathrm{Fex}}$). Hence

$$W_{ij,hk}^{\mathrm{Fex}}(t+T_S) = W_{ij,hk}^{\mathrm{Fex}}(t) + \beta_{ij,hk}^{\mathrm{Fex}} \cdot x_{ij}^{L}(t) \cdot m_{hk}(t) \tag{12}$$

$$\beta_{ij,hk}^{\mathrm{Fex}} = \beta_0^{\mathrm{Fex}} \left(W_{\mathrm{sumMax}}^{\mathrm{Fex}} - \sum_{lm} W_{ij,lm}^{\mathrm{Fex}} \right) \tag{13}$$

where the average activity $m_{hk}(t)$ is defined as in Eq. 9, and the sum in the right-hand member of Eq. 13 is extended to all synapses from the semantic network entering the neuron *ij* in the lexical network. The inhibitory synapses start from a high value (say $W_{\mathrm{Max}}^{\mathrm{Fin}}$) and are progressively withdrawn using an Hebbian mechanism:

$$W_{ij,hk}^{\mathrm{Fin}}(t+T_S) = \left[W_{ij,hk}^{\mathrm{Fin}}(t) - \beta_{ij,hk}^{\mathrm{Fin}} \cdot x_{ij}^{L}(t) \cdot m_{hk}(t) \right]^{+} \tag{14}$$

The function "positive part" ($[]^{+}$) is used in the right-hand member of Eq. 14 to avoid that these synapses become negative (i.e., that inhibition is converted to excitation).

All equations have been numerically solved in the software environment MATLAB with the Euler integration method (step 0.2 ms) and using the parameter values reported in **Table 2**. All inter-area weights were initially set at zero (that is a common choice in auto-associative networks).

REFERENCE

Ursino, M., Magosso, E., and Cuppini, C. (2009). Recognition of abstract objects via neural oscillators: interaction among topological organization, associative memory and gamma band synchronization. *IEEE Trans. Neural Netw.* 20, 316–335.

frontiers in
PSYCHOLOGY

PERSPECTIVE ARTICLE
published: 02 May 2011
doi: 10.3389/fpsyg.2011.00069

The other half of the embodied mind

Domenico Parisi*

Laboratory of Artificial Life and Robotics, Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

Edited by:
Diane Pecher, Erasmus University Rotterdam, Netherlands

Reviewed by:
Costantini Marcello, University of Chieti, Italy
Constantin A. Rothkopf, Goethe University, Germany

****Correspondence:***
Domenico Parisi, Laboratory of Artificial Life and Robotics, Institute of Cognitive Sciences and Technologies, National Research Council, Viale Marx 15, 00137 Rome, Italy.
e-mail: domenico.parisi@istc.cnr.it

Embodied theories of mind tend to be theories of the cognitive half of the mind and to ignore its emotional half while a complete theory of the mind should account for both halves. Robots are a new way of expressing theories of the mind which are less ambiguous and more capable to generate specific and non-controversial predictions than verbally expressed theories. We outline a simple robotic model of emotional states as states of a sub-part of the neural network controlling the robot's behavior which has specific properties and which allows the robot to make faster and more correct motivational decisions, and we describe possible extensions of the model to account for social emotional states and for the expression of emotions that, unlike those of current "emotional" robots, are really "felt" by the robot in that they play a well-identified functional role in the robot's behavior.

Keywords: embodied mind, emotion, robots

INTRODUCTION

In the Western cultural tradition the mind tends to be viewed as separated from the body and, in accordance with this tradition, the sciences of the mind try to understand the mind with no reference to the body. In the last few decades, however, this has changed. The cumulative and fast advances of the sciences of the body (neurosciences, evolutionary biology, genetics, the biological sciences more generally) make all attempts at studying the mind while ignoring the body less and less plausible. In fact, the idea that the mind is embodied and that to understand the mind it is necessary to take the body into consideration is being accepted by an increasing number of researchers and constitutes the premise of many important current investigations (Barsalou, 1999, 2008; Robbins and Aydede, 2009). The embodied view of the mind has led to a recognition of the importance of the actions with which the organism responds to the stimuli in determining how the world is represented in the organism's mind, in contrast to the traditional emphasis on mental representations as either entirely abstract or derived only from sensory input. This action-based view of the mind underlies a number of important ideas such as the grounding of symbols in the interactions of the organism with the physical environment (Harnad, 1990), the mental (neural) "simulation" of actions as a crucial component of all sorts of understanding (Gallese et al., 1996; Rizzolatti and Craighero, 2004), the mental representation of objects in terms not of their sensory properties but of the actions that the objects make possible (affordances; Gibson, 1977), the action-based nature of categories (Borghi et al., 2002; Di Ferdinando and Parisi, 2004). The embodied view of the mind is also reflected in computational models which reproduce the part of the body more directly linked to the mind, i.e., the brain (artificial neural networks) and, more recently, the entire body of the organism (robotics), and which have abandoned the disembodied view of the mind which is at the basis of artificial intelligence and of conceptions of the mind as symbol manipulation.

However, although the study of the mind can greatly benefit from an embodied conception of the mind, it still has to free itself from another tradition of Western culture which constitutes an obstacle to a complete understanding of the mind: the mind tends to be identified with cognition, that is, with knowing, reasoning, deciding, and acting. But cognition is only half of the mind. The other half of the mind is its emotional half and, although the two halves of the mind continuously interact and behavior is a result of both halves, no satisfactory account of the mind can be provided if the science of the mind is only "cognitive" science. Today one speaks of "embodied cognition," "grounded cognition," and the mental representation of objects in terms of the actions with which organisms respond to them. But organisms, including humans, do not only have knowledge, goals, and the capacity to act. They also have motivations and emotional states which play a crucial role in their behavior. Current embodied views of the mind tend to be concerned with the cognitive half of the mind but an embodied account of the mind must be extended to the other half of the mind, its emotional half and, in fact, some psychologists and neuroscientists are trying to extend the embodied conception of the mind to emotions (see, for example, Gallese, 2008; Freina et al., 2009; Glenberg et al., 2009).

Even if one assumes that the mind generally does not contain anything which is unrelated to sensory input and motor output (which is what embodied theories assume), the cognitive and the emotional halves of the mind may not function in the same way. Input to the brain can be input from the environment but also input from inside the body, and output from the brain can be external motor output but also changes in the internal organs and systems of the body, and these different sensory inputs and motor outputs may have different characteristics and consequences. This is why we need models that capture both the cognitive half and the emotional half of the mind and their interactions. These models should explicitly indicate both similarities and differences between embodied cognition and embodied emotion.

This also applies to the study of the mind through the construction of computational models or robots. Robots are the most appropriate tools for exploring embodied theories of the mind because, although in a very simplified form, they reproduce the body of organisms and the physical organ that controls the organisms' behavior (neuro-robots), and this is true for both physically realized robots and for robots which are simulated in a computer. However, current robots mostly try to reproduce the cognitive half of the mind but they ignore its emotional half. The robots displace themselves in the environment, move their arms and reach for objects, turn their eyes and their face, but they do not have emotions. Some current robots produce postures and movements of their bodies (mostly, the face) that in humans express emotions and they can recognize the expressed emotions of humans as a purely perceptual task, but they cannot be said to really have emotions and to really understand the emotions of others (Picard, 2000, 2003; Breazeal, 2002; Adolphs, 2005; Canamero, 2005; Dautenhahn et al., 2009; Robinson and el Kaliouby, 2009; cf. Arbib and Fellous, 2004; Fellous and Arbib, 2005). (For an attempt at understanding the functional role of emotions in behavior, see Ziemke, 2008.). The reason is quite simple. The cognitive half of the mind is the result of the interactions of the brain with the external environment or of processes self-generated inside the organism's brain (mental life). Current robots have artificial brains which interact with the external environment and, in some cases, can even self-generate inputs and respond to these self-generated inputs (Mirolli and Parisi, 2006, 2009; Parisi, 2007). But current robotics is an *external* robotics: robots reproduce the external morphology of an organism's body, the organism's sensory and motor organs, and the interactions of the organism's brain with the external environment. In contrast, the emotional half of the mind is the result of the interactions of the organism's brain with the organism's body and with the organs and systems that are inside the body. If we want to construct robots that can be said to really have emotions, what is needed is an *internal* robotics, that is, robots that have internal organs and systems with which the robot's brain can interact (Parisi, 2004). Only an internal robotics can help us to better understand the emotional half of the mind and to construct a complete embodied theory of the mind.

Computational models and, more specifically, robotic models are important to understand the mind. Theories in psychology tend to be expressed verbally but verbally expressed theories have limitations because words often have different meanings for different people and because verbally expressed theories may be unable to generate specific, detailed, and non-controversial predictions. Robots are an alternative way of expressing theories. The theory is used to construct a robot and therefore, in a sense, it can be directly observed and it can contain no ambiguity because otherwise the robot cannot be constructed. Furthermore, the theory generates many specific, detailed, and uncontroversial predictions which are the behaviors exhibited by the robot. These predictions can be empirically validated by comparing them with all sorts of empirical facts: the results of behavioral experiments, data on the ecology and past evolutionary history of the organism, and data on the organism's body and brain.

As we have said, robotic models are especially appropriate for formulating embodied theories of the mind because, by definition, a robot has a body and the robot's behavior clearly depends on its having a body. Furthermore, since the brain is part of the body, to be consistent robots should be neuro-robots, that is, robots whose behavior is controlled by a system that resembles the structure and functioning of the brain, i.e., an artificial neural network. This has the advantage that it becomes possible to examine the internal representations contained in the robot's "brain" (the patterns of activation and successions of patterns of activation in the robot's neural network) and to determine if they are embodied or non-embodied representations, i.e., if they reflect the robot's actions and the reactions of the robot's internal organs and systems to sensory input rather than the sensory input. (This is more difficult to do with "emotional" robots which are not controlled by neural networks but by symbolic systems such as those of Breazeal and Brooks, 2005.)

What we will do in this paper is describe a number of simple robots that may help us to construct an entire theory of mind as made up of a cognitive half and an emotional half.

THE STRATEGIC AND THE TACTICAL LEVEL OF FUNCTIONING OF ORGANISMS

The first step toward the construction of robots which have both a cognitive and an emotional mind is to construct robots that have many different motivations which cannot all be satisfied at the same time and therefore the robots have to decide at any given time which motivation to pursue with their behavior. Current robots tend to have only one motivation. In some cases their behavior is complex: for example, they may approach an object with their legs or wheels, reach the object with their arm, grasp the object with their hand, and put the object in their mouth. But behind this behavior there is a single motivation, say, the motivation to eat. In contrast, a robot's behavior may be very simple but the robot has many independent motivations. An example is a robot that has both a motivation to eat and a motivation to drink. The behavior which is needed to satisfy the motivation to eat or the motivation to drink may be very simple but the robot has to decide which motivation to satisfy if the two motivations cannot be both satisfied at the same time. Some current robots do have more than one motivation but it is their user which decides which motivation should control their behavior at any given time, i.e., what they must do, and in this sense current robots are not really autonomous. (For attempts at constructing robots that are motivationally autonomous or that take motivational decisions, cf. Brooks, 1986; Maes, 1990, although the last two references are to symbolically controlled robots. The difference between neural network and rule-based approaches to motivations and motivational decisions is discussed in Seth, 1998, 2007.)

Real organisms are different. They have many different and largely independent motivations and one of the most important aspects of their behavior is that they have to autonomously decide which motivation to pursue at any given time. In fact, the behavior of real organisms has two levels of functioning, the strategic or motivational level and the tactical or cognitive level (Ruini et al., 2010). At the strategic level the organism has to decide which one of its many different motivations will control its behavior at any particular time. At the tactical level, the organism has to execute the behavior which is appropriate to satisfy the motivation which has been decided at the strategic level. Imagine an organism which

lives in an environment containing both food and water and which to survive has to both eat and drink. Food and water are located in different places in the environment so that, at any given time, the organism has to decide whether to approach and reach food, and satisfy its motivation to eat, or to approach and reach water, and satisfy its motivation to drink. Once one of the two different motivations has been chosen at the strategic level, the tactical level of functioning of the organism generates the appropriate behavior which allows the organism to satisfy the motivation.

How is the decision taken at the strategic level? We assume that the decision is based on a simple mechanism. At any given time each motivation has an intensity which may vary from time to time, and the organism decides to pursue the motivation which currently has the highest intensity. We use the verb "to decide" but what we are talking about is implicitly expressed motivations and a purely physical mechanism that compares different motivational intensities, although in complex and verbal animals such as humans, the decision mechanism may involve talking to oneself, making explicit predictions, and producing explicit evaluations. The intensity of the different motivations is determined by two classes of factors: (1) the intrinsic intensity of the different motivations, which may depend on the overall adaptive pattern of the organism (e.g., the organism's body needs more food than water) or on the properties of the organism's environment (e.g., food is less abundant than water in the organism's environment), and (2) the current stimuli arriving to the organism's sensors from the external environment (e.g., the organism presently sees food rather than water) or from the organism's own body [e.g., the current level of nutrients (hunger) is lower than the current level of water (thirst)] or, in more complex animals such as humans, the stimuli which are self-generated (predicted, imagined, remembered) in the organism's brain. Once one motivation has been chosen at the strategic level, the organism produces the behavior that satisfies, or should satisfy, the motivation. Notice that since the intensity of the different motivations can change very rapidly because of the arrival of new stimuli and for other reasons, the organism should be able to shift very quickly from pursuing one motivation to pursuing another motivation, and this may happen even if the first motivation has not been satisfied or entirely satisfied.

That organisms function at two levels, the motivational and the cognitive level, is indicated by the fact that an individual may be good at the motivational level but not very good at the cognitive level while the opposite may be true for another individual. An individual may be good at choosing to satisfy its hunger rather than its thirst because its body needs energy rather than water but then it may not be very good at finding food. In contrast, another individual may be very good at finding both food and water but it makes the wrong motivational decisions, or is slow at deciding or, even more critically, is unable to decide and does nothing. Notice that to stay alive and possibly reproduce an individual should be sufficiently good at both levels since both being unable to decide correctly and rapidly and being unable to do what is necessary to satisfy the motivation decided at the motivational level may reduce the organism's chances of surviving and reproducing or the organism's well-being.

This indicates that a crucial component of behavior is that the organism's motivational decision mechanism must function effectively and efficiently. It must take the correct motivational decisions and it must take these decisions at the appropriate time, which in many cases means quickly. This is where emotions come in. Emotions are states of the organism's body that allow the motivational decision mechanism to function more effectively and more efficiently. Emotional states influence the current intensity of the organism's motivations in such a way that the organism functions better at the strategic level. The sight of a potential mate might cause an emotional state in the organism that increases the probability that the organism will pursue the motivation to mate rather than other motivations. The sight of a present danger may induce an emotional state that increases the probability that the organism will stop pursuing other motivations and try to avoid the danger, and will do this fast. In more complex organisms such as humans, even the thought of a mate or of a danger may increase the importance of the motivation to mate or the motivation to avoid the danger, with the consequence that the organism will be less likely to pursue other motivations.

In the next section we describe some simple robots that have to decide among different motivations and we show that if the robots have emotions their motivational decisions are more correct and efficient.

ROBOTS THAT HAVE EMOTIONS

We have constructed a robot which lives in an environment containing both food and a predator (Parisi and Petrosino, 2010; Ruini et al., 2010). To remain alive the robot has both to eat the food and to avoid being killed by the predator. The predator is not always present but when it appears the robot has to cease looking for food and escape from the predator. The neural network controlling the robot's behavior has sensory input units encoding the presence and location of both food and predator, motor output units encoding the movements that allow the robot to displace itself in the environment, and an intermediate layer of internal units. To eat the robot has to approach and reach the food elements while to avoid being killed by the predator the robot has to avoid physical contact with the predator. The connection weights of the robot's neural network are developed using a genetic algorithm with a population of robots that reproduce selectively and with the constant addition of random variations to the inherited connection weights (Mitchell, 1998). The robots are simulated Khepera robots (Nolfi and Gigliotta, 2010). They have a circular body with two wheels allowing the robot to displace itself in the environment and sensory organs allowing the robot to perceive what is in front of the robot within a given distance from the robot.

We compare two populations of robots. In one population the robots' neural network has the architecture we have already described. In the other population we add to the robot's neural network an "emotional circuit" made up of a certain number of units which receive activation from the sensory units encoding the presence of the predator and send their activation to either the internal units or the output units and therefore influence the robot's behavior. These emotional units have special properties compared to the other internal units of the robots' neural network. They have no "bias," they have an activation threshold, and an emotional unit is not only active in the particular cycle in which activation arrives to the emotional unit from the input units but its activation may persist in subsequent cycles. The unit's activation threshold and

the parameters that control how the activation of a unit persists in subsequent cycles all have evolved values. (For a more detailed description of the emotional circuit and of the results of the simulations, see Parisi and Petrosino, 2010.)

The results of the simulations show that the robots with the emotional circuit reach higher level of performance (they live longer) than the robots which do not have the emotional circuit in their brain. If we look at the robots' behavior, we see that they immediately cease looking for food and fly away when the predator appears, thereby reducing the probability of being reached and killed by the predator. The robots without the emotional circuit are less fast at shifting from the motivation to eat to the motivation to avoid being killed by the predator, and this leads to shorter lives.

In other simulations the robots have to make other motivational decisions: they have to decide whether to eat or drink, whether to eat or look for a mate, whether to eat or take care of their offspring, or whether to eat or rest when their body incurs some physical damage that can be healed by resting. The neural network of the robots has sensory units, motor units, and an intermediate layer of internal units and, in addition, it has internal inputs units encoding the current level of energy inside their body (hunger sensors) and, for the robots that have to both eat and drink, also internal input units encoding the level of water (thirst sensors), while the robots which can incur physical damage have internal input units encoding the presence of physical damage (pain sensors).

All these robots have to take motivational decisions, and their "fitness" depends on both their ability to take the appropriate motivational decisions with the required rapidity and their ability to produce the behavior which satisfies the motivation which has been chosen. The results of the simulations indicate that for all robots the possession of an emotional circuit allows them to take better and more rapid motivational decisions so that their fitness is higher than the fitness of the robots lacking the emotional circuit.

The robots we have described respond to the input from the external environment and/or from within their body not only with actions but also with emotional states, i.e., states of their emotional circuit, that make their motivational decisions more effective and efficient and therefore their behavior more "fit." The emotional circuit of these robots is very simple. The emotional units receive activation from the external environment (e.g., the sight of a predator) and/or from inside the body (e.g., the current level of energy in the robot's body) and they send their activation to the internal units or to the motor units of the robot's neural network, thereby influencing the robot's behavior. Real organisms are more complex. The equivalents of our robots' emotional units send their activation to various internal organs and systems such as the heart, the gut, the hormonal system, to the muscles of the face, and to other parts of the body, and receive activation from all these parts of the body, and it is this activation which influences the organism's motivational decisions and therefore the organism's behavior. In fact, as originally proposed by James and Lange (1922), and recently elaborated by Damasio (1994, 2004) and LeDoux (1996, 2000), felt emotional states are largely the product of these interactions. However, although very simple, our robots can be said to have emotions and to implement an embodied theory of the other half of the mind.

EMOTIONS AND SOCIALITY

The robots we have described in the preceding section do not have sociality. They live alone in their environment, and predators, mates, and offspring are not robots themselves but are only objects which are present in the environment and are perceived by the robots. Many animals, and especially humans, are social, which means that they live with conspecifics and a large portion of their adaptive pattern consists in being able to interact appropriately with conspecifics. What is the relation of emotions to sociality?

Emotions and sociality are linked in two distinct ways. Many motivations in social animals can be satisfied only with the participations of conspecifics, i.e., they are social motivations. For example, in species that reproduce sexually, an individual cannot satisfy its motivation to have offspring without the participation of an individual of the opposite sex. Since emotional states help the strategic level of functioning of the organism to take better motivational decisions, social animals tend to have social emotional states that allow them to take better motivational decisions when these decisions concern social motivations. The other way in which emotions and sociality are related is that emotional states are associated with postures and movements of the organism's body which can be perceived by other individuals and therefore can inform these individuals about the emotional states of the organism.

Examples of social motivations are the motivation to mate with another individual, the motivation to take care of one's offspring, and the motivation to be helped or to avoid being damaged by another individual. All these motivations enter into the motivational decisions of the organism which has to decide which motivation to pursue at any given time. Hence, social organisms tend to have social emotional states which regulate the current intensity of their social motivations and positively influence their motivational decisions that have to take into account social motivations.

It may not be too difficult to construct robots that have social emotions. In some recently completed simulations, the robots can be either female or male and to reproduce a robot has to mate with a robot of the opposite sex. The robots live in an environment with both food and other robots of the two sexes and to leave their genes to the next generation they have both to eat to remain alive and to approach and reach a robot of the opposite sex to reproduce. Female and male robots have different colors so that they can be recognized as females or males by conspecifics. But there is a further complication. After reproductively mating with a male robot, a female robot cannot reproduce for a certain number of time steps and during this non-reproductive period the female robot changes its color and this change of color is perceived by the other robots. A pregnant female knows its current state because its neural network has internal input units encoding the pregnant state of its body. The results show that while the average number of offspring is necessarily identical for females and males, males have more reproductive variability than females, that is, there are males with many offspring and males with very few or no offspring while females all have more or less the same number of offspring. Furthermore, the robots alternate appropriately between looking for food and looking for mates, with reproductive females behaving differently toward males compared to non-reproductive females, and males approaching only reproductive females.

The next step is to add an emotional circuit to the neural network of these robots and to see if in this case too the possession of the emotional circuit leads to better performance. What would make the states of this circuit *social* emotional states is that, unlike the emotional states of the robots described in the preceding Section, the emotional states of these robots would be activated by the sight of a conspecific and they would allow a robot to take better motivational decisions about what to do socially.

This leads us to the second aspect that links emotions to sociality. The emotional states of the robots we have just described are social only in the sense that they are associated with social motivations but they are not social in the sense that they are expressed, that is, they are communicated to other individuals. What appears to be really important to understand the role of emotions in sociality is to construct robots that express their emotional states and in this manner cause other individuals to know their emotional states and to be influenced in their behavior by this knowledge.

How can we construct robots that express emotions they really have? As we have said, current "emotional" robots express emotions that they do not have, that is, that do not play any functional role in their behavior. To construct robots that express emotional states which they really have it is necessary to link the emotional circuit of the robots we have described in the preceding section to postures and movements of the robot's body that can be perceived by other robots. In the robots described in the preceding section, the emotional circuit influences the manner in which the robot responds to both external and internal inputs but it has no interactions with the rest of the robot's body. As we have said, this is not so in real organisms, in which the emotional circuit of the brain sends its activation to other parts of the body and these parts of the body respond by sending activation to the brain. As we have also said, some of the parts of the body activated by the emotional circuit are external and therefore their state or change of state (postures and movements) can be perceived by another individual. In this way an individual may know the emotional states of another individual by observing the postures and movements of the body of the other individual that result from the activations sent to the body by its emotional units.

Why should the emotional states of an individual be reflected (expressed) in the postures and movements of the individual's body and therefore be accessible to the sensors of another individual? Why should the other individual be able to understand the emotional states of the individual by perceiving the postures and movements of its body? Postures and movements of the body may be simply a by-product of having emotional states, with no specific adaptive value. But the richness of the expression of emotions in some animals, and especially humans, seems to indicate that expressing one's emotions has adaptive value and has evolved for this reason. This adaptive value is informing other individuals of one's emotional states. The adaptive value appears to consist in two things: predictability and manipulation. For the other individual it may have adaptive value to know the emotional states of the first individual by perceiving the postures and movements of its body. Adapting to the environment is to a large extent to be able to predict the future state of the environment given its present state in order to prepare for the future state. For social animals an important component of their adaptive environment is constituted by conspecifics in that many of the motivations of an individual can only be satisfied with the participation of other individuals. Therefore, for a social individual it is important to be able to predict the behavior of other individuals, and knowing the emotional states of other individuals is a powerful predictor of their behavior. This may explain why social animals (of some complexity) tend to attend to and to be able to understand the emotional states of other individuals as these emotional states are expressed by the postures and movements of certain parts of their body. But why should an individual care to express its emotional states through the postures and movements of its body so that another individual can know them? The explanation in this case is not in terms of predictability but in terms of manipulation, where manipulation is behaving in such a way that the other individual will behave in some desired manner. In other words, an individual will express its emotional states because this will induce another individual which is informed of such emotional states to behave in ways which are desirable for the first individual.

Both functions of the expression of emotional states can be simulated with robots possessing an emotional circuit which interacts with the robot's body so that specific states of the circuit will cause specific postures and movements in the robot's body which can be perceived by other robots. This will allow the first robot to let its emotional state be known by other robots and therefore to influence their behavior and will allow the other robots to predict the behavior of the first robot.

Being able to predict X is an important component of (or perhaps the same thing as) understanding X. What is for individual A to understand individual B? The embodied theory of sociality says that A understands B by "simulating" B's perceived actions in its own brain (Gallese, 2010). But the behavior of B does not only consist in doing actions but also in expressing emotions, and sociality implies understanding not only B's actions but also B's expressed emotions. Hence, a more complete theory of sociality proposes that A understands B not only by observing B's actions and "simulating" these actions in its own brain but also by observing B's expressed emotions and "simulating" these emotions in its own brain–body. (Remember that, while the cognitive half of the mind is a product of the brain, the emotional half is a product of the interaction between the brain and the body.) There is an accumulating experimental literature on the role of expressed emotions in sociality. For example, Bayliss et al. (2007) have shown that how objects are affectively evaluated by an individual is influenced by the gaze and emotional expression of another individual, and Ferri et al. (2010) that the emotional expression of another individual may affect the goal-directed behavior of the observer. These are among the experimental results which our robots that have emotions should be able to replicate.

EMOTIONS AND LANGUAGE

An extension of the embodied view of mind to language proposes, and tries to show with experiments, that we understand nouns by internally representing in our brain the action with which we respond to the object designated by the noun and we understand verbs by "simulating" in our brain the action designated by the verb (Fischer and Zwaan, 2008). However, if the mind has two halves, the cognitive and the emotional half, an embodied conception of

language should give an account of both the cognitive and emotional meanings of words. What we do in this section is describe some simple robots that (begin to) have language, and to suggest how words can evoke in these robots not only a cognitive meaning but also an emotional meaning.

Most words are sounds, or phono-articulatory movements that produce acoustic sounds, which co-vary with specific objects (nouns) or actions (verbs). If objects and actions evoke emotional states, words will also evoke emotional states. The "cognitive" orientation of most current theories of the mind can also be seen in the privilege accorded to the "cognitive" component of the meanings of words rather than to their "emotional" component, where the cognitive component refers to the perceptual properties of the object designed by a noun or to the action designed by a verb while the emotional component refers to the emotional states evoked by the noun or verb. But if we want to construct robots that can be said to have language, it will be necessary that their words evoke emotional states, not only "cognitive meanings," in other robots. This requires an appropriate robotic model of the meanings of words. We will now briefly describe such a model. (For a more detailed description of the model, see Parisi, 2010. For robotic models of nouns and verbs, see Cangelosi and Parisi, 2001.)

Our robots live in an environment with other robots and their neural network is made up of two sub-networks, the non-linguistic sub-network (NoL) and the linguistic sub-network (L). NoL is made of input units encoding non-linguistic sensory input (perceived objects and perceived actions of another robot) and output units encoding non-linguistic actions (e.g., reaching and grasping an object). L has input units encoding linguistic sounds (produced by another robot) and output units encoding phono-articulatory movements that produce linguistic sounds. Both NoL and L have a layer of internal units which connects the input units to the output units. In addition, the two internal layers also have horizontal connections linking the internal units of NoL to those of L, and vice versa. Given these horizontal connections, perceiving an object can lead to executing a non-linguistic action but also to executing a phono-articulatory action, i.e., producing the word which designates the object. And hearing a word may lead to reproducing the word by executing the appropriate movements of one's phono-articulatory organs (imitating the sound of the word) or to executing a non-linguistic action. When the robot learns language (which, in children, begins at 1 year of age), the robots learns the appropriate connection weights for these horizontal connections. What is the role of these connection weights? In the robot's social experience one specific linguistic sound tends to co-vary with one specific object or action, and vice versa. The robot incorporates these co-variations in the connection weights of its neural network so that when the robot hears a linguistic sound and activation spreads from the internal layer of L to the internal layer of NoL, the pattern of activation evoked in the internal units of NoL is more or less the same pattern of activation evoked by perceiving the object or action which co-varies with the linguistic sound. When the robot perceives an object or action, activation spreads from the internal units of NoL to the internal units of L, which causes the robot to produce the sound which in its experience co-varies with that object or action. In other words, the robot is able to both produce and understand language.

A crucial assumption of the model we have described is that there is no entity in the robot's neural network which can be called *the* meaning of a word. If we assume that the internal units of NoL have internal connections linking the units of the *same* layer, when an activation pattern is evoked in the internal units of NoL, this activation pattern will evoke a second activation pattern in the same or other units, then a third, and so on, so that the initial activation pattern is only the first step of a process which has no natural end and no fixed boundaries and is influenced by a number of factors such as the linguistic and non-linguistic context in which the word is being experienced by the robot, inter-individual differences among the robots, the frequency with which a word has been experienced by the robot, and others. Notice that the internal layer of NoL is not a specialized "semantic module" but is the entire brain (minus the L sub-network). This implies that the spreading of activation in NoL can invade the emotional circuit of the robot's neural network and in this manner it can trigger emotional states in the robot. These emotional states triggered by words constitute the emotional meaning of words. Words have emotional meaning if a word co-varies in the robot's experience with non-linguistic experiences that cause emotional states in the robot.

This of course is only a first step toward a robotic account of the emotional meaning of words. Here is a list of interesting questions that this account should be able to answer. Why, while all words have a cognitive meaning, only some words appear to have an emotional meaning? Why the cognitive meaning of words appears to be more well-defined, articulated, and specific than their emotional meaning? How can we capture this fact with our robots? Words often appear to have emotional meaning not in isolation but in the context of other words, i.e., in sentences (Havas et al., 2007). Why this is so? In some experiments it has been shown that abstract words tend have more emotional meaning than concrete words (Kousta et al., 2009a). Why? Other experiments indicate that emotional words are processed more rapidly than non-emotional words (Kousta et al., 2009b). Why?

SUMMARY

Theories expressed as robots have two important advantages compared to verbally expressed theories: they are unambiguous because they refer to things that can observed and measured and they generate many detailed and non-controversial empirical predictions, which are the behaviors of the robot and what happens in the robot's "brain." We have described some robotic models that begin to address the question of how to develop a theory of mind that takes into consideration both the cognitive and the emotional halves of the mind. We have shown that if we add an emotional circuit to the neural network that controls a robot's behavior, the robot's behavior becomes more effective because the emotional circuit allows the robot to take more correct and faster motivational decisions. Unlike current "emotional" robots, robots that possess this circuit can be said to actually *have* emotions in that one can show that the circuit plays a clear functional role in the robot's behavior. We have then addressed the question of how emotions are related to sociality and we have distinguished two aspects of this relation. Robots can have social motivations, that is, motivations that can only satisfied with the participation of conspecifics, and emotions can be said to be social if they cause the robot to make better motivational decisions when the choice set includes

social motivations. The other link between emotions and sociality is the expression of emotions. The emotional circuit allows the brain (neural network) to interact not only with what is inside the body but also with the external body, causing postures and movements of some parts of the body (especially the face) that can be perceived by other robots. We have advanced the hypothesis that this expression of emotions has two adaptive advantages: it allows a robot that expresses its emotions to manipulate the behavior of other robots that perceive the expressed emotions and it allows the robot which perceives the emotions expressed by another robot to predict and anticipate the behavior of the other robot. Finally, we have briefly discussed which type of neural network should control the behavior of robots that have language and how words can have both cognitive and emotional meanings for a robot.

The robotic models we have described are very simple and very tentative and one still has to show that they are able to explain (reproduce) the constantly accumulating empirical facts about brain and behavior that involve the other half of the mind. (For some attempts at simulating the brain involved in emotions, see Rolls and Treves, 1998; Rolls, 1999; Avila-Garcia and Canamero, 2004.) Furthermore, we have said that current "emotional robots" do not actually have emotions because emotions do not play any clearly identifiable functional or adaptive role in their behavior, but the results obtained with our robotic models should be compared with those of alternative models, both robotic and non-robotic. An important test of the models will be their ability to reproduce not only the results of experiments but also other empirical facts such as inter-individual differences in the emotional half of the mind, i.e., differences in personality and character rather than in cognitive ability, the pathologies of the emotional half of the mind, i.e., psychiatric and psychological disturbances rather than neurological ones (Stein and Ludik, 2008), and the expression of the emotional half of the mind in the production and appreciation of works of art.

REFERENCES

Adolphs, R. (2005). "Could a robot have emotions? Theoretical perspectives from social cognitive neuroscience," in *Who Needs Emotions? The Brain Meets the Robot*, eds J.-M. Fellous and M. A. Arbib (Oxford: Oxford University Press), 9–25.

Arbib, M. A., and Fellous, J. M. (2004). Emotions: from brain to robot. *Trends Cogn. Sci.* 8, 554–561.

Avila-Garcia, O., and Canamero, L. (2004). "Using hormonal feedback to modulate action selection in a competitive scenario," in *From Animals to Animats 8: Proceedings of the 8th International Conference on Simulation of Adaptive Behavior*, eds S. Schaal, A. J. Ijspeert, A. Billard, S. Vijayakumar, J. Hallam, and J. A. Meyer (Cambridge, MA: Bradford Book), 243–252.

Barsalou, W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609.

Barsalou, W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bayliss, A. P., Frischen, A., Fenske, M. J., and Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition* 104, 644–653.

Borghi, A. M., Di Ferdinando, A., and Parisi, D. (2002). "The role of perception and action in object categorization," in *Connectionist Models of Cognition and Perception*, eds J. A. Bullinaria and W. Lowe (Singapore: World Scientific), 40–50.

Breazeal, C. (2002). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Breazeal, C., and Brooks, R. (2005). "Robot emotion. A functional perspective," in *Who Needs Emotions? The Brain Meets the Robot*, eds J.-M. Fellous and M. A. Arbib (Oxford: Oxford University Press), 271–310.

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE J. Robot. Autom.* 2, 14–23.

Canamero, L. (2005). Emotion understanding from the perspective of autonomous robots research. *Neural Netw.* 4, 445–455.

Cangelosi, A., and Parisi, D. (2001). "How nouns and verbs differentially affect the behaviour of artificial organisms," in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, eds J. D. Moore and K. Stenning (Hillsdale, NJ: Erlbaum), 170–175.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Grosset/Putnam.

Damasio, A. (2004). "William James and the modern neurobiology of emotion," in *Emotion, Evolution, and Rationality*, eds D. Evans and P. Cruse (Oxford: Oxford University Press), 175–188.

Dautenhahn, K., Nehaniv, C. L., Walters, M. L. Robins, B., Kose-Bagci, H., Assif Mirza, N., and Blow, M. (2009). KASPAR: a minimally expressive humanoid robot for human–robot interaction research. *Appl. Bionics Biomech.* 6, 369–397.

Di Ferdinando, A., and Parisi, D. (2004). "Internal representation of sensory input reflects the motor output with which organisms respond to the input," in *Seeing and Thinking*, ed. A. Carsetti (Boston, Kluwer), 115–141.

Fellous, J.-M., and Arbib, M. A. (2005). *Who Needs Emotions? The Brain Meets the Robot*. Oxford: Oxford University Press.

Ferri, F., Stoianov, I. P., Gianelli, C., D'Amico, L., Borghi, A. M., and Gallese, V. (2010). When action meets emotions: how facial displays of emotion influence goal-related behavior. *PLoS ONE* 10, e13126. doi: 10.1371/journal.pone.0013126

Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol.* 61, 225–250.

Freina, L., Baroni, G., Borghi, A. M., and Nicoletti, R. (2009). Emotive concept-nouns and motor responses: attraction or repulsion? *Mem. Cogn.* 37, 493–499.

Gallese, V. (2008). Empathy, embodied simulation and the brain. *J. Am. Psychoanal. Assoc.* 56, 769–781.

Gallese, V. (2010). "Embodied simulation and its role in intersubjectivity," in *The Embodied Self: Dimensions, Coherence, and Disorders*, eds T. Fuchs, H. C. Sattel, and P. Henningsen, (Stuttgart: Schattauer), 78–92.

Gallese, V., Fadiga, L. Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609.

Gibson, J. J. (1977). "The theory of affordances," in *Perceiving, Acting, and Knowing. Toward an Ecological Psychology*, eds R. Shaw and J. Bransford (Hillsdale, NJ: Erlbaum), 67–82.

Glenberg, A. M., Webster, B. J., Mouilso, E., Havas, D., and Lindeman, L. M. (2009). Gender, emotions, and language comprehension. *Emot. Rev.* 1, 151–161.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Havas, D. A., Glenberg, A. M., and Rinck, M. (2007). Emotion simulation during language comprehension. *Psychon. Bull. Rev.* 14, 436–441.

James, W., and Lange, C. (1922). *The Emotions*. Baltimore: Williams and Wilkins.

Kousta, S., Vigliocco, G., Vinson, D., and Andrews, M. (2009a). "Happiness is an abstract word: the role of affect in abstract knowledge representation," in *Proceedings of the 31st Meeting of the Cognitive Science Society*, Amsterdam.

Kousta, S., Vinson, D. P., and Vigliocco, G. (2009b). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition* 112, 473–481.

LeDoux, J. E. (1996). *The Emotional Brain. The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184.

Maes, P. (1990). "A bottom-up mechanism for behavioural selection in an artificial creature," in *From Animals to Animats*, eds J.-A. Meyer and S. W. Wilson (Cambridge, MA: MIT Press), 169–175.

Mirolli, M., and Parisi, D. (2006). "Talking to oneself as a selective pressure for the emergence of language," in *Proceedings of the 6th International Conference on the Evolution of Language*, eds A. Cangelosi, A. D. M. Smith, and K. Smith (Singapore: World Scientific), 214–221.

Mirolli, M., and Parisi, D. (2009). Language as a cognitive tool. *Minds Mach.* 14, 517–528.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Nolfi, S., and Gigliotta, O. (2010). "Evorobot: a tool for running experiments on the evolution of communication," in *Evolution of Communication and Language in Embodied Agents*, eds S. Nolfi and M. Mirolli (Berlin: Springer), 297–301.

Parisi, D. (2004). Internal robotics. *Connection Sci.* 16, 325–338.

Parisi, D. (2007). "Mental robotics," in *Artificial Consciousness*, eds A.

Chella and R. Mazzotti (Thorverton: Imprinting Academic), 191–211.
Parisi, D. (2010). Robots with language. *Front. Neuorobot.* 4:10. doi: 10.3389/fnbot.2010.00010.
Parisi, D., and Petrosino, G. (2010). Robots that *have* emotions. *Adapt. Behav.* 18, 453–469.
Picard, R. W. (2000). *Affective Computing*. Cambridge, MA: MIT Press.
Picard, R. W. (2003). "What does it mean for a computer to "have" emotions?" in *Emotions in Humans and Artifacts*, eds R. Trappl, P. Petta, and S. Payr (Cambridge, MA: MIT Press), 87–102.
Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
Robbins, P., and Aydede, M. (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.
Robinson, P., and el Kaliouby, R. (2009). Computations of emotions in man and machines. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3441–3447.
Rolls, E. T. (1999). *The Brain and Emotion*. Oxford: Oxford University Press.
Rolls, E. T., and Treves, A. (1998). *Neural Networks and Brain Function*. Oxford: Oxford University Press.
Ruini, F., Petrosino, G., Saglimbeni, F., and Parisi, D. (2010). "The strategic level and the tactical level of behaviour," in *Advances in Cognitive Systems*, eds J. Gray and S. Nefti-Meziani (Herts: IET Publisher), 271–299.
Seth, A. K. (1998). "Evolving action selection and selective attention without action, attention, or selection," in *Proceedings of the 5th Conference on the Simulation of Adaptive Behaviour*, eds R. Pfeifer, B. Blumberg, J-A. Meyer, and S. Wilson (Cambridge, MA: MIT Press).
Seth, A. K. (2007). The ecology of action selection: insights from artificial life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1545–1558.
Stein, D. J., and Ludik, J. (eds). (2008). *Neural Networks and Psychopathology*. Cambridge: Cambridge University Press.
Ziemke, T. (2008). On the role of emotion in biological and robotic autonomy. *Biosystems* 91, 410–408.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 August 2010; accepted: 04 April 2011; published online: 02 May 2011.
Citation: Parisi D (2011) The other half of the embodied mind. Front. Psychology ***2****:69. doi: 10.3389/fpsyg.2011.00069*
This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.
Copyright © 2011 Parisi. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 15 November 2010
doi: 10.3389/fpsyg.2010.00195

The neuroscience of storing and molding tool action concepts: how "plastic" is grounded cognition?

J. C. Mizelle and Lewis A. Wheaton*

School of Applied Physiology, Georgia Institute of Technology, Atlanta, GA, USA

Edited by:
Anna M. Borghi, University of Bologna, Italy

Reviewed by:
Binkofski Ferdinand, University of Luebeck, Germany
Maurizio Gentilucci, University of Parma, Italy

****Correspondence:***
Lewis A. Wheaton, School of Applied Physiology, Georgia Institute of Technology, 281 Ferst Drive, Room 104, Atlanta, GA 30332-0356, USA.
e-mail: lewis.wheaton@ap.gatech.edu

Choosing how to use tools to accomplish a task is a natural and seemingly trivial aspect of our lives, yet engages complex neural mechanisms. Recently, work in healthy populations has led to the idea that tool knowledge is grounded to allow for appropriate recall based on some level of personal history. This grounding has presumed neural loci for tool use, centered on parieto-temporo-frontal areas to fuse perception and action representations into one dynamic system. A challenge for this idea is related to one of its great benefits. For such a system to exist, it must be very plastic, to allow for the introduction of novel tools or concepts of tool use and modification of existing ones. Thus, learning new tool usage (familiar tools in new situations and new tools in familiar situations) must involve mapping into this grounded network while maintaining existing rules for tool usage. This plasticity may present a challenging breadth of encoding that needs to be optimally stored and accessed. The aim of this work is to explore the challenges of plasticity related to changing or incorporating representations of tool action within the theory of grounded cognition and propose a modular model of tool–object goal related accomplishment. While considering the neuroscience evidence for this approach, we will focus on the requisite plasticity for this system. Further, we will highlight challenges for flexibility and organization of already grounded tool actions and provide thoughts on future research to better evaluate mechanisms of encoding in the theory of grounded cognition.

Keywords: grounded cognition, MOSAIC, apraxia, aging, development, tools, action

INTRODUCTION

The theory of grounded cognition has offered a fascinating window into mechanisms of storage and recall of concepts. Generally, the act of simulation of past events is an important aspect of current theories of grounded cognition (Decety and Grezes, 2006). For a given object-related interaction, information from the modal senses for perception (e.g., vision, somatosensation), action (e.g., kinethesis, proprioception), and introspection (e.g., affect), are blended into a representation of that experience (Barsalou, 2008). As knowledge is needed to later represent that object, this multimodal information (perception, action, introspection) is recalled to simulate the brain states associated with that object (Barsalou, 2008). Through this process, both cognitive and motoric functions can take advantage of the ideas offered by this theory. Here, we will focus on a specific aspect of motoric function: implementing tools and objects in various actions. For the purposes of this work, we will use the word "tool" to define non-animate artifacts that operate on something by a user (e.g., spoon). "Object" will be that which is operated on by a tool (e.g., mug). "Action" will generally involve mutual use of the tool and object to accomplish a goal (e.g., stir coffee). We will go further to differentiate knowledge of tool manipulation from knowledge of tool function, where manipulation is "how" to use a tool and function is "why" to use a tool (Buxbaum et al., 2000). Some of the principles presented can be applicable for many types of movements, such as communicative gesture, but we are limiting the scope of this to tool use alone.

GROUNDED COGNITION RELEVANT TO TOOL USE

An idea that is being pursued is that acts relevant to tool use are grounded, where tool knowledge is grounded based on past acts. Thus, hammer does not come with an operant functional definition, but more precisely is based on how we have used a hammer in our lives. This will "ground" a representation of the tool in a state that encodes usage. Of use is also to note that this grounding may differ from person to person, although possibly in subtle ways. For example, the two co-authors of this work have slightly distinct histories of a hammer in multiple obvious and atypical contexts. Further, actions can become encoded deeper in context, based on situational rules and tendencies. For example, occasionally a spoon is more likely/useful than a plastic stirrer for coffee (more will follow of this example). The act of stirring coffee is grounded with multiple potential items based on typical contexts.

Basic aspects of tools and objects have a clear foundation, making the idea of grounding tool use attractive. Related to grounded views of tool usage, vital work in this field was performed that demonstrated how objects and actions are coupled based on perception of how the object is held (Tucker and Ellis, 1998). Such an idea has been advanced by suggesting that seen objects can activate motor representations as a result of functional knowledge of the seen object (Anderson et al., 2002). Many other studies have established this important idea (Grezes et al., 2003; Handy et al., 2003; Tucker and Ellis, 2004; Rice et al., 2007). A strong basis for the argument of grounded cognition in tool and object-related action is seen in

a review by Grafton (2009) which clearly illustrates that action can be recognized using a system where subjects can effectively match seen actions with their own representations for action understanding (Grafton, 2009). Further, theories have been well proposed to suggest how vision and action knowledge are integrated to allow for the fusing of perception and action (Buxbaum and Kalenine, 2010). Similar ideas are given in the context of action. It is plausible that a key basis for our ability to act based on prior experiences through simulation (Barsalou, 1999). Here, we can recall our prior experiences and apply them as we are carrying out an act (Barsalou, 2008). A similar mechanism can be used when we have to represent a seen action, by recalling the states of the observer in tool use (Kellenbach et al., 2003; Jarvelainen et al., 2004), object grasping (Pierno et al., 2009; Valyear and Culham, 2010), and action (Frey and Gerry, 2006; Evangeliou et al., 2009) studies.

A POTENTIAL LIMITATION? – "PLASTICITY" OF GROUNDED COGNITION

While grounded cognition is attractive for the ideas above, a common thread is that sensorimotor experience drives this process for tools/object action knowledge. However, two aspects of this theory should be addressed related to this idea of how we actually ground tool/object/action knowledge.

First, what if we have limited or no prior experience with a tool? How well is it grounded? The recall of a familiar tool (e.g., hammer) for action or recognition is clearly left parietofrontal involved (Lewis, 2006). However, there is suggestion that tool representations differ based on our relative history with a tool (Vingerhoets et al., 2009). Our questions regarding grounded cognition come in consideration of the process and neural architecture used to *encode* a tool/object representation, then how this grounded representation may change over time, particularly to contract into a limited-use tool representation.

To our second question, how can a grounded tool/object/action be modified? As another example, how might we come to the idea that a hammer can be used to re-seal a paint can (without any sensorimotor experience of this), when we likely first learned it as an implement to drive in a nail? Similar examples are provided in other work, where grounded views may result in a need to break out of "functional fixedness" on a regular basis (Wilson, 2002). Thus, we focus on the idea of the "plasticity of grounded cognition." That is, how do these parietofrontal tool–object representations develop, grow, and change over time based on our experiences and creativity? Further, what is the neuroscience of this process? As will be seen, these two questions are not fully segregated from each other.

CREATION/MATURATION OF A GROUNDED REPRESENTATION

In consideration of the first question (progression of novel tool/object/action representations), the study of Vingerhoets (2008) offers an intriguing point. Here, unlike familiar tools, recognition of unfamiliar and rarely used tools heavily engages left temporo-occipital areas. As unfamiliar tools natively have no clear action (specifically sensorimotor) representation, this will clearly prevent simulation-based action knowledge, and should fail to generate clear action–observation knowledge related to such tools (as in Grafton, 2009). However, rarely used tools likely would have some limited action (sensorimotor) knowledge, and may engage action–observation related networks. Yet, the recognition of both classes of tools falls outside of the suggested tool-related network for knowledge (that codes recognition and action). Similarly, others have recently shown that prefrontal and mediotemporal areas of the left hemisphere are important for understanding the use of a new tool (Menz et al., 2010).

Although the exact length of time required for a tool or object to become grounded is unknown, we can look to the motor control and learning literature for insight. It is well known that motor learning in humans occurs in three distinct phases: Initial stage, where trial-and-error is required to establish new sensory information with correct motor commands (e.g., derivation of a novel sensorimotor association); Intermediate stage, where the newly acquired sensorimotor association is learned through practice (e.g., consolidation of the sensorimotor map); Advanced stage, where the working memory access of the sensorimotor association is no longer required and movements can be performed with less reliance on sensory feedback processing and attention (Halsband and Lange, 2006). These stages of learning, although variable, begin immediately and may continue to develop for days to weeks (Karni et al., 1998), at which point the skill is retained (Karni and Sagi, 1993). Some interesting work suggests that, in humans, development of tool-specific functional knowledge and skilled usage (for previously non-tool objects) occurs within the first training session and continues over successive sessions (Weisberg et al., 2007). However, at which stage an object becomes grounded as a tool remains an open question.

Clearly, sensorimotor knowledge is gained (albeit in a limited degree) in rarely used tools and one might presume that it *became* grounded. One could argue that the process of grounding the tool was "incomplete" for the rarely used tools, as action encoding and grounding seems to reflect left parietofrontal activation (Buxbaum and Kalenine, 2010). Though, if unfamiliar or very rarely used tools are not grounded and activate temporo-occipital regions, perhaps this is a site of initial storage of new/atypical tools before grounded in a sensorimotor experience. Here, a storage area is defined that represents an object that remains low level (perhaps visual) and shares some tool features, while a clear tool/action representation (left parietofrontal) has yet to be defined. This effect was seen in the study of Weisberg et al. (2007) on training the use of novel tools in actions. This study nicely demonstrated that before training, object–action matching for novel tools generated activity limited to occipital cortex followed by left temporo-parieto-frontal areas after training.

The spirit of this theory is seen in studies of recognition of visually presented objects. A study demonstrated that the occipitotemporal areas (lateral occipital–posterior fusiform) process recognized pictures with more activity than non-recognized images (Grill-Spector et al., 2000). Further, a more concise review demonstrates that such areas do not only respond to presence of stimuli, but to our perception of the stimuli (Grill-Spector and Malach, 2004). This garners some speculation that our perception (perhaps including familiarity, environment, etc…) may affect the meaningful processing of a tool or object (Lin et al., 2008). Similarly, activity within primary visual cortex can be modulated by crossmodal sensory processing during visual–auditory illusion (Watkins et al., 2006) and in the visual–tactile domain (Ramos-Estebanez et al., 2007),

suggesting a complex link between low-level visual structures and perception. Closer to the current topic, studies have shown that perceptual awareness of two objects is increased when the objects form a familiar tool–object action pair and are positioned, within a visual array, in a plausible way for action (Riddoch et al., 2003; Humphreys et al., 2006). More recently, Roberts and Humphreys (2010) observed specific activations of lateral occipital areas (e.g., early visual processing areas) for familiar objects positioned correctly for action as compared to those positioned incorrectly for action (Roberts and Humphreys, 2010). This object-orientation/object-position effect seems to be independent of visual attention (Riggio et al., 2008), and suggests that low-level visual processes directly contribute to our understanding of action-related object features.

Higher-level visual representations of tools and objects have been the subject of study for many years, and the neural mechanisms of tool identification and basic tool use are well known. Previous work has identified class-specific mechanisms of object recognition in the brain (Martin and Chao, 2001), and it is known that viewing tools activates inferior regions of the left intraparietal sulcus and ventral premotor cortex (Chao et al., 2002). Extensive tool-specific activation is also commonly seen in temporal regions (Beauchamp and Martin, 2007). Although commonly active in response to tool stimuli, there does appear to be some specialization in the type of information processed at these regions based on familiarity. Posterior parietal and premotor activation in response to tools may be specialized to convey information related to the motor affordance of a tool rather than its identity (Jeannerod et al., 1995; Johnson-Frey, 2004), while posterior and inferior temporal activation seems to be of particular importance in tool identification rather than understanding motoric qualities (Martin, 2007). Further, the anterior frontal regions, especially the ventrolateral aspect of the prefrontal cortex (Ranganath et al., 2004), are thought to form a network with regions of the temporal cortex (Mayes et al., 2007) to mediate the semantic and associative memory used in object representations (Martin, 2007). This concept is supported by lesion studies, which have shown that disrupted communication between the prefrontal and temporal regions impairs the recall of visual associative information (Tomita et al., 1999). As such, specialized mechanisms seem to link the identification of manipulable objects with information about the actions and context associated with their use, an idea that is well supported by previous literature. Grounding of action associated with a tool should engage extensive areas that affect perception. This infers that a strong test of tool grounding is more action based, and should be reflected in low and high level visual areas.

MODIFYING WHAT'S GROUNDED

In regards to the second question (modifying already grounded tools/objects/action), this speaks to a consideration that the parieto-temporo-frontal system related to tool action is very plastic. In this construct, new functions are attributed to existing tools or a task action is learned using a new set of tools. Such learning is typical of our daily lives, as rarely do we use a tool in only the capacity in which we have learned it. Routinely, tools are used to accomplish a task by the nature of basic aspects that the tool can afford. Here, if one needs to stir paint in absence of a paint stirrer, objects can be used that bear some similarity to the necessary tool. The issue of tools being grounded based on prior action history does not preclude one from using them apart from their intended purpose. Though, certain tools may be favored. The handle of a hammer could be used to stir paint, though this might make latter use of the hammer more difficult. An alternative may be a disposable pencil to accomplish the same task. The important fact is that neither the pencil nor the hammer was grounded for stirring a can of paint.

This issue bears similarities to ideas proposed in the study of retaining and recalling motor memories, based on our ability to store, recall, and modifying motor plans for varying environmental contexts. One theory, MOSAIC (modular selection and identification for control), takes into account our ability to operate in the world in multivariate environments (Wolpert and Kawato, 1998). Interestingly, this work introduces the idea of object use to promote how behavior depends on many internal and external variables. Hence, we are not scripted to perform movements only in the state that we learned the movement (e.g., only moving a full can from a neutral arm posture). In the MOSAIC model, modules of "controllers" for motor behavior account for some number of potential contexts. When the motor context is realized, the appropriate control is implemented. Thus, once context is realized, we can modify our movements based on existing modular states. Similarly, the vast array of tools and objects (and the possible interactions between them) creates a tremendous number of potential contexts that would have to be learned. Here, when a tool is needed but unavailable, other known tools may be imported to replace the needed one to afford action.

Herein, we propose a Modular Selection for Action Goals (MSAG) model. While the MOSAIC model is focused at the controller level and primarily describes the adaptive nature of sensorimotor control based on multiple pairs of forward and inverse models, we approach motor control from a more global level and propose a modular structure for the realization of action goals. In our model, an action goal is fixed (grounded), while the selection of an appropriate tool, context of use (of the tool), and neurobiomechanical parameters are contextually dependent on the action goal. A modular organization of tool/object action affords similar flexibility for rules in tool/object usage that allow deviation from established norms into new contexts, just as the MOSAIC model affords similar optimization of "situational" motor control. In this construct, once a tool is learned for a given task, that tool can be moved with relative ease in and out of possible contexts and become potentially "wired" to related/alternative tasks. In essence, a tool can be grounded to a particular sensorimotor and conceptual experience, while being embedded into a larger module accessible for many potential uses, which may or may not share similarities with past actions. Thus, we learn to use a knife to cut something *in particular*, then to cut *in general*, and, last, we extend the representation of a knife to a myriad of other potential uses (e.g., stir). This theory is represented in **Figure 1**. Three distinct modules are represented: a Tools Module, a Usage Contexts Module, and a Potential Neurobiomechanics Module. Each module provides unique information to a controller, which then derives a motor plan for the execution of behavior. Here, an array of *canonical and available* tools is identified for a given action, along with the expected neurobiomechanics. In this array, the spoon is available and selected, paired

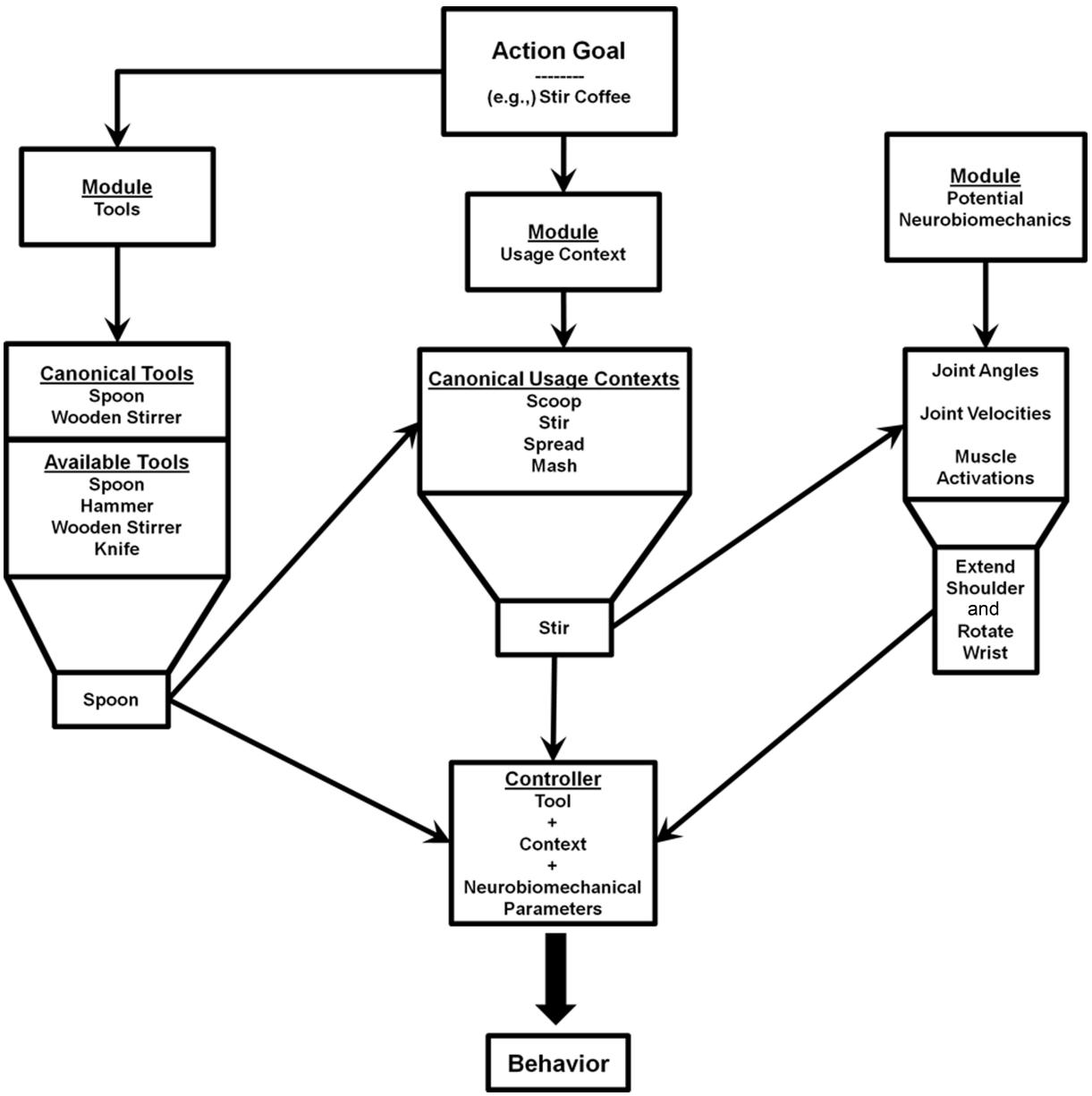


FIGURE 1 | Proposed MSAG model of tool, neurobiomechanical, and action representations for driving appropriate tool–object behaviors. In this example, canonical tools are available and spoon is selected for achieving the action goal of stirring coffee.

with the motoric demands of the action for successful accomplishment. As will be discussed later (Section "Relevance in Apraxia") the modular organization of MSAG helps to tentatively explain unique aspects of tool action impairment.

Existing theories of the structure of the encoding of tools/objects may afford this. The proposed "superordinate" representations of tool knowledge relate to how we categorize tools/objects etc. around themes. Superordinate classes of tools may reflect an organization that allows us to group things based on similarities, which may be functional (Murphy and Wisniewski, 1989), and has been used to address developmental theories (see next section). Behavioral evidence indicates that there is a potentiating of judgment when a current image bears similarities with a previous one (Myung et al., 2006). Similarly, behavioral evidence would also suggest that action can be afforded when presented with a superordinate category by way of activating a myriad of potential actions related to that category (Heit and Barsalou, 1996; Borghi and Caramelli, 2003; Marques, 2006). This last evidence may be important to the MSAG theory of defining novel tool uses (of familiar tools) for an action.

So, consider if a tool is desired but not available (e.g., **Figure 2**; no spoons or stirrers). Seeking the desired tool may, in effect, "prime" a supraordinate class of tools that are the best-fit plausible alternatives, if any are available. Here, eye glasses may not be primed, but

the knife would be (**Figure 2**) as it would be a part of a "utensil" supraordinate class. Once selected, the knife *may* become grounded for that new action upon use, while not impairing tool knowledge of the former action. Here, selection of knife is driven by functional similarity (e.g., used when eating) to the canonical (but unavailable) tool. Similarly, key to the MOSAIC model is that learning a new movement repertoire does not impede modules for other actions (Wolpert and Kawato, 1998). Further, interference from other potential usage contexts is avoided (slice, spread, chop), as they automatically do not meet the action goals (stir coffee). We highlight that at this stage, we are back to the first question, as actually grounding the "best-fit" alternative tool to a new action may require more than just one exposure. Under MSAG for tool selection and action, the "plasticity" of tools for many potential tasks is optimized.

It is important to emphasize that we are not fully regarding the MSAG model to tool–hand action alone (i.e., kinematic strategies for hammers of different weights). Much previous work has identified a potential role for the MOSAIC model in this capacity (Wolpert and Kawato, 1998; Haruno et al., 2001; Imamizu et al., 2003, 2007a,b; Imamizu and Kawato, 2009; Shah and Barto, 2009). This MSAG proposal extends our theory into conceptual tool use in variable situational states where a known tool takes on a "new" function. We will further advance theories related to MSAG into

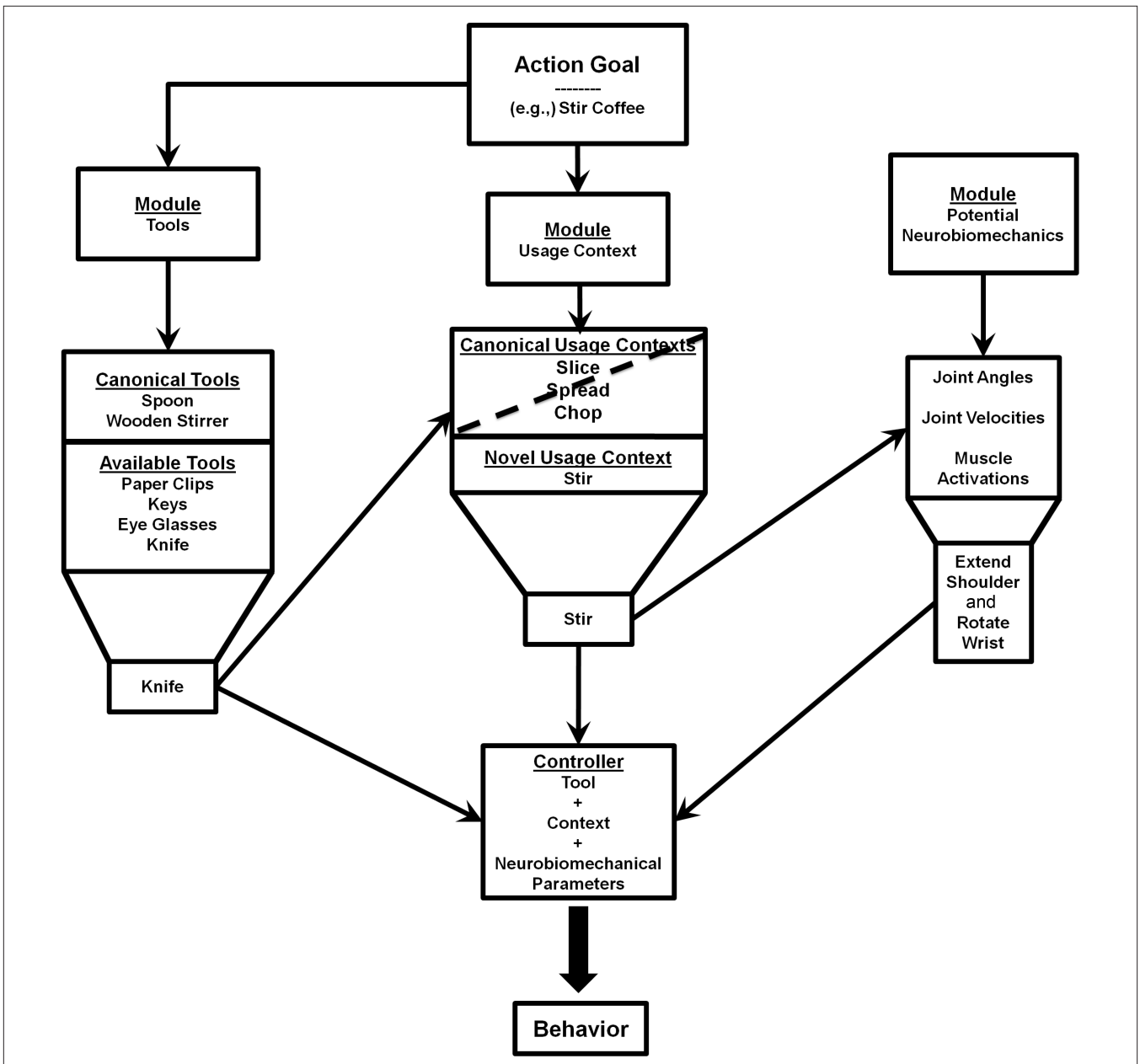


FIGURE 2 | Proposed MSAG model of tool, neurobiomechanical, and action representations for driving appropriate tool–object behaviors. In this example, canonical tools are not available and the novel usage context of stir is applied to a best-fit-tool (knife) to achieve the action goal of stirring coffee.

the neurological disorder of apraxia (see Section "Relevance in Apraxia"), offering a potential mechanistic description of deficits common to motor and conceptual apraxias.

IDEAS FROM DEVELOPMENTAL STUDIES

Adolescent behavioral neuroscience literature offers insight into the development of tool/object/action representations. Behavioral studies have demonstrated the development of functional associations between items in the formation of concepts related to manipulation knowledge (Kalenine and Bonthoux, 2008). Further, action priming is shown to enable object recognition (Mounoud et al., 2007). Particularly at a young age, knowledge of tool-based relationships is critical above more abstract, categorical relationship knowledge (Perraudin and Mounoud, 2009). Age plays a dramatic role in establishing how we categorize objects, where at very young ages (7 years old) object categorization is best at a basic level than at a superordinate level, which is no longer noticeable by age 9 (Kalenine et al., 2009a). Even infants display basic categorizations of visually presented objects at anterior–posterior brain areas (Grossmann et al., 2009). In keeping with this evidence, it is suggested that learning categories of tools and objects based on a mutual sense of action between the two is possible (Gershkoff-Stowe and Rakison, 2005; Smith, 2005) and may reflect similar, ongoing process in adulthood. Observational-based tool learning is robust in children (Hopper et al., 2010), though with age this capability likely introduces more motor error than necessary (McGuigan et al., 2010) and is suggestive of a need for sensorimotor experience for learning. Nevertheless, these developmental data represent a basis that the acquisition of tool concepts is heavily related to binding with objects or action. Thus, a spoon is a type of item that matches other spoons (belongs to a class of spoons based on structural characteristics), but is also associated with bowls (based on functional characteristics). Interestingly, knowledge that "spoon" is a member of the larger class of "utensil" requires more time. Evidence in adults suggests that such categorization knowledge of tool with a functional unit is represented, as expected, in pareito-temporal areas that have a high correspondence to action knowledge areas (Kalenine et al., 2009b). Thus, to learn new tool/object/action representations, or particularly in modifying existing ones, we heavily utilize pre-established, grounded concepts. The advantage of this concept would be in using existing templates that can be modified over time (see Section "Modifying What's Grounded"). To our knowledge this is still speculative, although further work considering developmental models may provide greater insight into the mechanisms of formation, storage, and modification of tool/object representations.

RELEVANCE IN APRAXIA

Of interest to the theory of grounded cognition (relevant to tool use) is apraxia. Apraxia is a deficit commonly arising after stroke that will impair tool-related behavior. This can include performance of tool movements, selection and ordering of tools for a task, and the general understanding of tools. Apraxia also extends to communicative gesture impairments, though we will not consider this feature in the current discussion. Here, we will focus on two forms of apraxia to highlight the implications of grounded cognition in explaining apraxia-related deficits, and we will offer an alternative theory based on the MSAG model.

We will consider two types of apraxia in this new model: ideomotor apraxia and conceptual apraxia. The first form, ideomotor apraxia, is commonly characterized by deficits in pantomime of tool use (i.e., demonstrating tool use without the actual tool). In this case, a patient is deficient at pantomiming tool usage out of context (e.g., "Show me how to use a hammer.") (Wheaton and Hallett, 2007). The second form, conceptual apraxia, is manifest as the inability to select tools adequate for a task. Thus, selecting tools in task-driven ways is impaired though it is possible to retain functional knowledge of the tool even when removed from natural settings (Heilman et al., 1997). Here, appropriate knowledge of tools and objects for a given task may become disrupted.

If we consider the nature of performing a praxis motor task, say driving a nail, there are many potential solutions to the goal. These solutions vary based on a number of factors, including the exact tool used (e.g., hammer vs. other possible implements), neurobiomechanical constraints to human motion (e.g., kinetics and kinematics, neuromuscular synergies, etc.), and contextual parameters (e.g., roofing vs. a more delicate task). The neurologically intact person is capable of deriving an appropriate set of motor commands from the combination of all of these factors. We first identify that a hammer is a "better" tool for driving a nail as compared to a shoe, for example. Then, we integrate the properties of the hammer with our sensorimotor system to develop and execute a motor plan, thus allowing us to drive the nail. Although seemingly automatic for many of us, patients with ideomotor and conceptual apraxia seem to show failures at different levels of this type of task.

In the case of ideomotor apraxia, there seems to be a disruption in translating the correctly identified tool (**Figure 1**, left side, many tools to one tool solution remains intact) into a single sensorimotor control parameters set (resulting in a one tool – appropriate sensorimotor parameters failure). In this case, the MSAG controller may be given the correct tool-related input, yet these parameters fail to converge at an appropriate, single set of sensorimotor tool-task control parameters. For example, a patient may be able to identify a fork as the most appropriate tool to eat a meal, but is unable to derive the neurobiomechanical control strategy to utilize the fork in the correct manner. In this framework, ideomotor apraxia is related to a disruption of the input or output of the Neurobiomechanical Module on the right side of **Figure 1**, while the Available Tools Module (left side of **Figure 1**) functions normally.

Conceptual apraxia, however, seems to be a disruption of the opposite phenomenon. There is a failure in deriving the many tools to one tool solution, yet the one tool – appropriate sensorimotor parameters solution is intact. Interestingly, here the MSAG controller may be given faulty tool-related information, yet appropriate tool-task sensorimotor control parameters are often realized which results in successful motoric (although contextually incorrect) use of a tool to achieve the desired behavior. For example, a patient may be able to successfully eat a meal using a toothbrush instead of a fork. In our proposed framework, conceptual apraxia is related to a disruption of the Available Tools Module (left side of **Figure 1**), while the Neurobiomechanical Module (right side of **Figure 1**) functions normally based on the action goals.

According to MSAG, the detriments in apraxia are based on separately damaged modules. Hence, a modular organization of MSAG is plausible. Thus, tool selection module fails in conceptual apraxia,

while the tool action module fails in ideomotor apraxia. At the same time, the tool action module survives in conceptual apraxia while the tool selection module survives in ideomotor apraxia. Lesion profiles for the two types of apraxia potentially support this notion of multiple modules, as primary damage to left parietofrontal regions is seen in ideomotor apraxia, while conceptual apraxia tends to result from bilateral damage to temporoparietal areas (Heilman and Gonzalez Rothi, 2003).

AGING

Aside from apraxia, evidence suggests specific processes related to tool knowledge may be affected by aging. It is important to note that these studies identify praxis impairments in a non-neuropathological aging population. Early work into the effect of healthy aging on praxis function revealed age-specific reductions in performance of pantomimes executed to verbal command. In healthy subjects, Ska and Nespoulous (1987) evaluated self-oriented (e.g., brushing the teeth) and external (e.g., tearing a piece of paper) pantomimes. One-handed repetitive and non-repetitive pantomimes were most affected, where older subjects often committed Body Part as Object (BPO) errors (e.g., instead of shaping the hand as if to use the absent object, the hand becomes the acting object).

More recently, Rodrigues Cavalcante and Caramelli (2009) evaluated gesture production to verbal command and imitation in healthy older subjects. As before, older subjects showed reduced performance of pantomime to verbal command, primarily due to commission of BPO errors. Further, Mozaz et al. (2009) evaluated arm–hand postural knowledge of tool use and communicative gestures in healthy older adults. Although knowledge related to both types of postures were reduced at a similar rate with increasing age, tool-related knowledge showed greater overall declines. This is in line with previous reports also suggesting dissociations between tool use and communicative gestures in apraxia (Bartolo et al., 2001; Villarreal et al., 2008; Bohlhalter et al., 2009). Mozaz et al. (2009) went further to suggest that loss of tool-use postural knowledge was unrelated to deficits in basic visual function, associative agnosia or semantic processing, but may be accounted for by the domain-specific representational hypothesis (Buxbaum and Kalenine, 2010). Evidence does not suggest that older subjects have "apraxia" *per se*, but that domain-specific processes are impaired in these subjects similar to stroke patients with apraxia, where tool-use knowledge is largely affected. Hence, aging may play a role in the grounded tool knowledge state. The mechanisms of what is changing are of importance and have a high relevance to understanding neural disease states common in advancing age. As well, it is clear that normal aging alters many neuroanatomical, cognitive and physiological brain processes (Lu et al., 2002; Davis et al., 2009; Hutton et al., 2009; Park and Reuter-Lorenz, 2009; Zahr et al., 2009) which may affect the neural architecture and function of the grounded cognition model as well. Adjustments under advancing age may represent another aspect of essential plasticity of tool knowledge states.

ARE WE FOR, OR AGAINST, GROUNDED COGNITION?

This section is of importance to correctly represent the above discussion. We do not feel that the grounded theory is invalid, but that there are challenging issues that must be addressed when considering whether and how something is "grounded." The main issue we hope to make clear is in the essentials of plasticity of the system. That is, how flexible are the functional representations of grounded tool/object/action to multiple variables (e.g., environment)? To make the point clear, we rely on our previous example of **Figure 2**, stirring coffee. If you need to prepare coffee but no stirrers or spoons are readily available, what do you do? In strict form, you may only have two alternatives (spoon or stirrer) with which to complete the action based on your sensorimotor history. Their absence would end in a "fault" and thus you would not be able to prepare your coffee. However, we know that other stirrer-like or spoon-like objects may be available which can be used in the place of a spoon or stirrer, and so we are able to advance and finish preparing our coffee. These alternatives may exist based on functional or structural similarity.

This adaptability presents a computational challenge to our neural architecture. How can we have nearly limitless adaptability of committing an act regardless of setting and the most basic tools at hand? While behaviorally robust, the neural representation of this is worth understanding. We propose a modular organization (MSAG) of tool/object ordinates as a method to accomplish the goal of importing and applying tools/objects in variable action states which afford for the plasticity of tool/object based decisions for action, and also utilize MOSAIC-like principles for tool-related motor behavior. Although seemingly oxymoronic at first, we propose this "plasticity" of grounded elements is an essential characteristic. Our neural systems for tool use must have the powerful ability to adapt, incorporating new acts and goals of various tools and objects into our motor repertoire. Such a system would strengthen our ability to map familiar tools to new actions, and afford importing new tools into the same neural architecture.

This proposal stems from the theory of *ad hoc* categorization (Barsalou, 1983) and is supported by recent literature in grounded and embodied cognition. It has been proposed that a "strong" and "weak" version of embodiment exists, with the latter described as a form of "graded" grounding (Chatterjee, 2010). Chatterjee (2010) goes further to suggest that "Referring to graded grounding invites consideration of continua and trade-offs between what is lost and what is gained. Representations by virtue of being less grounded in sensory and motor details lose some of their referential power. But, by virtue of being less grounded they also gain generative and flexible power." Other evidence specifically suggests that a distributed brain network abstracts action representations away from actors (Kable and Chatterjee, 2006), thereby diminishing sensory and motor details of specific actors and objects and allowing for relational plasticity. Our model is in agreement with this "graded grounding" conceptual approach, where tool–object relational processing may be removed from strict sensorimotor grounding and instead be governed by the context-dependent abstracted (e.g., *ad hoc*) category "Action Goal." Here, the original tool and object percepts remain intact, yet novel, context-driven relational interactions can form. In keeping with our knife-stir example, a knife is capable of adopting the function of a stirring implement with respect to a cup of coffee, yet the knife remains a knife (e.g., it is not perceived as a spoon).

POINTS FOR FURTHER REFINEMENT

Here we suggest a few thoughts to consider in support of the advancement of the theories that pertain to the storage of action information relevant to tools and objects:

1. Consideration of how new tools/objects are incorporated into our motor repertoire. Understanding of this "entry mechanism" will enable us to uniquely observe grounding taking place. Further, the time-course of this process can help to directly elaborate on the creation/maturation of grounded acts.
2. When might the action–observation network suggest that grounding has occurred? Based on the review of Grafton (2009) and considering the compelling evidence of action–observation network in matching a template of seen to understood action, this would suggest that this network would be a strong marker to detect grounding.
3. Address how tool/object relationships are weighted (i.e., is one tool/object more grounded than another?). Further understanding as to whether there is differential weighting of tools/objects to action may help to understand the development and structure of tools/objects to actions.
4. Computational models. Such models have been well used to drive improved theories of motor control, such as the MOSAIC model (Haruno et al., 2001) and our proposed MSAG model. These models have also been well argued for advancing theories of motoric aspects tool use (Krakauer and Shadmehr, 2007).

CONCLUSIONS AND SUMMARY

While an appealing notion, the mechanism of grounded cognition in tool use still remains an unclear. Of note, we demonstrate two main ideas that remain to be explored. Work on the mechanisms of grounding from a new tool/object and modifying grounded tools/objects to new actions perhaps sheds light on the mechanisms of this system. Consideration of computational and human research may suggest an alternative strategy, where the MSAG processes allow for behavioral optimization and (chiefly) operational flexibility without being computationally daunting. Fortunately, there are promising ideas to pursue that will help in better formulating ideas of how the system functions and offer insights to understand the neuroscience of this process from pediatric and adolescent development, adulthood, aging, and neural pathology.

REFERENCES

Anderson, S. J., Yamagishi, N., and Karavia V. (2002). Attentional processes link perception and action. *Proc. Biol. Sci.* 269, 1225–1232.

Barsalou, L. W. (1983). Ad hoc categories. *Mem. Cogn.* 11, 211–227.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609; discussion 610–560.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bartolo, A., Cubelli, R., Della Sala, S., Drei, S., and Marchetti, C. (2001). Double dissociation between meaningful and meaningless gesture reproduction in apraxia. *Cortex* 37, 696–699.

Beauchamp, M. S., and Martin, A. (2007). Grounding object concepts in perception and action: evidence from fMRI studies of tools. *Cortex* 43, 461–468.

Bohlhalter, S., Hattori, N., Wheaton, L., Fridman, E., Shamim, E. A., Garraux, G., and Hallett, M. (2009). Gesture subtype-dependent left lateralization of praxis planning: an event-related fMRI study. *Cereb. Cortex* 19, 1256–1262.

Borghi, A. M., and Caramelli, N. (2003). Situation bounded conceptual organization in children: from action to spatial relations. *Cogn. Dev.* 18, 49–60.

Buxbaum, L. J., and Kalenine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann. N. Y. Acad. Sci.* 1191, 201–218.

Buxbaum, L. J., Veramonti, T., and Schwartz, M. F. (2000). Function and manipulation tool knowledge in apraxia: knowing 'what for' but not 'how'. *Neurocase* 6, 83–97.

Chao, L. L., Weisberg, J., and Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cereb. Cortex* 12, 545–551.

Chatterjee A. (2010). Disembodying cognition. *Lang. Cogn.* 2, 79–116.

Davis, S. W., Dennis, N. A., Buchler, N. G., White, L. E., Madden, D. J., and Cabeza R. (2009). Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *Neuroimage* 46, 530–541.

Decety, J., and Grezes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain Res.* 1079, 4–14.

Evangeliou, M. N., Raos, V., Galletti, C., and Savaki, H. E. (2009). Functional imaging of the parietal cortex during action execution and observation. *Cereb. Cortex* 19, 624–639.

Frey, S. H., and Gerry, V. E. (2006). Modulation of neural activity during observational learning of actions and their sequential orders. *J. Neurosci.* 26, 13194–13201.

Gershkoff-Stowe, L., and Rakison, D. H. (eds) (2005). *Building Object Categories in Developmental Time*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N. Y. Acad. Sci.* 1156, 97–117.

Grezes, J., Tucker, M., Armony, J., Ellis, R., and Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *Eur. J. Neurosci.* 17, 2735–2740.

Grill-Spector, K., Kushnir, T., Hendler, T., and Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3, 837–843.

Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677.

Grossmann, T., Gliga, T., Johnson, M. H., and Mareschal, D. (2009). The neural basis of perceptual category learning in human infants. *J. Cogn. Neurosci.* 21, 2276–2286.

Halsband, U., and Lange, R. K. (2006). Motor learning in man: a review of functional and clinical studies. *J. Physiol. Paris* 99, 414–424.

Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., and Gazzaniga, M. S. (2003). Graspable objects grab attention when the potential for action is recognized. *Nat. Neurosci.* 6, 421–427.

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220.

Heilman, K. M., and Gonzalez Rothi, L. J. (2003). "Apraxia," in *Clinical Neurophysiology*, eds K. M. Heilman and E. Valenstein (New York: Oxford University Press), 215–235.

Heilman, K. M., Maher, L. M., Greenwald, M. L., and Rothi, L. J. (1997). Conceptual apraxia from lateralized lesions. *Neurology* 49, 457–464.

Heit, E., and Barsalou, L. W. (1996). The instantiation principle in natural categories. *Memory* 4, 413–451.

Hopper, L. M., Flynn, E. G., Wood, L. A. N., and Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *J. Exp. Child. Psychol.* 106, 82–97.

Humphreys, G. W., Riddoch, M. J., and Fortt, H. (2006). Action relations, semantic relations, and familiarity of spatial position in Balint's syndrome: crossover effects on perceptual report and on localization. *Cogn. Affect. Behav. Neurosci.* 6, 236–245.

Hutton, C., Draganski, B., Ashburner, J., and Weiskopf, N. (2009). A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage* 48, 371–380.

Imamizu, H., Higuchi, S., Toda, A., and Kawato, M. (2007a). Reorganization of brain activity for multiple internal models after short but intensive training. *Cortex* 43, 338–349.

Imamizu, H., Sugimoto, N., Osu, R., Tsutsui, K., Sugiyama, K., Wada, Y., and Kawato, M. (2007b). Explicit contextual information selectively contributes to predictive switching of internal models. *Exp. Brain Res.* 181, 395–408.

Imamizu, H., and Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. *Psychol. Res.* 73, 527–544.

Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T., and Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5461–5466.

Jarvelainen, J., Schurmann, M., and Hari, R. (2004). Activation of the human primary motor cortex during observation of tool use. *Neuroimage* 23, 187–192.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* 18, 314–320.

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends Cogn. Sci.* 8, 71–78.

Kable, J. W., and Chatterjee, A. (2006). Specificity of action representations in the lateral occipitotemporal cortex. *J. Cogn. Neurosci.* 18, 1498–1517.

Kalenine, S., and Bonthoux, F. (2008). Object manipulability affects children's and adults' conceptual processing. *Psychon. Bull. Rev.* 15, 667–672.

Kalenine, S., Bonthoux, F., and Borghi, A. M. (2009a). How action and context priming influence categorization: a developmental study. *Br. J. Dev. Psychol.* 27, 717–730.

Kalenine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., and Baciu, M. (2009b). The sensory-motor specificity of taxonomic and thematic conceptual relations: a behavioral and fMRI study. *Neuroimage* 44, 1152–1162.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., and Ungerleider, L. G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proc. Natl. Acad. Sci. U.S.A.* 95, 861–868.

Karni, A., and Sagi, D. (1993). The time course of learning a visual skill. *Nature* 365, 250–252.

Kellenbach, M. L., Brett, M., and Patterson, K. (2003). Actions speak louder than functions: the importance of manipulability and action in tool representation. *J. Cogn. Neurosci.* 15, 30–46.

Krakauer, J. W., and Shadmehr, R. (2007). Towards a computational neuropsychology of action. *Prog. Brain Res.* 165, 383–394.

Lewis, J. W. (2006). Cortical networks related to human use of tools. *Neuroscientist* 12, 211–231.

Lin, Z., Lin, Y., and Han, S. (2008). Self-construal priming modulates visual activity underlying global/local perception. *Biol. Psychol.* 77, 93–97.

Lu, L. H., Crosson, B., Nadeau, S. E., Heilman, K. M., Gonzalez-Rothi, L. J., Raymer, A., Gilmore, R. L., Bauer, R. M., and Roper, S. N. (2002). Category-specific naming deficits for objects and actions: semantic attribute and grammatical role hypotheses. *Neuropsychologia* 40, 1608–1621.

Marques, J. F. (2006). Specialization and semantic organization: evidence for multiple semantics linked to sensory modalities. *Mem. Cogn.* 34, 60–67.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201.

Mayes, A., Montaldi, D., and Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends Cogn. Sci.* 11, 126–135.

McGuigan, N., Makinson, J., and Whiten, A. (2010). From over-imitation to super-copying: adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *Br. J. Psychol.* doi: 10.1348/000712610X493115

Menz, M. M., Blangero, A., Kunze, D., and Binkofski, F. (2010). Got it! Understanding the concept of a tool. *Neuroimage* 51, 1438–1444.

Mounoud, P., Duscherer, K., Moy, G., and Perraudin, S. (2007). The influence of action perception on object recognition: a developmental study. *Dev. Sci.* 10, 836–852.

Mozaz, M. J., Crucian, G. P., and Heilman, K. M. (2009). Age-related changes in arm–hand postural knowledge. *Cogn. Neuropsychol.* 26, 675–684.

Murphy, G. L., and Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *J. Exp. Psychol. Learn Mem. Cogn.* 15, 572–586.

Myung, J. Y., Blumstein, S. E., and Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition* 98, 223–243.

Park, D. C., and Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annu. Rev. Psychol.* 60, 173–196.

Perraudin, S., and Mounoud, P. (2009). Contribution of the priming paradigm to the understanding of the conceptual developmental shift from 5 to 9 years of age. *Dev. Sci.* 12, 956–977.

Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., and Castiello, U. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cereb. Cortex* 19, 367–374.

Ramos-Estebanez, C., Merabet, L. B., Machii, K., Fregni, F., Thut, G., Wagner, T. A., Romei, V., Amedi, A., and Pascual-Leone, A. (2007). Visual phosphene perception modulated by subthreshold crossmodal sensory stimulation. *J. Neurosci.* 27, 4178–4181.

Ranganath, C., Cohen, M. X., Dam, C., and D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J. Neurosci.* 24, 3917–3925.

Rice, N. J., Valyear, K. F., Goodale, M. A., Milner, A. D., and Culham, J. C. (2007). Orientation sensitivity to graspable objects: an fMRI adaptation study. *Neuroimage* 36(Suppl. 2), T87–T93.

Riddoch, M. J., Humphreys, G. W., Edwards, S., Baker, T., and Willson, K. (2003). Seeing the action: neuropsychological evidence for action-based effects on object selection. *Nat. Neurosci.* 6, 82–89.

Riggio, L., Iani, C., Gherri, E., Benatti, F., Rubichi, S., and Nicoletti, R. (2008). The role of attention in the occurrence of the affordance effect. *Acta Psychol. (Amst)* 127, 449–458.

Roberts, K. L., and Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *Neuroimage*. 52, 1541–1548.

Rodrigues Cavalcante, K., and Caramelli, P. (2009). Evaluation of the performance of normal elderly in a limb praxis protocol: influence of age, gender, and education. *J. Int. Neuropsychol. Soc.* 15, 618–622.

Shah, A., and Barto, A. G. (2009). Effect on movement selection of an evolving sensory representation: a multiple controller model of skill acquisition. *Brain Res.* 1299, 55–73.

Ska, B., and Nespoulous, J. L. (1987). Pantomimes and aging. *J. Clin. Exp. Neuropsychol.* 9, 754–766.

Smith, L. B. (2005). Action alters shape categories. *Cogn. Sci.* 29, 665–679.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., and Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401, 699–703.

Tucker, M., and Ellis, R. (1998). On the relations between seen objects and components of potential actions. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 830–846.

Tucker, M., and Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychol. (Amst)* 116, 185–203.

Valyear, K. F., and Culham, J. C. (2010). Observing learned object-specific functional grasps preferentially activates the ventral stream. *J. Cogn. Neurosci.* 22, 970–984.

Villarreal, M., Fridman, E. A., Amengual, A., Falasco, G., Gerscovich, E. R., Ulloa, E. R., and Leiguarda, R. C. (2008). The neural substrate of gesture recognition. *Neuropsychologia* 46, 2371–2382.

Vingerhoets, G. (2008). Knowing about tools: neural correlates of tool familiarity and experience. *Neuroimage* 40, 1380–1391.

Vingerhoets, G., Acke, F., Vandemaele, P., and Achten, E. (2009). Tool responsive regions in the posterior parietal cortex: effect of differences in motor goal and target object during imagined transitive movements. *Neuroimage.* 47, 1832–1843.

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., and Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31, 1247–1256.

Weisberg, J., van Turennout, M., and Martin, A. (2007). A neural system for learning about object function. *Cereb. Cortex* 17, 513–521.

Wheaton, L. A., and Hallett, M. (2007). Ideomotor apraxia: a review. *J. Neurol. Sci.* 260, 1–10.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329.

Zahr, N. M., Rohlfing, T., Pfefferbaum, A., and Sullivan, E. V. (2009). Problem solving, 5orking memory, and motor correlates of association and commissural fiber bundles in normal aging: a quantitative fiber tracking study. *Neuroimage* 44, 1050–1062.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 July 2010; accepted: 21 October 2010; published online: 15 November 2010.

Citation: Mizelle JC and Wheaton LA (2010) The neuroscience of storing and molding tool action concepts: how "plastic" is grounded cognition? Front. Psychology 1:195. doi: 10.3389/fpsyg.2010.00195

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.

Copyright © 2010 Mizelle and Wheaton. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.