HOW HUMANS RECOGNIZE OBJECTS: SEGMENTATION, CATEGORIZATION AND INDIVIDUAL IDENTIFICATION

EDITED BY: Chris Fields
PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

frontiers

Frontiers Copyright Statement

© Copyright 2007-2016 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, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence 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.

ISSN 1664-8714
ISBN 978-2-88919-940-2
DOI 10.3389/978-2-88919-940-2

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

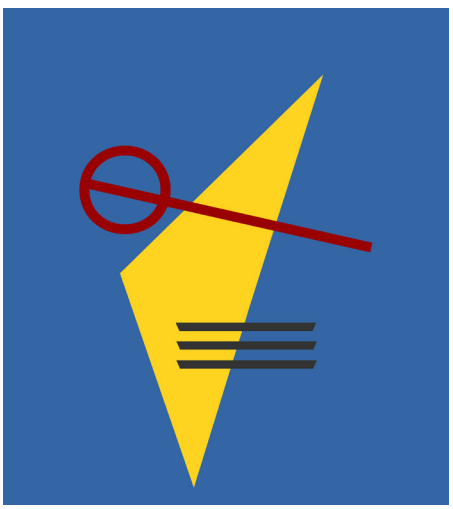
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - 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**

HOW HUMANS RECOGNIZE OBJECTS: SEGMENTATION, CATEGORIZATION AND INDIVIDUAL IDENTIFICATION

Topic Editor:
Chris Fields, New Mexico State University, USA (retired)



Assemblage: Bronwyn Sleeping, C. Fields, 2003

Human beings experience a world of objects: bounded entities that occupy space and persist through time. Our actions are directed toward objects, and our language describes objects. We categorize objects into kinds that have different typical properties and behaviors. We regard some kinds of objects – each other, for example – as animate agents capable of independent experience and action, while we regard other kinds of objects as inert. We re-identify objects, immediately and without conscious deliberation, after days or even years of non-observation, and often following changes in the features, locations, or contexts of the objects being re-identified.

Comparative, developmental and adult observations using a variety of approaches and methods have yielded a detailed understanding of object detection and recognition by the visual system and an advancing understanding of haptic and auditory information processing. Many fundamental questions, however, remain unanswered. What, for example, physically constitutes an "object"? How do specific, classically-characterizable object boundaries emerge from the physical dynamics described by quantum theory, and can this emergence process be described independently of any assumptions regarding the perceptual capabilities of observers? How are visual motion and feature information combined to create object information? How are the object trajectories that indicate persistence to human observers implemented, and how are these trajectory representations bound to feature representations? How, for example, are point-light walkers recognized as single objects? How are conflicts between trajectory-driven and feature-driven identifications of objects resolved, for example in multiple-object tracking situations? Are there separate "what" and "where" processing streams for haptic and auditory perception? Are there haptic and/or auditory equivalents of the visual object file? Are there equivalents of the visual object token? How are object-identification conflicts between different perceptual systems resolved?

Is the common assumption that "persistent object" is a fundamental innate category justified? How does the ability to identify and categorize objects relate to the ability to name and describe them using language? How are features that an individual object had in the past but does not have currently represented? How are categorical constraints on how objects move or act represented, and how do such constraints influence categorization and the re-identification of individuals? How do human beings re-identify objects, including each other, as persistent individuals across changes in location, context and features, even after gaps in observation lasting months or years? How do human capabilities for object categorization and re-identification over time relate to those of other species, and how do human infants develop these capabilities? What can modeling approaches such as cognitive robotics tell us about the answers to these questions?

Primary research reports, reviews, and hypothesis and theory papers addressing questions relevant to the understanding of perceptual object segmentation, categorization and individual identification at any scale and from any experimental or modeling perspective are solicited for this Research Topic. Papers that review particular sets of issues from multiple disciplinary perspectives or that advance integrative hypotheses or models that take data from multiple experimental approaches into account are especially encouraged.

Citation: Fields, C., ed. (2016). How Humans Recognize Objects: Segmentation, Categorization and Individual Identification. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-940-2

Table of Contents

06 ***Editorial: How Humans Recognize Objects: Segmentation, Categorization and Individual Identification***
Chris Fields

08 ***The development of contour processing: evidence from physiology and psychophysics***
Gemma Taylor, Daniel Hipp, Alecia Moser, Kelly Dickerson and Peter Gerhardstein

18 ***Low-level and high-level modulations of fixational saccades and high frequency oscillatory brain activity in a visual object classification task***
Maciej Kosilo, Sophie M. Wuerger, Matt Craddock, Ben J. Jennings, Amelia R. Hunt and Jasna Martinovic

35 ***Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions***
Haline E. Schendan and Giorgio Ganis

69 ***Affective and contextual values modulate spatial frequency use in object recognition***
Laurent Caplette, Gregory West, Marie Gomot, Frédéric Gosselin and Bruno Wicker

78 ***The neural mechanisms for the recognition of face identity in humans***
Stefano Anzellotti and Alfonso Caramazza

84 ***The transition in the ventral stream from feature to real-world entity representations***
Guy A. Orban, Qi Zhu and Wim Vanduffel

93 ***Visuo-haptic multisensory object recognition, categorization, and representation***
Simon Lacey and K. Sathian

108 ***Short-term plasticity of visuo-haptic object recognition***
Tanja Kassuba, Corinna Klinge, Cordula Hölig, Brigitte Röder and Hartwig R. Siebner

124 ***Cortical processing of object affordances for self and others' action***
Monica Maranesi, Luca Bonini and Leonardo Fogassi

134 ***Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI***
Ricarda I. Schubotz, Moritz F. Wurm, Marco K. Wittmann and D. Yves von Cramon

149 ***Prediction-learning in infants as a mechanism for gaze control during object exploration***
Matthew Schlesinger, Scott P. Johnson and Dima Amso

161 ***Binocular fusion and invariant category learning due to predictive remapping during scanning of a depthful scene with eye movements***
Stephen Grossberg, Karthik Srinivasan and Arash Yazdanbakhsh

202 ***Perceptions of document relevance***
Peter Bruza and Vivien Chang

210 ***Quantum theory and human perception of the macro-world***
Diederik Aerts

229 ***Sameness and the self: philosophical and psychological considerations***
Stanley B. Klein

244 ***Objects of consciousness***
Donald D. Hoffman and Chetan Prakash

frontiers in Psychology

EDITORIAL
published: 22 March 2016
doi: 10.3389/fpsyg.2016.00400

Editorial: How Humans Recognize Objects: Segmentation, Categorization and Individual Identification

Chris Fields *

New Mexico State University, Las Cruces, NM, USA (Retired)

Keywords: space, time, object persistence, individual identity

The Editorial on the Research Topic

How Humans Recognize Objects: Segmentation, Categorization and Individual Identification

What does it mean to say that something is an *object*? How do we recognize objects as such, picking them out from any non-objects that might happen to be present? What, indeed, does it mean to say that something is *not* an object? Is it even possible to recognize a non-object?

What, moreover, does it mean to say that something is a specific, *individual* object. Suppose you are handed 10 brand-new 1 € coins, each of which looks and feels exactly like the others. How do we recognize one of them as exactly the same individual 1 € coin we were looking at a moment ago? How does this process change if we've looked away for a few seconds, a minute, an hour? What if we have not seen the coin since last year? How does the individual recognition process change if, instead of coins, we are talking about 10 new colleagues encountered at a meeting 1 year ago?

The "what does it mean" versions of these questions have been with us since antiquity, in the form of philosophical musings about the nature of or evidence for an external world. The "how" versions have been asked for slightly over a century, and a detailed picture has begun to emerge only in the past two decades. Schneider's (1969) suggestion that two distinct pathways support visual orientation toward object features and locations was a watershed event in this growing understanding (see Goodale and Milner, 1992 for an early review). Research stemming from this idea has inextricably linked object recognition to the experiences of space, time, and persistence over time, i.e., individual identity (see Scholl, 2007; Fields, 2012 for review). Without a spacetime "container" and individual, time-persistent objects, motion and causation cannot be defined; hence object recognition underlies these experiences as well.

The papers in this Research Topic provide a glimpse of the current state of understanding the "how" of object recognition. Beginning with the most concrete, Taylor et al. review the development of contour detection and integration in humans, relating the functional trajectory from infancy to adolescence to the increasing range of horizontal connectivity within areas V1 and V2 during the same period. Kosilo et al. then describe new experiments designed to tease apart the effects of low-level (color and contrast) and high-level (identifiability as an object) stimulus features on the control of visual saccades. Schendan and Ganis show that object recognition exerts top-down effects on visual processing within 250 ms; Caplette et al. demonstrate the influence of top-down affective and contextual expectations on the precision with which objects are represented. Anzellotti and Caramazza review evidence suggesting that human face identity is selectively encoded in the right-hemisphere anterior temporal pole (ATP), an area generally implicated in semantic memory. Orban et al. review the functional anatomy of the ventral stream, and suggest that fully-defined individual entities of all types are represented in ATP.

OPEN ACCESS

Edited and reviewed by:
Rufin VanRullen, Centre de Recherche Cerveau et Cognition, France

***Correspondence:**
Chris Fields
fieldsres@gmail.com

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *02 March 2016*
Accepted: *06 March 2016*
Published: *22 March 2016*

Citation:
Fields C (2016) Editorial: How Humans Recognize Objects: Segmentation, Categorization and Individual Identification. Front. Psychol. 7:400. doi: 10.3389/fpsyg.2016.00400

Lacey and Sathian review visuo-haptic integration, focusing on the role of lateral occipital cortex (LOC); Kassuba et al. describe downstream effects on visual and haptic processing following disruption of LOC activity by transcranial magnetic stimulation. Maranesi et al. review the representation of motor affordances and their activation by object recognition, while Schubotz et al. present new results on the representation of action expectations. Schlesinger et al. address the key question of how infants learn to generate expectations that predict the behavior of the visual world.

The remaining five papers address fundamental theoretical issues. Grossberg et al. address the question of scene stability across eye movements using the Adaptive Resonance Theory framework. Bruza and Chang investigate the utility of quantum probabilities for explaining relevance judgments. Aerts reviews quantum theory itself, explaining why it renders the existence of the separate, bounded entities that we call "objects" mysterious. Klein examines the human perception of a time-persistence self and suggests that sameness is a pre-evidential "default mode" of the self representation. Hoffman and Prakash review evidence suggesting that neither objects nor their spacetime "container" objectively exist, but must instead be considered to be emergent from multi-agent interactions.

Beyond the leading edge represented by these papers lie questions for further research, many of which concern the development, especially during early infancy, of object-recognition capabilities. Three of the most significant, in my opinion, are the following.

1. How malleable are the human representations of space and time? Are particular motor capabilities essential to the development of these representations? What is the role of sensory-motor correlations in representing perceived space? Would an organism inhabiting a world devoid of manipulable objects be able to develop a 3d spatial representation?
   Recent developments in quantum theory have led to a new emphasis among physicists on reference frames as physical objects, not just abstract coordinate systems, with respect to which quantities are measured: examples include clocks and gyroscopes used as reference frames to measure time and spatial orientation, respectively (Bartlett et al., 2007). What are the earliest-developing reference frames for space and time in humans? By what age do infants perceive objects as embedded in a containing space that imposes relationships upon them, as opposed to just perceiving objects?
2. How do causal reasoning and object recognition ability co-develop? Is there some particular level of predictability that is required? What kind of predictability—predictable locations or motions, predictable static features, or both? What would happen in an environment in which the predictability of locations and motions was uncorrelated with the predictability of static features?
   Any object that serves as a reference frame must be unproblematically recognizable as such: a clock, for example, can only serve as a clock if its identity over time is not in question. What level of predictability must the infant environment have in order for typical space and time reference frames to develop? What level of predictability must it have in order for typical object categories to develop? What happens in environments with less than this critical level of predictability?
3. How does the subjectively-accessible sense of the body as a time-persistent object and hence of the stably-embodied self develop? Rochat (2012) suggests that a rudimentary embodied-self representation is present at birth. How is this representation implemented? How is this implementation constructed prenatally?
   If Hoffman and Prakash are right in stating that a shared external world of objectively-defined objects cannot be assumed, the infant's representation of itself and its capabilities for action becomes the only reference frame from which a perceived world of persistent objects can be constructed. What level of coherence must the world provide, whatever its structure, for this process of construction to be feasible?

These questions cannot, clearly, be fully answered by experiments with human infants. Combining experiments that are feasible with infants with experiments carried out on validated computational models, as in the work of Schlesinger et al. promises to become even more important as questions such as those contemplated here are addressed.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

REFERENCES

Bartlett, S. D., Rudolph, T., and Spekkens, R. W. (2007). Reference frames, superselection rules, and quantum information. *Rev. Mod. Phys.* 79, 555–609. doi: 10.1103/RevModPhys.79.555

Fields, C. (2012). The very same thing: extending the object token concept to incorporate causal constraints on individual identity. *Adv. Cogn. Psychol.* 8, 234–247. doi: 10.2478/v10053-008-0119-8

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25.

Rochat, P. (2012). "Primordial sense of embodied self-unity," in *Early Development of Body Representations*, eds V. Slaughter and C. A. Brownell (Cambridge: Cambridge University Press), 3–18.

Schneider, G. E. (1969). Two visual systems. *Science* 163, 895–902.

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind Lang.* 22, 563–591. doi: 10.1111/j.1468-0017.2007.00321.x

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.

Copyright © 2016 Fields. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 08 July 2014
doi: 10.3389/fpsyg.2014.00719

The development of contour processing: evidence from physiology and psychophysics

Gemma Taylor[1], Daniel Hipp[1], Alecia Moser[1], Kelly Dickerson[2] and Peter Gerhardstein[1]*

[1] *Department of Psychology, Binghamton University, State University of New York, Binghamton, NY, USA*
[2] *US Army Research Laboratory, Department of the Army, RDRL-HRS-D, Aberdeen Proving Grounds, MD, USA*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Chris Fields, New Mexico State University, USA (retired)
Bat-Sheva Hadad, University of Haifa, Israel

****Correspondence:***
Peter Gerhardstein, Department of Psychology, Binghamton University, State University of New York, Binghamton, NY 13902-6000, USA
e-mail: gerhard@binghamton.edu

Object perception and pattern vision depend fundamentally upon the extraction of contours from the visual environment. In adulthood, contour or edge-level processing is supported by the Gestalt heuristics of proximity, collinearity, and closure. Less is known, however, about the developmental trajectory of contour detection and contour integration. Within the physiology of the visual system, long-range horizontal connections in V1 and V2 are the likely candidates for implementing these heuristics. While post-mortem anatomical studies of human infants suggest that horizontal interconnections reach maturity by the second year of life, psychophysical research with infants and children suggests a considerably more protracted development. In the present review, data from infancy to adulthood will be discussed in order to track the development of contour detection and integration. The goal of this review is thus to integrate the development of contour detection and integration with research regarding the development of underlying neural circuitry. We conclude that the ontogeny of this system is best characterized as a developmentally extended period of associative acquisition whereby horizontal connectivity becomes functional over longer and longer distances, thus becoming able to effectively integrate over greater spans of visual space.

Keywords: contour detection, closure, horizontal connections, development, visual development

INTRODUCTION

The early visual system is one of the first avenues by which infants begin to learn about the world around them (Piper and Darrah, 1994). Visual capabilities begin developing before birth (Alberts, 1984), undergo considerable maturation in the first few months after birth (Johnson, 2001; Lewis and Maurer, 2005; Atkinson and Braddick, 2007), and continue into adolescence (see Slater and Johnson, 1998; Pennefather et al., 1999; Hadad et al., 2010b). Visual development has been characterized with varying degrees of specificity across several domains, including: sensitivity to spatial frequency (Patel et al., 2010), orientation (Braddick et al., 1986; Morrone and Burr, 1986; Candy et al., 2001), motion (Johnson, 2001; Wattam-Bell et al., 2010), color perception (Bornstein et al., 1976; Gerhardstein et al., 1998), and facial recognition (Bushnell et al., 1989; Johnson et al., 1992; for a recent review, see Braddick and Atkinson, 2011). However, many descriptions of the mechanisms through which infants begin to make sense of their visual world and how these mechanisms might change across ontogeny are somewhat sparse.

The goal of the present paper is to review, discuss, and integrate findings from across infancy and childhood in order to shed light on the development of contour detection and integration from first emergence to adult-level function. Throughout this review, psychophysical data will be augmented by data from physiological and theoretical studies, and adult data will be used to inform the examination of the developmental path where possible. We will focus on how the visual pathway implements initial contour processing across development. Therefore, we will not discuss the role of top-down processing in modulating object perception in depth, as that topic is beyond the scope of this review. We conclude with a discussion of how to interpret what appears to be quite protracted unfolding of this system, and with a call to action for further research in areas where data is lacking.

PATH TO OBJECT PERCEPTION

Construction of a clear and meaningful percept of a visual scene is a demanding computational problem. Developing basic acuity in infancy and orientation sensitivity (Banks and Salapatek, 1981; Morrone and Burr, 1986; Sireteanu et al., 1994; Candy et al., 2001) is an important first step toward the development of pattern and object perception in the visual world (see also Wattam-Bell et al., 2010). Detecting regions within the visual field that contain points of locally high contrast and then integrating these early representations into a contour-level description of the scene (e.g., Marr, 1982) can then be used to infer object edges, surfaces, and depth boundaries (Peterson, 2001). Although a number of theoretical models for object perception have been proposed (e.g., Marr, 1982; Biederman, 1987; Dickinson et al., 1992; Ullman, 2007), the ontogeny of object perception is still not well understood (e.g., Kovács et al., 1999; Hou et al., 2003; Gerhardstein et al., 2004; Hadad et al., 2010a).

Gestalt theorists have proposed that proximity (elements that are close together tend to be grouped together), collinearity or good continuation (elements that are aligned with one another will

be grouped into the same contour), common fate (elements that move along the same path likely belong to the same contour), and closure (a closed contour is easier to detect than an open one) are processing heuristics for contour detection and contour integration (Köhler, 1947; Wertheimer, 1958). Within the adult literature, a substantial body of research describing contour perception suggests that contour or edge-level processing reflects the heuristics of proximity, collinearity, common fate and closure (for a review, see Wagemans et al., 2012).

Importantly, the low-level characteristics of natural scenes in the visual world have been shown to be statistically regular; this regularity has been taken as support for the suggestion that Gestalt heuristics may be used for contour detection. Geisler (2008), Geisler and Perry (2009), and Geisler et al. (2001) in particular noted that contours in natural scenes are relatively smooth and therefore heuristics such as proximity and collinearity have a statistical basis in natural scenes. This regularity scaffolds numerous aspects of visual perception including the use of proximity information (Brunswick and Kamiya, 1953), proximity interacting with curvature/collinearity (Geisler et al., 2001; Tversky et al., 2004; Lawlor and Zucker, 2013), figure-ground segmentation (Fowlkes et al., 2007) and closure (for reviews see Kovács, 1996; Pettet et al., 1998; Mathes and Fahle, 2007; Geisler, 2008; Loffler, 2008; Geisler and Perry, 2009). Gestalt heuristics therefore take advantage of this natural order. How the mature observer acquires the mechanisms underlying these heuristics, however, is unclear. In nature, proximity and collinearity are highly correlated (Geisler et al., 2001) even in natural scenes in which partial occlusions are frequent (although contrast polarity also plays a role in contour detection in such instances; see Geisler and Perry, 2009).

CONTOUR PROCESSING – ELEMENTAL DETECTION TO INTEGRATION, IN BRAIN AND BEHAVIOR

The integration of spatially disparate but organizationally related visual information is a fundamental component of object perception, and has been highlighted in the adult psychophysics literature (Field et al., 1993; Kovács and Julesz, 1993; Mathes and Fahle, 2007; for review, see Loffler, 2008), the neurophysiological literature (Nelson and Frost, 1985; Ts'o et al., 1986; Gilbert and Wiesel, 1989; Gilbert et al., 1996; Bosking et al., 1997; Li, 1998; Stettler et al., 2002; Cass and Spehar, 2005), and in modeling work (Yen and Finkel, 1998; Grossberg and Williamson, 2001; Voges et al., 2010; Gintautas et al., 2011; Piëch et al., 2013). Following detection of contour segments, integrating these segments into a larger whole, or contour, is generally seen as the next step toward detecting individual objects. While much work has been done on object perception (Johnson, 2001), the present review focuses on low- and intermediate-level studies regarding contour processing to determine the relation between physiology and perceptual capabilities in this domain across development. The next section discusses the lowest level of spatial integration – collinear facilitation in flanker tasks – in terms of physiology and perception. Our discussion then extends up the visual hierarchy, to similarly elucidate larger-scale visuo-spatial integration underpinning higher-order contour processing. Again, this relationship is examined in terms of research from both the psychophysical and physiological perspectives. At its terminus, this section relates the discussed work to higher-level object perception across development.

Physiology for elemental detection and integration

The rudiments of object perception begins when light from the visual scene falls on the photoreceptors in the retina. Each photoreceptor detects light from a small fraction of the visual scene. From the photoreceptors, information is sent via ganglion cells to the lateral geniculate nucleus (LGN) and then to area V1 (followed by V2, V3, V4, and V5 via feedforward and feedback connections) in the primary visual cortex. Neurons in area V1 are dedicated to the detection of segments of specific orientations and spatial frequencies (among other visual attributes, Hubel and Wiesel, 1959, 1968; Hubel et al., 1977), referred to as the neuron's *classical receptive field* (CRF). However, more recent work has shown that neurons in area V1 are also influenced by input from areas outside the CRF. Specifically, detection of a foveated Gabor target (a Gaussian-modulated sinusoidal luminance distribution) is influenced by proximity and collinearity of the flanking elements in a flanker facilitation task (Polat and Sagi, 1993; Shani and Sagi, 2005; Lev and Polat, 2011). When flankers were presented in the 2–6λ range (where λ equals the wavelength of the Gabor itself) and were collinear with the target element, a flanker facilitation effect occurred, reducing the detection threshold for the target element (Polat and Sagi, 1993). This *contextual modulation* of neurons in area V1 can be explained by excitatory and inhibitory long-range horizontal interconnections between neurons. Early reports of the existence of these connections (Rockland and Lund, 1982; Gilbert and Wiesel, 1983, 1989; Nelson and Frost, 1985; Ts'o et al., 1986) have been clearly confirmed (Gilbert et al., 1996; Bosking et al., 1997; Kapadia et al., 2000; Stettler et al., 2002; Gilad et al., 2012). Research suggests that the horizontal connections in the visual cortex underlie at least some Gestalt processes (Field et al., 1993; Kovács and Julesz, 1993; Tversky et al., 2004; Mathes and Fahle, 2007; for review, see Loffler, 2008). Information detected by neurons in area V1 must, however, be integrated into more global-level contours that can be used to detect objects and subsequently, form a meaningful percept of the visual scene.

Two complimentary, but computationally quite opposite processes appear to occur via these connections in V1 (and perhaps in V2; Polat, 2013). The first is a process of object boundary detection supported by iso-orientation inhibition, whereby cortical columns sensitive to a particular orientation inhibit nearby regions sharing orientation information. This inhibition occurs less at object edges than inside or outside these boundaries, making the enclosing regions of the visual field that denote objects explicit and salient. This process appears to occur early, and does not appear to require top-down input to operate, functioning instead as part of an initial bottom-up process. The second is a process of attention-mediated region-filling, whereby regions sharing orientation information propagate an excitatory signal that fills in textures and stops at object boundaries (similar to classical grassfire algorithms, e.g., Blum, 1967; Kovács et al., 1998). This process appears to occur following the boundary detection process, and indeed may depend on it, as the boundaries discovered in the first process designate for the second process which

regions of the visual field need filling in. Anatomically, superficial layers in V1 columns receive feed-forward inputs and perform pre-attentive boundary detection, whereas region filling appears to be triggered in deeper layers (layers IV and V) as a function of top-down attentional feedback from higher layers (Polat, 2013).

The physiology supporting a mechanism for contour detection appears to be present early in infancy, at least in a rudimentary form (see Burkhalter et al., 1993; Kovács et al., 1999; Gerhardstein et al., 2004; Hadad et al., 2010a). Using human brains ranging from 24 weeks gestational age to those of children up to 5 years of age, Burkhalter et al. (1993) documented that the basic structures of V1 in the primary visual cortex are in place early in life. However, the vertical connections between layers and horizontal connections within layers of the visual cortex show protracted development. Specifically, Burkhalter et al. (1993) describe a dense network of horizontal connections that first emerges prenatally around 37 weeks gestation. The patchiness characteristic of the horizontal connections in adults (Gilbert et al., 1996; Stettler et al., 2002) begins emerging at 7 weeks post-natal and is anatomically "adult-like" by 24 months (Burkhalter et al., 1993; also see Galuske and Singer, 1996 for a similar description of the development of horizontal connections in cats). Computational models of development in the visual system strongly suggest that the spatial distribution of horizontal connections in the cortex can arise from self-organization following visual input (Voges et al., 2010) and from processing "real" images (Prodöhl et al., 2003). For example, Grossberg and Williamson (2001) implemented a (modeled) period of exuberant growth and a period of refinement for horizontal connections following initial visual input by emphasizing the role of balance between excitation and inhibition. Similarly, Choe (2001) demonstrated that these horizontal connections link columns whose orientations are collinear, and that the connection statistics match the edge co-occurrence statistics in natural scenes (Geisler et al., 2001). It appears, therefore, that considerable visual development occurs during the post-natal period, including the development of contour detection capabilities.

Perceiving contours embedded in noise

Prior to beginning our review of the influence of Gestalt principles on element detection and contour integration, we first present a summary of approaches and stimuli used in the more recently emergent literature investigating these questions. When perceiving natural scenes, contours must be detected despite the high degree of visual noise obscuring the signal at the retina. For example, within natural scenes such as a field of flowers there are typically multiple overlapping contours referring to multiple different objects, patterns or depth information. Careful psychophysical methods analogous to this signal extraction problem have been developed using Gabor patch contours embedded in noise. Gabor elements are ideal stimuli with which to measure contour detection in the visual system since the Gabor elements model the orientation selective cells in V1. Perception of a contour composed of Gabor elements relies on the long-range horizontal connections between these orientation selective cells. Using Gabor patches to study contour detection visual noise is done by manipulating relative noise density, or the ratio of the density (D) of surrounding noise elements over the density of elements on a contour. For example, $D = 1.0$ means that the density of elements on the contour matches that of the noise elements, while $D < 1.0$ means that the density of the contour elements is less than those on the contour and $D > 1.0$ means that the density of the contour is greater than the density of the noise. Adult participants are relatively good at detecting contours embedded in noise, the minimum noise density ratio at which a contour can still be detected is $D = 0.67$ (Kovács et al., 1999).

Developmental work has started to document contour detection thresholds, and thus the functionality of long-range horizontal connectivity, in children. Using a mobile conjugate reinforcement procedure in which infants learn to kick to move a mobile consisting of three cards displaying either Gabor contours embedded in noise or only noise (e.g., circle vs. noise), Gerhardstein et al. (2004) assessed contour detection in 3-month old infants (see **Figure 1**). Infants were trained with one stimulus and tested with the other 24 h after training; baseline kick rate in response to the (new) test stimulus was taken as evidence that infants could discriminate between the two. Gerhardstein et al. (2004) found that for circular contours, at 3-months of age $D = 0.9$ was the minimum noise density ratio for contour detection. In other words, infant kick rate was greater than baseline in the immediate test, demonstrating that the infants could discriminate the stimulus from noise and no different from baseline in the discrimination test 24 h later demonstrating that the infants could discriminate between the stimuli. The applicability of the mobile conjugate reinforcement procedure for studying contour detection across older ranges of development, however, is limited.

Alternative procedures have been developed to study contour detection abilities across development. Baker et al. (2008) used a visual expectation cueing paradigm and an eye-tracker to assess detection in 6-month old infants, in a procedure in which the presentation of a square composed of Gabor elements predicted the subsequent appearance of a target on one side of the screen and a circle composed of Gabor elements predicted the subsequent appearance of a target on the other side of the screen. Predictive

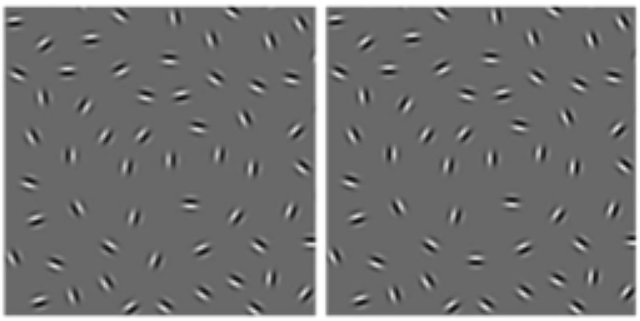


FIGURE 1 | Example of the open (left) and closed (right) contours composed of Gabor elements used in Gerhardstein et al. (2004) at $D = 0.09$.

(anticipatory) looks to the correct side for the target stimulus following the presentation of the contour (square vs. circle) were evidence that infants could detect and discriminate between the contours. Overall, Baker et al. (2008) found that 6-month-old infants could accurately detect and discriminate the shape of a contour embedded in noise only when the noise density ratio was $D = 0.90$ or higher, similar to 3-month-olds, suggesting that little functional development of this ability takes place in the first 6 months.

Research with older children and adults suggests that noise density continues to play a role in contour detection across a much longer range of development (Kovács et al., 1999; see also Benedek et al., 2010). When participants are asked to point at a contour presented in a Gabor patch on a card held in front of them, the minimum noise density ratio at which a contour can still be detected is $D = 0.84$ at 5–6 years, $D = 0.70$ at 13–14 years and $D = 0.67$ into adulthood (Kovács et al., 1999). In contrast, when the Gabor patch contours were presented on a computer screen until the participant responded or for a maximum of 15 s, Hipp et al. (2014) found that children under 9 years of age could not perform above a 75% accuracy threshold at noise density ratios of $D = 0.90$. Importantly, Kovács et al. (1999) determined the minimum D for each age group by the last correctly identified card, while Hipp et al. (2014) used a more conservative threshold measure of responding correctly 75% of the time to a given D to control for chance. Nevertheless, noise density plays an important role in contour detection during development and the tolerance for noise density when detecting contours increases across development.

Gestalt principles for elemental detection and integration

Separating the proximity and collinearity principles functionally is difficult by some definitions. Indeed, it may be prudent to consider them as aspects of a single description of the relation between two or more parts of the visual scene. Given this, it is perhaps no surprise that much of the behavioral research on perceptual grouping manipulates both proximity and collinearity. Following the work on flanker facilitation (e.g., Polat and Sagi, 1993), the role of collinearity in contour integration has been determined by *jittering* Gabor elements along a contour (Field et al., 1993) as well as through the use of noise manipulations. Jitter refers to a manipulation in which a contour is first rendered using co-aligned, identical Gabor elements that fall on a (typically curved) path embedded in noise (Gabor elements of the same spatial frequency and phase, but random orientation and position). Elements on the contour are then jittered by a manipulated amount in a random direction, to reduce the extent to which contour elements follow the true path of the contour, and the level of such jitter at which detection ceases is the threshold. Field et al. (1993) found that adult contour detection dropped off rapidly after about 15° of orientation disparity between elements, suggesting that the greater the collinearity from element to element on the contours, the more easily they were detected from a field of random noise elements. Similarly, participants can perform contour detection even over the relatively large inter-element distances of 0.9° (Field et al., 1993), suggesting that spatial integration can occur over large areas of the visual cortex. Indeed, the long-range horizontal interconnections between neurons span cortical distances of up to 8 mm (Gilbert et al., 1996). Overall, contours are easier to detect from a background of randomly oriented noise elements of the same size and shape if elements are proximal and coaligned elements (Field et al., 1993; for a more recent example see also Beaudot and Mullen, 2001).

Early in development, proximity between the elements on a contour plays a larger role in determining the detectability of the contour. Using Gabor stimuli, Hipp et al. (2014) noted that when inter-element spacing was 9λ (which is quite far apart, such that spacing is analogous to object contours that are partly occluded in the visual scene) 7–9 year olds only detected contours when $D = 1.00$, and 5–6 year old children failed to detect the contour reliably even at that level. However, when the inter-element spacing was reduced from 9 to 4.5λ, 7–9 year olds performance was nearly adult-like, and 5–6 year old children were able to detect the contour at the $D = 0.90$ level. Performance was also improved even in 3–4 year olds, who improved from not being able to detect the contour at all to being able to detect the contour at $D = 1.0$ at 4.5λ. In other words, doubling proximity while keeping relative noise ratio constant dramatically improved performance across a broad span of developmental time. Importantly, in adults the noise density tolerated for contour detection is relatively independent of the proximity between elements (Kovács et al., 1999).

Other research in developmental psychophysics investigating the use of local heuristics in contour detection supports the adult data, and suggests that the effects of collinearity and proximity are not independent. Hadad et al. (2010a) measured the ability to detect an egg-shaped contour constructed of Gabor elements by adults and children aged 7–14 years. Overall, adults and older children demonstrated a higher tolerance for noise density as collinearity increased, while proximity played more of a role when collinearity decreased (increased jitter between contour elements). In contrast, in 7-year-old children both proximity and collinearity play a significant role such that even when collinearity is high, children were hindered by low proximity. By 14 years of age, children rely less on proximity when collinearity is high, but are not yet adult-like. Notably, greater reliance on proximity for contour integration early in childhood may reflect functionally shorter-range horizontal connections early in development (for a similar argument, see Kovács et al., 1999; Kovács, 2000; Hipp et al., 2014). If so, it may be the case that the protracted development of contour integration is potentially sourced in the extended development of this aspect of the physiology of the visual system (see also Benedek et al., 2010).

It appears, then, that developing humans acquire correlations in orientation information (i.e., collinearity) within a limited spatial extent around a particular location (i.e., proximity). This spatial extent appears to expand with age and experience. The development of these proximity and collinearity heuristics in the visual system is suggestive of developmental statistical learning, progressing at a rate that depends on the robustness of the natural correlations that support it. Indeed, Geisler et al. (2001) and Geisler and Perry (2009) documented the edge co-occurrence statistics in natural scenes which suggested

that, in natural scenes, the rate at which edge elements share orientation drops off rapidly with distance from a target. Behaviorally, Hall et al. (2010) reported increased detectability for targets whose temporal presentation sequence mirrored statistical regularities as outlined by Geisler et al. (2001). That is, collinearity in nature weakens with increased spatial-temporal distance.

The use of proximity and collinearity heuristics for contour detection and integration appear to have different developmental trajectories. The use of proximity information appears to begin early in development (Hipp et al., 2014). However, the distances required for successful detection and the noise levels tolerated are greatly reduced in infants and children compared to adults, and develop gradually throughout ontogeny (Hipp et al., 2014). In contrast, the use of proximity information appears to begin later on in childhood (e.g., Hadad et al., 2010a). With respect to the physiological development of the visual system, these results support the neurophysiological data suggesting significant developments in axonal lengths and neuron density facilitating the development of long-range horizontal connections in V1 occurring across the first several years of life (Burkhalter et al., 1993). Moreover, it may be that these studies also index the development of horizontal connectivity in V2, where receptive field sizes are greater, but this remains an open question.

Physiology for higher-order contour integration

Although the processes fundamental to spatial integration of disparate contour elements likely occur in V1 (Polat, 2013), recent research suggests that the likely cortical site of larger-scale contour representation is V2 (Huang et al., 2006) indicating that these integrative processes might scale with receptive field size. The proximity and collinearity effects found in flanker facilitation tasks extend to larger-scale contour integration (Polat, 1999; Polat and Bonneh, 2000; Cass and Spehar, 2005; Zhaoping, 2011), such that elements are grouped into contours if they share orientation information and are sufficiently close together (see Geisler et al., 2001). Like V1, excitatory and inhibitory long-range horizontal connections in area V2 are likely to be the physiological source for the implementation of a contour integration mechanism and are invoked by multiple models of contour integration in vision (e.g., Li, 1998, 2002; Yen and Finkel, 1998; Usher et al., 1999; Gintautas et al., 2011; Zhaoping, 2011; Piëch et al., 2013).

Evidence of differential processing of lower-level properties and higher-level properties in the visual system has been demonstrated using a monoptic/dichoptic masking procedure to test adult participants for perceptual after-effects of closed and open contours (Sweeny et al., 2011). Monoptic masking is known to disrupt lower-level visual processing and spare higher-order processing, while dichoptic masking affects processing in the opposite way. Sweeny et al. (2011) found closed contour after-effects were evoked following monoptic, but not dichoptic masking, while the opposite pattern was found for open contours. This result supports the idea that contour integration via a closure mechanism is implemented in visual areas beyond V1 in the pathway. Specifically, implementation of the global closure heuristic during visual processing likely occurs in either area V2, thought to be the site of global contour integration (Huang et al., 2006), or area V4, which performs population coding of shape (Pasupathy and Connor, 2002). Nevertheless, long-range connections within and between cortical sites provide a mechanism through which the input from several receptive fields can interact and bind together spatially disparate segments of a contour using a global closure heuristic. Neural synchrony resulting from the oscillation of these excitatory neurons is argued to be the binding mechanism (Kovács, 1996; Yen et al., 1998; Sweeny et al., 2011; see also Gilad et al., 2013). The idea is that a reciprocal relation exists between the strength of neural synchrony and the salience of the contours. Global closure may therefore influence local level feature enhancement in a top-down fashion (Mathes and Fahle, 2007).

In adults, a delicate balance between neural synchrony-mediated excitation and surround suppression-mediated inhibition controls the characteristics of local and global contextual modulation found in various perceptual grouping tasks (Yen and Finkel, 1998). This design inherently requires neural responses to balance the involvement of excitatory and inhibitory circuits simultaneously (Grossberg and Williamson, 2001). Developmentally, acquiring this essential balance is critical for flexible perceptual learning and achievement of reliable perceptual grouping in adulthood (Grossberg and Williamson, 2001; Pinto et al., 2010). One mechanism responsible for achieving balance in neural synchrony is GABAergic expression responsible for local inhibition in the visual cortex, which is known to develop throughout the lifespan (Pinto et al., 2010). This inhibition is thought to underpin the oppositely signed surround portion of the oriented center-surround receptive fields in early visual cortex. This GABAergic expression undergoes three "main transition stages" in which rapid switches in GABAergic signaling in visual cortex occur – one in early childhood, another in early teenage years and yet another as signs of aging commence (Pinto et al., 2010). Given the developmental psychophysics research described above, it seems likely that similar developmental neurochemical foundations underlie the development of excitatory circuits.

Gestalt principles for higher-order contour integration

Closure represents a global heuristic for contour integration, depending on the higher-order pattern of relations between more than two elements. Psychophysical studies show that adults exhibit a *closure superiority effect*; that is, detectability of closed figures is enhanced relative to open figures (Kovács and Julesz, 1993; Mathes and Fahle, 2007; Machilsen and Wagemans, 2011; Gerhardstein et al., 2012). For instance, using a contour detection task with adults, Kovács and Julesz (1993) incrementally added co-aligned elements to a circular contour and found that performance was not enhanced until the contour was closed. Closure therefore elicited a pop-out effect, by their interpretation. While there has been some contention regarding whether a global heuristic such as closure needs to be invoked to explain the closure superiority effect (Tversky et al., 2004), recent research (Gerhardstein et al., 2012) strongly suggests that such a mechanism does operate in the visual system. By separately manipulating collinearity and closure using circles and S contours, Gerhardstein et al. (2012) showed that closure

enhances detectability of a contour separate from local grouping heuristics.

Closure facilitates contour integration (Pettet et al., 1998; Mathes and Fahle, 2007; Gerhardstein et al., 2012), object detection (Machilsen and Wagemans, 2011), texture-segmentation (Atkinson and Braddick, 1992; Norcia et al., 2005; Machilsen and Wagemans, 2011), and figure-ground segmentation (Field et al., 1993; Kovács and Julesz, 1993; Kovács, 1996). To date, few studies have explored the development of such a closure mechanism across childhood (Gerhardstein et al., 2004; Hadad and Kimchi, 2006; Baker et al., 2008; Hadad et al., 2010a; Hipp et al., 2014). Using a mobile conjugate reinforcement procedure, Gerhardstein et al. (2004) found that unlike adults, 3- to 4-month-old infants show no evidence of a closure superiority effect when detecting contours embedded in noise regardless of noise density; manipulation of contour type (open or closed) did not affect sensitivity to the contour at this age. Moreover, at 3–9 years of age children appear to use the local proximity heuristic rather than closure when detecting closed and open contours composed of Gabor elements and embedded in noise (Hipp et al., 2014). Specifically, children failed to show a closure superiority effect at 4.5λ or 9λ, although overall contour detection performance was better when proximity was 4.5λ rather than 9λ. Adults, in contrast, demonstrated a closure superiority effect at both 4.5 and 9λ and at the highest noise density level, $D = 0.80$. Thus, the presence of closure information for contour integration does not appear to compensate for children's dependence on proximity information early in development.

The interaction between the local heuristics of proximity and collinearity and the global closure heuristic appears to change across development. Using a different procedure, Hadad and Kimchi (2006) tested children aged 5 and 10 years and adults on their ability to detect a concave shape embedded among convex shapes in a visual display. The shapes were composed of disconnected line segments that were either open or closed. Notably, this procedure was a visual search task to determine the role of closure on visual search efficiency. Overall, performance by 10 year old children and adults was unaffected by changes in proximity when closure and collinearity information was available. However, at 5 years of age, children's concave contour detection performance was affected by decreasing proximity between line segments regardless of whether closure alone or closure and collinearity information was available. Overall, research with children suggests that a closure mechanism may not function at adult levels until into adolescence (e.g., Pennefather et al., 1999). Thus, it appears that the global closure mechanism also undergoes a prolonged developmental trajectory, gradually evoked and tuned across childhood and into adolescence. In sum, the reviewed psychophysics data (Kovács et al., 1999; Gerhardstein et al., 2004; Hadad and Kimchi, 2006; Hadad et al., 2010a; Hipp et al., 2014) suggest an extended developmental trajectory of the visual system that may be explained by physiological development (e.g., Burkhalter et al., 1993).

This interaction between proximity and collinearity also affects perception of the illusory contours formed by Kanizsa squares. To perceive the illusory contour created by Kanizsa elements, the perceiver needs to bind the Kanizsa elements into an object contour by filling in the gaps of the Kanizsa elements. It is perhaps not surprising that although when bound together into an illusory contour, the elements form a closed contour, the proximity heuristic is particularly important. Proximity within Kanizsa squares is defined by a support ratio, the length of the contour specified by the Kanizsa elements to the total length of the illusory contour. Higher support ratios typically result in stronger illusory contour perception given that the observer must traverse a smaller gap to perceive the contour. For example, Watanabe and Oyama (1988) found that Kanizsa illusory squares were perceived as stronger (e.g., greater contrast and clarity) when proximity between the four elements was high (see also Shipley and Kellman, 1992; Hadad et al., 2010b). Indeed, 4-month old infants perceive an illusory contour formed by a Kanizsa square as an occluding object only when proximity was high and the square formed a narrow occluder (Bremner et al., 2012). Thus, the greater dependence upon the proximity heuristic for illusory contours is may reflect limitations in the distance projected by the horizontal connections in the visual system.

Within the context of whole object perception, for young infants, contour integration may be achieved by a greater reliance on the grouping heuristic of common fate. Indeed, sensitivity to motion develops around 3- to 4-months and may provide a scaffold for the use of proximity and collinearity heuristics in later infancy (Johnson and Aslin, 1996, 1998; Smith et al., 2003; Johnson et al., 2012). Using occluded objects on a textured background, Johnson and Aslin (1996, 1998), Smith et al. (2003), and Johnson et al. (2012) found that 3- to 4-month old infants could perceive object unity when the two visible portions of an object were moving together. In contrast, when there was no motion information available infants did not perceive object unity for a partly occluded object (Kellman and Spelke, 1983). Importantly, common motion is not the sole factor for perceiving object unity when objects are partly occluded. For example, Johnson (2004) found that infants were better able to perceive object unity when the occluding object was narrow, compared to a wide occluding object. The early use role of motion for contour integration consistent with the earlier development of the M-pathway in the infant visual system compared to the horizontal connections (Burkhalter et al., 1993).

FUTURE DIRECTIONS

Taken together, the findings discussed in the present review inform research on the development of object perception in a number of ways. With respect to distinguishing a stationary object from the background, the principles of proximity (which will likely be high if the object is not occluded), collinearity (depending upon the object's shape), and the emergent property of closure all appear to play a role. Moreover, according to the research reviewed (e.g., Johnson et al., 2012), for infants, a moving object is clearly easier to segment from the background than a static object, demonstrating the importance of the motion-based "common fate" heuristic. Importantly, the research in the present review informs the development of bottom-up processes for object perception and does not consider the role of top-down processes (e.g., Needham et al., 2005; for review, see Quinn and Bhatt, 2009), although as with the development of horizontal connections, physiological findings also suggest a protracted development

of feedback connections in the visual system (Burkhalter, 1993). However, many studies on object perception lack the low-level control employed in the contour detection and integration psychophysics studies discussed in the present review, for example controlling color, background noise, brightness, and depth cues. Thus, to more accurately map the findings discussed here onto those investigating the development of object perception, a set of studies marrying the methods of the lower-level psychophysics studies with higher-level object perception investigations would be informative.

Within the psychophysics literature on contour detection and integration, developmental studies are relatively sparse and as such, there has been very little systematic documentation on the development of these abilities. The role of noise density on contour detection when stimuli are composed of Gabor elements has been systematically studied, documenting a progressive increase in the tolerance for noise elements across development and into adulthood (Kovács et al., 1999; Gerhardstein et al., 2004; Baker et al., 2008; Hipp et al., 2014). The use of Gestalt heuristics for contour detection across development, however, has not been documented systematically. For example, studies investigating the use of the closure heuristic leap from investigating 3- to 4-month old infants (Gerhardstein et al., 2004) to 3–9 year old children (Hipp et al., 2014). Additionally, studies investigating proximity begin with investigation of children from 3 to 4 years (Kovács et al., 1999; Hipp et al., 2014) and studies investigating collinearity start with investigation of children at 7 years of age (Hadad et al., 2010a). Moreover, the terms "contour detection" and "contour integration" have been used to refer to a number of different tasks from detecting contours composed of Gabor elements (e.g., Baker et al., 2008; Hadad et al., 2010a; Hipp et al., 2014), illusory contours using Kanizsa squares (e.g., Hadad et al., 2010b) and a visual display of concave and convex shapes (e.g., Hadad and Kimchi, 2006). While each task clearly calls upon the long-range horizontal connections in the visual system, a systematic investigation considering the differences between the tasks is needed. Future work should focus on a systematic study of the development of contour detection across development from infancy and childhood, through adolescence and into adulthood.

By systematically tracking the development of the visual system from functional onset early in infancy to adult-level functioning in adolescence and into adulthood, we can begin to infer how the visual system continues to develop physiologically. Eye tracking methodology may provide one means by which the development of contour detection can be systematically documented given that this method can be used across development (e.g., Taylor and Herbert, 2014). Furthermore, although it is clear that contour detection occurs early on in the visual system (e.g., Huang et al., 2006), it is not possible to conclude whether the majority of the contour detection mechanisms are implemented in V1 or in V2, a region containing cells with a larger receptive field (e.g., Smith et al., 2001).

CONCLUSION

While the visual system appears to be functional early on in development, it is clear from the present review that adult-level functionality does not begin to emerge until late in childhood and early adolescence (Kovács et al., 1999; Gerhardstein et al., 2004; Hadad and Kimchi, 2006; Hadad et al., 2010a,b; Hipp et al., 2014). Specifically, Burkhalter et al. (1993) note that the patchiness characteristic of the horizontal connections is anatomically "adult-like" by 24-months (also see Burkhalter, 1993). In contrast, psychophysics data demonstrates that while 3- to 6-month old infants are capable of detecting contours embedded in noise (e.g., Gerhardstein et al., 2004; Baker et al., 2008), the use of proximity, collinearity and closure information apparently does not become adult-like until preadolescence or later (e.g., Kovács et al., 1999; Hadad and Kimchi, 2006; Hadad et al., 2010a; Hipp et al., 2014). Thus, the developmental time course for physiology and psychophysics appear to differ considerably but nonetheless suggest a protracted development for contour processing.

The difference between functional physiological development of the visual system in childhood and a functionally mature physiological visual system in adulthood may explain the disparity between behavioral and physiological data. In addition, the extended physiological development of the visual system may be related to the extent and features of the visual input (see Gilbert et al., 2001). For example, by exploiting congenital cataract, Maurer et al. (1999) found that visual acuity begins developing within the first hour of receiving visual input, but not before. Importantly, in adulthood, short exposure to visual input that includes edges with orthogonal alignments facilitates orthogonal contour detection as mediated by changes in the neural representation (Schwarzkopf et al., 2009). Visual input therefore remains an important tool for mediating contour detection in the visual system (Gilbert et al., 2001; Sagi, 2011) and may account in part for the protracted development of the visual system.

To conclude, contour detection appears to become increasingly sensitive to long-range correlations in the visual world as development proceeds, with the eventual magnitude of this span not fully realized until at least adolescence. Physiologically, ontogeny is likely characterized by increases in efficiency of the plexus of horizontal connectivity connecting cortical columns in V1 and V2 in the visual cortex. This intrinsic connectivity thus becomes increasingly effective at integrating representations over greater and greater cortical distances as expertise with short-range pairings based on orientation is achieved. This process likely proceeds into adulthood, as experience is gleaned with less common – but still robust – longer-range correlations present in nature.

REFERENCES

Alberts, J. R. (1984). "Sensory-perceptual development in the Norway rat: A view towards comparative studies," in *Comparative Perspectives on Memory Development*, eds R. Kail and N. S. Spear (Hillsdale, NJ: Erlbaum), 65–101.

Atkinson, J., and Braddick, O. (1992). Visual segmentation of oriented textures by infants. *Behav. Brain Res.* 49, 123–131. doi: 10.1016/S0166-4328(05)80202-5

Atkinson, J., and Braddick, O. (2007). Visual and visuocognitive development in children born very prematurely. *Prog. Brain Res.* 164, 123–149. doi: 10.1016/S0079-6123(07)64007-2

Baker, T. J., Tse, J., Gerhardstein, P., and Adler, S. A. (2008). Contour integration by 6-month-old infants: discrimination of distinct contour shapes. *Vision Res.* 48, 136–148. doi: 10.1016/j.visres.2007.10.021

Banks, M., and Salapatek, P. (1981). Infant pattern vision: a new approach based on the contrast sensitivity function. *J. Exp. Child Psychol.* 31, 1–45. doi: 10.1016/0022-0965(81)90002-3

Beaudot, W. H., and Mullen, K. T. (2001). Processing time of contour integration: the role of colour, contrast, and curvature. *Perception* 30, 833–854. doi: 10.1068/p3164

Benedek, K., Janáky, M., Braunitzer, G., Rokszin, A., Kéri, S., and Benedek, G. (2010). Parallel development of contour integration and visual contrast sensitivity at low spatial frequencies. *Neurosci. Lett.* 472, 175–178. doi: 10.1016/j.neulet.2010.02.001

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147. doi: 10.1037/0033-295X.94.2.115

Blum, H. (1967). "A transformation for extracting new descriptors of shape," in *Models for the Perception of Speech and Visual Form*, ed. W. Wathen-Dunn (Amsterdam: MIT Press), 362–380.

Bornstein, M. H., Kessen, W., and Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 115–129. doi: 10.1037/0096-1523.2.1.115

Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* 17, 2112–2127.

Braddick, O., and Atkinson, J. (2011). Development of human visual function. *Vision Res.* 51, 1588–1609. doi: 10.1016/j.visres.2011.02.018

Braddick, O. J., Wattam-Bell, J., and Atkinson, J. (1986). Orientation-specific cortical responses develop in early infancy. *Nature* 320, 617–619. doi: 10.1038/320617a0

Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U. C., and Spring, J. (2012). Illusory contour figures are perceived as occluding contours by 4-month-old infants. *Dev. Psychol.* 48, 398–405. doi: 10.1037/a0024922

Brunswick, E., and Kamiya, J. (1953). Ecological cue-validity of proximity and of other gestalt factors. *Am. J. Psychol.* 66, 20–32. doi: 10.2307/1417965

Burkhalter, A. (1993). Development of forward and feedback connections between areas V1 and V2 of human visual cortex. *Cereb. Cortex* 3, 476–487. doi: 10.1093/cercor/3.5.476

Burkhalter, A., Bernardo, K. L., and Charles, V. (1993). Development of local circuits in human visual cortex. *J. Neurosci.* 13, 1916–1931.

Bushnell, I. W. R., Sai, F., and Mullin, J. T. (1989). Neonatal recognition of the mothers face. *Br. J. Dev. Psychol.* 7, 3–15. doi: 10.1111/j.2044-835X.1989.tb00784.x

Candy, T. R., Skoczenski, A. M., and Norcia, A. M. (2001). Normalization models applied to orientation masking in the human infant. *J. Neurosci.* 21, 4530–4541.

Cass, J. R., and Spehar, B. (2005). Dynamics of collinear contrast facilitation are consistent with long-range horizontal striate transmission. *Vision Res.* 45, 2728–2739. doi: 10.1016/j.visres.2005.03.010

Choe, Y. (2001). *Perceptual Grouping in a Self-organizing Map of Spiking Neurons. Unpublished Doctoral Dissertation.* Austin: University of Texas.

Dickinson, S. J., Pentland, A. P., and Rosenfeld, A. (1992). From volumes to views: an approach to 3-D object recognition. *Comput. Vis. Graph. Image Process. Image Underst.* 55, 130–154. doi: 10.1016/1049-9660(92)90013-S

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field." *Vision Res.* 33, 173–193. doi: 10.1016/0042-6989(93)90156-Q

Fowlkes, C. C., Martin, D. R., and Malik, J. (2007). Local figure–ground cues are valid for natural images. *J. Vis.* 7, 1–9. doi: 10.1167/7.8.2

Galuske, R. A., and Singer, W. (1996). The origin and topography of long-range intrinsic projections in cat visual cortex: a developmental study. *Cereb. Cortex* 6, 417–430. doi: 10.1093/cercor/6.3.417

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167–192. doi: 10.1146/annurev.psych.58.110405.085632

Geisler, W. S., and Perry, J. S. (2009). Contour statistics in natural images: grouping across occlusions. *Vis. Neurosci.* 26, 109–121. doi: 10.1017/S0952523808080875

Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* 41, 711–724. doi: 10.1016/S0042-6989(00)00277-7

Gerhardstein, P., Kovács, I., Ditre, J., and Feher, A. (2004). Detection of contour continuity and closure in three-month-olds. *Vision Res.* 44, 2981–2988. doi: 10.1016/j.visres.2004.06.023

Gerhardstein, P., Lui, J., and Rovee-Collier, C. (1998). Perceptual constraints on infant memory retrieval. *J. Exp. Child Psychol.* 69, 109–131. doi: 10.1006/jecp.1998.2435

Gerhardstein, P., Tse, J., Dickerson, K., Hipp, D., and Moser, A. (2012). The human visual system uses a global closure mechanism. *Vision Res.* 71, 18–27. doi: 10.1016/j.visres.2012.08.011

Gilad, A., Meirovithz, E., Leshem, A., Arieli, A., and Slovin, H. (2012). Collinear stimuli induce local and cross-areal coherence in the visual cortex of behaving monkeys. *PLoS ONE* 7:e49391. doi: 10.1371/journal.pone.0049391

Gilad, A., Meirovithz, E., and Slovin, H. (2013). Population responses to contour integration: early encoding of discrete elements and late perceptual grouping. *Neuron* 78, 389–402. doi: 10.1016/j.neuron.2013.02.013

Gilbert, C. D., Das, A., Ito, M., Kapadia, M., and Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 93, 615–622. doi: 10.1073/pnas.93.2.615

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *J. Neurosci.* 3, 1116–1133.

Gilbert, C. D., and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* 9, 2432–2442.

Gilbert, C. D., Sigman, M., and Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron* 31, 681–697. doi: 10.1016/S0896-6273(01)00424-X

Gintautas, V., Ham, M. I., Kunsberg, B., Barr, S., Brumby, S. P., Rasmussen, C., et al. (2011). Model cortical association fields account for the time course and dependence on target complexity of human contour perception. *PLoS Comput. Biol.* 7:e1002162. doi: 10.1371/journal.pcbi.1002162

Grossberg, S., and Williamson, J. R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cereb. Cortex* 11, 37–58. doi: 10.1093/cercor/11.1.37

Hadad, B., Maurer, D., and Lewis, T. L. (2010a). The effects of spatial proximity and collinearity on contour integration in adults and children. *Vision Res.* 50, 772–778. doi: 10.1016/j.visres.2010.01.021

Hadad, B. S., Maurer, D., and Lewis, T. L. (2010b). The development of contour interpolation: evidence from subjective contours. *J. Exp. Child Psychol.* 106, 163–176. doi: 10.1016/j.jecp.2010.02.003

Hadad, B. S., and Kimchi, R. (2006). Developmental trends in utilizing perceptual closure for grouping of shape: effects of spatial proximity and collinearity. *Percept. Psychophys.* 68, 1264–1273. doi: 10.3758/BF03193726

Hall, S., Pollux, P. M., and Guo, K. (2010). Exploitation of natural geometrical regularities facilitates target detection. *Vision Res.* 50, 2411–2420. doi: 10.1016/j.visres.2010.09.011

Hipp, D., Moser, A., Dickerson, K., and Gerhardstein, P. (2014). Age-related changes in visual contour integration: implications for physiology from psychophysics. *Dev. Psychobiol.* doi: 10.1002/dev.21225 [Epub ahead of print].

Hou, C., Pettet, M. W., Sampath, V., Candy, T. R., and Norcia, A. M. (2003). Development of the spatial organization and dynamics of lateral interactions in the human visual system. *J. Neurosci.* 23, 8630–8640.

Huang, P. C., Hess, R. F., and Dakin, S. C. (2006). Flank facilitation and contour integration: different sites. *Vision Res.* 46, 3699–3706. doi: 10.1016/j.visres.2006.04.025

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591.

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243.

Hubel, D. H., Wiesel, T. N., and LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 278, 377–409. doi: 10.1098/rstb.1977.0050

Johnson, M., Dziurawiec, S., Ellis, H., and Morton, J. (1992). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition* 40, 1–19. doi: 10.1016/0010-0277(91)90045-6

Johnson, S. P. (2001). Visual development in human infants: binding features, surfaces, and objects. *Vis. Cogn.* 8, 565–578. doi: 10.1080/13506280143000124

Johnson, S. P. (2004). Development of perceptual completion in infancy. *Psychol. Sci.* 15, 769–775. doi: 10.1111/j.0956-7976.2004.00754.x

Johnson, S. P., and Aslin, R. N. (1996). Perceptions of object unity in young infants: the rules of motion, depth and orientation. *Cogn. Dev.* 11, 161–180. doi: 10.1016/S0885-2014(96)90001-5

Johnson, S. P., and Aslin, R. N. (1998). Young infants' perception of illusory contours in dynamic displays. *Perception* 27, 341–354. doi: 10.1068/p270341

Johnson, S. P., Bremner, J. G., Slater, A. M., Shuwairi, S. M., Mason, U., Spring, J., et al. (2012). Young infants' perception of the trajectories of two- and three-dimensional objects. *J. Exp. Child Psychol.* 113, 177–185. doi: 10.1016/j.jecp.2012.04.011

Kapadia, M. K., Westheimer, G., and Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J. Neurophysiol.* 84, 2048–2062.

Kellman, P. J., and Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cogn. Psychol.* 15, 483–524. doi: 10.1016/0010-0285(83)90017-8

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Psychology*. New York: Liveright.

Kovács, I. (1996). Gestalten of today: early processing of visual contours and surfaces. *Behav. Brain Res.* 82, 1–11. doi: 10.1016/S0166-4328(97)81103-5

Kovács, I. (2000). Human development of perceptual organization. *Vision Res.* 40, 1301–1310. doi: 10.1016/S0042-6989(00)00055-9

Kovács, I., Fehér, Á., and Julesz, B. (1998). Medial-point description of shape: a representation for action coding and its psychophysical correlates. *Vision Res.* 38, 2323–2333. doi: 10.1016/S0042-6989(97)00321-0

Kovács, I., and Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proc. Natl. Acad. Sci. U.S.A.* 90, 7495–7497. doi: 10.1073/pnas.90.16.7495

Kovács, I., Kozma, P., Fehér, Á., and Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12204–12209. doi: 10.1073/pnas.96.21.12204

Lawlor, M., and Zucker, S. W. (2013). "Third-order edge statistics: contour continuation, curvature, and cortical connections," in *Proceedings of the Advances in Neural Information Processing Systems 26*, Chicago, 1763–1771.

Lev, M., and Polat, U. (2011). Collinear facilitation and suppression at the periphery. *Vision Res.* 51, 2488–2498. doi: 10.1016/j.visres.2011.10.008

Lewis, T. L., and Maurer, D. (2005). Multiple sensitive periods in human visual development: evidence from visually deprived children. *Dev. Psychobiol.* 46, 163–183. doi: 10.1002/dev.20055

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Comput.* 10, 903–940. doi: 10.1162/089976698300017557

Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cogn. Sci.* 6, 9–16. doi: 10.1016/S1364-6613(00)01817-9

Loffler, G. (2008). Perception of contours and shapes: low and intermediate stage mechanisms. *Vision Res.* 48, 2106–2127. doi: 10.1016/j.visres.2008.03.006

Machilsen, B., and Wagemans, J. (2011). Integration of contour and surface information in shape detection. *Vision Res.* 51, 179–186. doi: 10.1016/j.visres.2010.11.005

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* New York, NY: Henry Holt and Co. Inc.

Mathes, B., and Fahle, M. (2007). Closure facilitates contour integration. *Vision Res.* 47, 818–827. doi: 10.1016/j.visres.2006.11.014

Maurer, D., Lewis, T. L., Brent, H. P., and Levin, A. V. (1999). Rapid improvement in the acuity of infants after visual input. *Science* 286, 108–110. doi: 10.1126/science.286.5437.108

Morrone, M. C., and Burr, D. C. (1986). Evidence for the existence and development of visual inhibition in humans. *Nature* 321, 235–237. doi: 10.1038/321235a0

Needham, A., Dueker, G., and Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition* 94, 215–240. doi: 10.1016/j.cognition.2004.02.002

Nelson, J. I., and Frost, B. J. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Exp. Brain Res.* 61, 54–61. doi: 10.1007/BF00235620

Norcia, A., Pei, F., Bonneh, Y., Hou, C., Sampath, V., and Pettet, M. (2005). Development of sensitivity to texture and contour information in the human infant. *J. Cogn. Neurosci.* 7, 569–579. doi: 10.1162/0898929053467596

Pasupathy, A., and Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/nn972

Patel, P., Maurer, M., and Lewis, L. (2010). The development of spatial frequency discrimination. *J. Vis.* 10, 1–10. doi: 10.1167/10.14.41

Pennefather, P. M., Chandna, A., Kovács, I., Polat, U., and Norcia, A. M. (1999). Contour detection threshold: repeatability and learning with 'contour cards.' *Spat. Vis.* 12, 257–266. doi: 10.1163/156856899X00157

Peterson, M. A. (2001). "Object perception," in *Blackwell Handbook of Sensation and Perception*, Chap. 6, ed. E. B. Goldstein (Oxford: Blackwell), 168–203.

Pettet, M. W., McKee, S. P., and Grzywacz, N. M. (1998). Constraints on long range interactions mediating contour detection. *Vision Res.* 38, 865–879. doi: 10.1016/S0042-6989(97)00238-1

Piëch, V., Li, W., Reeke, G. N., and Gilbert, C. D. (2013). Network model of top-down influences on local gain and contextual interactions in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4108–E4117. doi: 10.1073/pnas.1317019110

Pinto, J. G., Hornby, K. R., Jones, D. G., and Murphy, K. M. (2010). Developmental changes in GABAergic mechanisms in human visual cortex across the lifespan. *Front. Cell. Neurosci.* 4:16. doi: 10.3389/fncel.2010.00016

Piper, M., and Darrah, J. (1994). *Motor Assessment of the Developing Infant.* Philadelphia: Saunders.

Polat, U. (1999). Functional architecture of long-range perceptual interactions. *Spat. Vis.* 12, 143–162. doi: 10.1163/156856899X00094

Polat, U. (2013). Spatial and temporal rules for contextual modulations. *Paper presented at the 13th Annual Meeting of the Vision Sciences Society*, Naples, FL.

Polat, U., and Bonneh, Y. (2000). Collinear interactions and contour integration. *Spat. Vis.* 13, 393–402. doi: 10.1163/156856800741270

Polat, U., and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res.* 33, 993–999. doi: 10.1016/0042-6989(93)90081-7

Prodöhl, C., Würtz, R. P., and Von Der Malsburg, C. (2003). Learning the gestalt rule of collinearity from object motion. *Neural Comput.* 15, 1865–1896. doi: 10.1162/08997660360675071

Quinn, P. C., and Bhatt, R. S. (2009). Perceptual organization in infancy: bottom-up and top-down influences. *Optom. Vis. Sci.* 86, 589–594. doi: 10.1097/OPX.0b013e3181a5238a

Rockland, K. S., and Lund, J. S. (1982). Widespread periodic intrinsic connections in the tree shrew visual cortex. *Science* 215, 1532–1534. doi: 10.1126/science.7063863

Sagi, D. (2011). Perceptual learning in vision research. *Vision Res.* 51, 1552–1566. doi: 10.1016/j.visres.2010.10.019

Schwarzkopf, D. S., Zhang, J., and Kourtzi, Z. (2009). Flexible learning of natural statistics in the human brain. *J. Neurophysiol.* 102, 1854–1867. doi: 10.1152/jn.00028.2009

Shani, R., and Sagi, D. (2005). Eccentricity effects on lateral interactions. *Vision Res.* 45, 2009–2024. doi: 10.1016/j.visres.2005.01.024

Shipley, T. F., and Kellman, P. J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Percept. Psychophys.* 52, 97–106. doi: 10.3758/BF03206762

Sireteanu, R., Fronius, M., and Constantinescu, D. H. (1994). The development of visual acuity in the peripheral visual field of human infants: binocular and monocular measurements. *Vision Res.* 34, 1659–1671. doi: 10.1016/0042-6989(94)90124-4

Slater, A. J., and Johnson, S. P. (1998). "Visual sensory and perceptual abilities of the newborn: beyond the blooming, buzzing confusion," in *The Development of Sensory, Motor and Cognitive Capacities in Early Infancy: From Perception to Cognition*, eds F. Simion and G. Butterworth (Hove: Psychology Press/Erlbaum), 121–141.

Smith, A. T., Singh, K. D., Williams, A. L., and Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* 11, 1182–1190. doi: 10.1093/cercor/11.12.1182

Smith, W. C., Johnson, S. P., and Spelke, E. S. (2003). Motion and edge sensitivity in perception of object unity. *Cogn. Psychol.* 46, 31–64. doi: 10.1016/S0010-0285(02)00501-7

Stettler, D. D., Das, A., Bennett, J., and Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* 36, 739–750. doi: 10.1016/S0896-6273(02)01029-2

Sweeny, T. D., Grabowecky, M., and Suzuki, S. (2011). Awareness becomes necessary between adaptive pattern coding of open and closed curvatures. *Psychol. Sci.* 22, 943–950. doi: 10.1177/0956797611413292

Taylor, G., and Herbert, J. S. (2014). Infant and adult visual attention during an imitation demonstration. *Dev. psychobiol.* 56, 770–782. doi: 10.1002/dev.21147

Ts'o, D. Y., Gilbert, C. D., and Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *J. Neurosci.* 6, 1160–1170.

Tversky, T., Geisler, W. S., and Perry, J. S. (2004). Contour grouping: closure effects are explained by good continuation and proximity. *Vision Res.* 44, 2769–2777. doi: 10.1016/j.visres.2004.06.011

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn. Sci.* 11, 58–64. doi: 10.1016/j.tics.2006.11.009

Usher, M., Bonneh, Y., Sagi, D., and Herrmann, M. (1999). Mechanisms for spatial integration in visual detection: a model based on lateral interactions. *Spat. Vis.* 12, 187–209. doi: 10.1163/156856899X00111

Voges, N., Schüz, A., Aertsen, A., and Rotter, S. (2010). A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Prog. Neurobiol.* 92, 277–292. doi: 10.1016/j.pneurobio.2010.05.001

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychol. Bull.* 138, 1172–1217. doi: 10.1037/a0029333

Watanabe, T., and Oyama, T. (1988). Are illusory contours a cause or a consequence of apparent differences in brightness and depth in the Kanizsa square. *Perception* 17, 513–521. doi: 10.1068/p170513

Wattam-Bell, J., Birtles, D., Nyström, P., von Hofsten, C., Rosander, K., Anker, S., et al. (2010). Reorganization of global form and motion processing during human visual development. *Curr. Biol.* 20, 411–415. doi: 10.1016/j.cub.2009.12.020

Wertheimer, M. (1958). "Principles of perceptual organization," in *Readings in Perception*, eds D. C. Beardslee and M. Wertheimer (Princeton, NJ: D. Van Nostrand Company), 115–135.

Yen, S. C., and Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Res.* 38, 719–741. doi: 10.1016/S0042-6989(97)00197-1

Yen, S. C., Menschik, E. D., and Finkel, L. H. (1998). "Cortical synchronization and perceptual salience," in *Computational Neuroscience: Trends in Research 1998*, ed. J. M. Bower (New York, NY: Plenum Publishing Co.), 125–130.

Zhaoping, L. (2011). Neural circuit models for computations in early visual cortex. *Elsevier* 21, 808–815. doi: 10.1016/j.conb.2011.07.005

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: 03 February 2014; accepted: 21 June 2014; published online: 08 July 2014.
*Citation: Taylor G, Hipp D, Moser A, Dickerson K and Gerhardstein P (2014) The development of contour processing: evidence from physiology and psychophysics. Front. Psychol. **5**:719. doi: 10.3389/fpsyg.2014.00719*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Taylor, Hipp, Moser, Dickerson and Gerhardstein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 18 December 2013
doi: 10.3389/fpsyg.2013.00948

Low-level and high-level modulations of fixational saccades and high frequency oscillatory brain activity in a visual object classification task

Maciej Kosilo[1,2], Sophie M. Wuerger[3], Matt Craddock[4], Ben J. Jennings[1,5], Amelia R. Hunt[1] and Jasna Martinovic[1]*

[1] School of Psychology, University of Aberdeen, Aberdeen, UK
[2] Department of Psychology, City University London, London, UK
[3] Department of Psychological Sciences, Institute of Psychology, Health and Society, University of Liverpool, Liverpool, UK
[4] Institute for Experimental Psychology and Methods, University of Leipzig, Leipzig, Germany
[5] Department of Ophthalmology, McGill Vision Research, McGill University, Montreal, QC, Canada

Edited by:
Chris Fields, Retired, USA

Reviewed by:
Carl M. Gaspar, University of Glasgow, UK
Susana Martinez-Conde, Barrow Neurological Institute, USA

***Correspondence:**
Jasna Martinovic, School of Psychology, University of Aberdeen, William Guild Building, Aberdeen, AB24 3FX, UK
e-mail: j.martinovic@abdn.ac.uk

Until recently induced gamma-band activity (GBA) was considered a neural marker of cortical object representation. However, induced GBA in the electroencephalogram (EEG) is susceptible to artifacts caused by miniature fixational saccades. Recent studies have demonstrated that fixational saccades also reflect high-level representational processes. Do high-level as opposed to low-level factors influence fixational saccades? What is the effect of these factors on artifact-free GBA? To investigate this, we conducted separate eye tracking and EEG experiments using identical designs. Participants classified line drawings as objects or non-objects. To introduce low-level differences, contours were defined along different directions in cardinal color space: S-cone-isolating, intermediate isoluminant, or a full-color stimulus, the latter containing an additional achromatic component. Prior to the classification task, object discrimination thresholds were measured and stimuli were scaled to matching suprathreshold levels for each participant. In both experiments, behavioral performance was best for full-color stimuli and worst for S-cone isolating stimuli. Saccade rates 200–700 ms after stimulus onset were modulated independently by low and high-level factors, being higher for full-color stimuli than for S-cone isolating stimuli and higher for objects. Low-amplitude evoked GBA and total GBA were observed in very few conditions, showing that paradigms with isoluminant stimuli may not be ideal for eliciting such responses. We conclude that cortical loops involved in the processing of objects are preferentially excited by stimuli that contain achromatic information. Their activation can lead to relatively early exploratory eye movements even for foveally-presented stimuli.

Keywords: visual object representation, parallel visual pathways, color, luminance, fixational saccades, microsaccades, EEG, gamma-band activity

INTRODUCTION

In order to acquire sufficient information from the complex and dynamically changing environment, the visual system implements various strategies. One such strategy is to perform eye movements in order to scan the visual scene, while intermittently maintaining gaze at objects of interest. The fovea is the central part of the retina with highest spatial acuity and is responsible for the acquisition of fine spatial details during fixations, making foveation an excellent strategy for acquiring visual information. Fixations themselves are dynamic events, during which different classes of small, involuntary eye movements have been recognized: these include microsaccades, drifts and tremors. Cornsweet (1956) suggested that the purpose of microsaccades is to counteract the effects of other fixational eye movements, such as tremor and drift - namely, to correct the eye position so that fixation returns to the target. Engbert and Kliegl (2004) refined Cornsweet's (1956) suggestions. Their analysis revealed that microsaccades operated on two time scales of different characteristics. On a short time scale (up to 20 ms), microsaccades increased fixation errors, thus increasing retinal image shifts. This most likely contributes to the prevention of perceptual fading (see Hubel and Wiesel, 1968). However, over longer time intervals (100–400 ms) microsaccades lead to a reduction of fixation errors so that fixation was maintained. A recent study by Mergenthaler and Engbert (2010) provided evidence for a microsaccade dichotomy of a different kind: a bimodal saccade amplitude distribution was observed when participants were asked to freely view natural scenes. Larger saccades ($>0.4°$) behaved differently than very small saccades ($<0.4°$), indicating that larger saccades during fixation could be inspection saccades rather than microsaccades. The purpose of these fixational saccades is likely to be selection or re-selection of scene attributes that are relatively close to fixation.

Saccades and microsaccades are both generally controlled by the superior colliculus (Hafed et al., 2009) which receives input directly from the retina, as well as cortical input from perceptual areas. Therefore, at the level of the superior colliculus subcortical low-level inputs converge with cortical loops that provide high-level information used for ocular control. Both bottom-up and top-down factors can modulate the rate of microsaccades (Betta and Turatto, 2006; Valsecchi et al., 2009; Laubrock et al., 2010; for reviews see Martinez-Conde et al., 2009; Rolfs, 2009; for a recent model see Engbert, 2012). In their study on low-level influences on microsaccade rates, Valsecchi and Turatto (2007) looked at microsaccadic responses to events often thought to be "invisible" to the superior colliculus since its superficial layers which receive direct retinal inputs do not support color-opponent processing (Marrocco and Li, 1977; also see White et al., 2009). Valsecchi and Turatto (2007) hypothesized that if microsaccades are generated solely by a low-level circuit involving the retina and the superior colliculus, microsaccadic rates should not be affected by the presentation of a stimulus which is isoluminant with the background. However, microsaccadic rates were very similar for both isoluminant and luminance-defined stimuli. They interpreted this as evidence that microsaccades elicited by isoluminant stimuli were driven by cortical loops. The idea that small fixational saccades can be modulated by cortical inputs was further supported by another study (Otero-Millan et al., 2008) which looked at microsaccadic responses in free-viewing and visual search tasks. In free exploration of a natural scene, the highest rates of microsaccades occurred during fixation of human faces. In the search task, large increases in microsaccade rates occurred in image regions containing identified targets. Otero-Millan et al.'s (2008) findings imply that foveation of targets is an essential determinant of microsaccadic behavior and that this is determined by high-level as well as low-level image content.

This line of research into the role of fixational saccades in object processing coincides with the findings reported by Yuval-Greenberg et al. (2008). These authors demonstrated that the brief broadband peak in the induced gamma-band frequency range in the electroencephalogram (EEG) actually reflects a peak in the rate of miniature fixational saccades. Induced GBA is high frequency (above 30 Hz) oscillatory activity which is neither time- nor phase-locked to stimulus onset, as opposed to stimulus-locked evoked activity. Until the publication of Yuval-Greenberg et al.'s (2008) study, iGBA was widely assumed to reflect a neural oscillation associated with higher-order cortical activity, including object representation, memory, attention and awareness (for more recent reviews see Tallon-Baudry, 2009; Herrmann et al., 2010; Rieder et al., 2011). However, saccades are also induced by the stimulus. Eye muscle movements associated with each saccade generate a spike in high-frequency electrical activity recorded from the scalp with EEG. Since microsaccades and induced gamma-band activity (iGBA) share similar temporal dynamics, the high-frequency output of these eye movements can be confused with a genuine cortical response. Engbert and Kliegl (2003) report a characteristic microsaccadic response after the onset of an event: the microsaccadic rate drops substantially below its normal rate, reaching a minimum at around 150 ms after event onset. This is followed by a substantial rate increase, which reaches a maximum at around 350 ms and returns to baseline level around about 500 ms after event onset. This "signature" has been consistently demonstrated in other studies in response to novel visual or auditory stimuli (for a review, see Rolfs, 2009). The timing of the broadband iGBA peak overlaps with this microsaccadic maximum, being most pronounced around 200–350 ms after the stimulus has been presented. Yuval-Greenberg et al. (2008) showed that the iGBA is time-locked to the onset of miniature saccades. However, iGBA may also coincide with microsaccades because both are triggered by similar perceptual processes (for reviews, see Melloni et al., 2009; Martinovic and Busch, 2011). Thus, iGBA is likely to contain both an artifactual, muscular component and an underlying genuine, cortically-generated oscillation. A recent study by Hassler et al. (2011) demonstrated just that: removal of the ocular artifact revealed an underlying iGBA which was still enhanced for object as opposed to non-object images.

Previous experiments on fixational saccades generally investigated low-level visual processing and its modulation by attention, while studies investigating the contribution of fixational saccades to iGBA looked at high-level vision. In this study, we aim to look at both low and high-level modulations of fixational saccades. We recorded fixational saccades using the paradigm from a previously reported EEG experiment on low and high-level factors in object classification (Martinovic et al., 2011). Since that study focused on event related potentials (ERPs), we reanalyzed its dataset to examine evoked and total GBA. Total GBA (tGBA) is a sum of both evoked and iGBA. To isolate iGBA, a common approach is to subtract the ERP from each single trial, theoretically removing evoked GBA. However, Truccolo et al. (2002) demonstrate that there is no way to remove evoked activity from the signal and be sure that what is remaining is only "induced," as removing the ERP from each trial relies on the inaccurate assumption that the evoked signal is completely stationary. This leaves residual "evoked" signals on each trial. As substantial contributions of the evoked signal to the gamma-band are largely centered in frequencies below 40 Hz, occurring before 150–200 ms, the contribution to the GBA after 200 ms is mainly driven by the induced part (e.g., see Fründ et al., 2007).

We added several additional participants in order to increase the power for the gamma frequency-band analyses, which were reliant on the algorithm for microsaccadic artifact removal proposed by Keren et al. (2010), applied successfully in a previous study by Craddock et al. (2013). Although we collected fixational saccade and tGBA data in separate experiments with different participants, which limits how strongly we can draw conclusions on their relation to each other, we were able to compare lower and higher-level influences on fixational saccades themselves and on tGBA after artifact correction. Finally, the study also aimed to examine evoked gamma-band activity (eGBA; 30–40 Hz at approx. 50–150 ms), which can be modulated by object class under specific circumstances (Herrmann et al., 2004a; Fründ et al., 2008; for a review see Martinovic and Busch, 2011) but is also highly influenced by low-level stimulus properties (Busch et al., 2004; Fründ et al., 2007). Evoked gamma-band activity has been hypothesized to reflect a memory match and to act as a precursor to iGBA by Herrmann et al. (2004b).

Participants responded to simple line-drawings presented on the screen, indicating whether these drawings showed familiar, nameable objects or novel, unnamable images (i.e., non-objects). The lines were defined along different directions in DKL color space (Derrington et al., 1984) to differentially excite post-receptoral mechanisms that are distinguished at the level of lateral geniculate nucleus. Luminance is defined as the weighted sum of L and M cone excitation, with S-cones contributing only at high levels of overall luminance (Ripamonti et al., 2009). The cone-opponent mechanisms process either the weighted difference between L and M cone excitation (L − M) or the weighted difference between S-cone excitation and a sum of L and M cone excitation [S − (L + M)]. These mechanisms roughly map onto the three visual pathways—the magnocellular pathway processes luminance information, while the parvo- and koniocellular pathways also subserve color processing (for a review, see Kulikowski, 2003). The parvocellular pathway receives L and M cone input, and is sensitive to chromatic but also to luminance information, depending on the spatial scale (Reid and Shapley, 2002). Physiological studies have revealed subdivisions within the koniocellular pathway, with its middle layers involved in S-cone information processing (Hendry and Reid, 2000; Tailby et al., 2008).

The decision to define object and non-object stimuli by signals from different post-receptoral mechanisms was motivated by predictions from Bar's (2003) model that the contribution of luminance and chromatic mechanisms to object classification is not equal. In this model, luminance information significantly contributes to the speed and efficiency of object categorization, over and above the contribution of chromatic mechanisms. Initial information on shape derived from luminance detectors is rapidly transmitted through the magnocellular pathway from early visual areas to the prefrontal cortex (PFC). In the PFC, those initial cues trigger top-down facilitation of object recognition by providing the visual system with an "initial guess" on stimulus identity. Feedback from the PFC is then transmitted to the temporal cortex where it is used to facilitate bottom-up processing. The whole process results in more rapid and efficient object categorization. A functional Magnetic Resonance Imaging (fMRI) study which looked at the processing of chromatic and achromatic object contours used dynamic causal modeling to demonstrate that achromatic stimuli triggered pathways from the visual cortex to orbitofrontal cortex and from orbitofrontal cortex to fusiform gyrus, which likely reflects the top-down facilitation in object recognition by the luminance information. On the other hand, chromatic stimuli activated a direct pathway from occipital cortex to the fusiform gyrus (Kveraga et al., 2007). We therefore compared full-color and reduced-color object (or non-object) contours. Full-color stimuli contained both chromatic and luminance information [L + M, L − M, S − (L + M)]. Luminance information was absent in the reduced-color stimuli, which either excited both of the chromatic mechanisms [S − (L + M) and L−M] or only excited the S − (L + M) mechanism. An earlier ERP study by Martinovic et al. (2011) used the same paradigm as we use here. After matching stimulus contrast across conditions by use of discrimination threshold units, they found that the inclusion of luminance information results in higher accuracy and faster reaction times for object as opposed to non-object images, as well as in a reduced N1 component for object images. These results are in line with Bar's model (2003) and Kveraga et al.'s (2007) findings. As mentioned above, Valsecchi and Turatto (2007) have demonstrated that microsaccade rates are the same for isoluminant red and green stimuli and stimuli with an additional luminance edge. Through the use of two types of contrast-matched isoluminant stimuli [S − (L + M); S − (L + M) & L − M], as well as a stimulus with both chromatic and luminance information, our study can further extend the findings of Valsecchi and Turatto (2007). There are several important methodological differences between the studies. In our study, we match contrast across different types of stimuli in terms of threshold units, while Valsecchi and Turatto (2007) used stimuli that were not matched in terms of contrast. We also further divide isoluminant contrast into contrast from two chromatic cone-opponent mechanisms. The intermediate isoluminant stimulus, which excites both L − M and S − (L + M) mechanisms, is probably similar to the stimulus from Valsecchi and Turatto (2007). However, the S − (L + M) defined stimulus is dissimilar and may be particularly interesting. Methodologically, it is less likely to contain residual luminance artifacts at the edges/lines of the stimulus, as S-cone contribution to luminance is quite limited (see Ripamonti et al., 2009). Theoretically, it is also interesting because the central fovea does not contain any Scones, so S − (L + M) signals may be less salient for the generation of microsaccades than L-M cone-opponent signals.

Isolating the S − (L + M) channel enabled us to make a specific prediction, based on the fact that the central part of the fovea, about 0.3°–0.4° in size in humans, is S-cone free (Bumsted and Hendrickson, 1999). Therefore, we expect that lower fixational saccade rates should be observed for S-cone isolating stimuli but not for tGBA. If tGBA reflects mainly higher-level, object representation processes, it should not differ between S − (L + M) and intermediate isoluminant or full color stimuli. This would in turn indicate that tGBA is predominantly reflecting higher-level, cortical mechanisms of object representation. Moreover, if fixational saccades and tGBA reflect object-sensitive mechanisms, they should be enhanced for objects, as in Hassler et al. (2011). If eGBA is absent while tGBA is present, this would signify that eGBA is not a necessary and sufficient precursor to iGBA, contrary to the model of Herrmann et al. (2004b). Existing evidence already indicates that eGBA is strongly related to luminance contrast (Schadow et al., 2007). We predicted that eGBA would be absent at least from the isoluminant conditions, as our paradigm used stimuli that should not strongly engage the magnocellular pathway which has previously been related to eGBA (Fründ et al., 2007).

MATERIALS AND METHODS

PARTICIPANTS

Twelve healthy participants (3 males, aged 20–35 years) with normal or corrected to normal vision volunteered and gave written informed consent to take part in the eye movement experiment. All participants had normal color vision as assessed with the Cambridge Color Test (Regan et al., 1994). The study was

approved by the ethics committee of the School of Psychology at the University of Aberdeen.

Eighteen healthy participants (11 males; aged 21–40 years) with normal or corrected-to-normal vision, as well as normal color vision as assessed with the Cambridge Color Test gave written informed consent to take part in the EEG experiment. One participant was subsequently removed from the sample, since more than 40% of trials were artifact-contaminated. Six further participants were removed as the ocular artifact could not be sufficiently removed from the tGBA (see section on EEG data acquisition and analysis). The participants received a small honorarium to compensate for their time. The study was approved by the ethics committee of the School of Psychology, University of Liverpool.

APPARATUS

The eye movement experiment was run on a Dell Precision PC equipped with a visual stimulus generator (Visage, Cambridge Research Systems, Ltd., Kent, UK). Stimulus presentation was controlled using Matlab (Mathworks, Natick, Massachusetts) and the stimuli were presented on a Sony GDM-520 21 inch CRT monitor. The chromatic and luminance outputs of the monitor were calibrated using the CRS calibration system (ColourCAL II, Cambridge Research Systems, Ltd., Kent, UK); the accuracy of the calibration was verified with a spectroradiometer (SpectroCal, Cambridge Research Systems, Ltd., Kent, UK). The monitor had been switched on for at least 30 min before any experiment. Participants responded via a button box (Cedrus RB-530, Cedrus Corporation, San Pedro, USA) and were seated 60 cm from the screen with their head placed in a chin rest. Binocular eye movements were recorded using an Eyelink 1000 system (SR Research, Mississauga, Ontario, Canada), which received stimulus-onset triggers from the Visage.

In the EEG experiment, an almost identical system was used for generation of stimuli and collection of responses (see Martinovic et al., 2011), with the Visage system sending triggers to a 32-electrode Biosemi Active-Two system (Biosemi, Amsterdam, Netherlands).

COLOUR SPACE

We use the DKL-color space (Derrington et al., 1984; Brainard, 1996), an extension of the MacLeod–Boynton chromaticity diagram (Macleod and Boynton, 1979), to describe the chromatic properties of our stimuli. In this space, any color is defined by modulations along three different "cardinal" axes. Along the achromatic axis, all three cone classes (L, M and S) are modulated such that the contrast is identical, that is, $\Delta L/L_{BG} = \Delta M/M_{BG} = \Delta S/S_{BG}$, where ΔL, ΔM, and ΔS denote the incremental cone excitations in three cone classes, respectively. L_{BG}, M_{BG} and S_{BG} indicate the L-, M-, and S-cone excitations of the background. The second direction refers to a modulation along a red–green axis; modulations in this direction leave the excitation of the S cones constant (i.e., $\Delta S = 0$), and the excitation of the L and M cones covaries as to keep their sum constant. Therefore, this axis is referred to as a "constant S-cone axis" (Kaiser and Boyton, 1996), or a "red–green isoluminant" axis (Brainard, 1996). Along the third axis, only the S cones are modulated, and $\Delta L = \Delta M = 0$. Therefore, this axis is often referred to as a "constant L & M cone" axis (Kaiser and Boyton, 1996), or as an "S-cone isoluminant" axis (Brainard, 1996) or as a "tritanopic confusion line."

Instead of defining the chromatic properties of a stimulus by their respective L-, M-, and S-cone modulations, the stimuli are often defined in terms of the responses of a set of hypothesized post-receptoral mechanisms that are isolated by these cardinal color modulations (Derrington et al., 1984; Brainard, 1996; Eskew et al., 1999; Wuerger et al., 2002, 2011). The three corresponding mechanisms are two cone-opponent color mechanisms and a luminance mechanism (see **Figure 1A**). One of the two cone-opponent mechanisms is a reddish–greenish mechanism that takes the weighted difference between the differential L- and the M-cone excitations. The second cone-opponent mechanism is a lime-violet mechanism that takes the weighted difference between the differential S-cone and the summed differential L- and M-cone excitations. The luminance mechanism sums the weighted differential L- and M-cone signals. These orthogonal mechanisms are often referred to as "L + M", "L − M", "S − (L + M)" (Derrington et al., 1984). For simplicity, we will define the chromatic properties of our stimuli in terms of their L,M,S cone excitations, that is, the achromatic direction as "L+M"; the reddish-greenish direction as "L−M," and the lime-violet direction as "S."

In the eye movement experiment, the CIE coordinates of the gray background were $x = 0.278$, $y = 0.298$ and Lum = 42.52 cd/m^2. The endpoints of the L-M and the S directions were defined by the available monitor gamut, but constrained to be symmetric around the gray background. In terms of cone contrast, stimuli at the endpoints of the S direction were defined as follows: S increments had L and M cone contrasts 0.0 and an S- cone contrast of 0.69 while S decrements had contrasts of 0.0 for both L and M cones and −0.68 for S-cones. Increments and decrements along the L − M direction resulted in an average cone contrast in the L and M cones of 0.16 and −0.16, respectively, and 0.0 for S cone contrast.

In the EEG experiment, the CIE coordinates of the gray background were $x = 0.296$, $y = 0.309$ and Lum = 46.3 cd/m^2. At the edge of the monitor's gamut, positive modulations along the S direction resulted in L and M cone contrasts of 0.0 and S-cone contrast of 0.89, while a negative excursion along the S direction resulted in zero contrasts for L and M cones and cone contrast of −0.89 for S-cones. The maximum incremental and decremental modulations along the L-M axis (within the available gamut) were as follows: 0.20 and -0.21 for the average LM cone contrast, and 0.0 for S cone contrast.

STIMULI

Stimuli were taken from existing stimulus sets that contain line drawings of common objects (International Picture Naming Project with 525 pictures, Bates et al., 2003; 400 pictures from a French-language naming study, Alario and Ferrand, 1999; 152 images used in object recognition studies, Hamm and McMullen, 1998). A set of 225 objects was selected for use in the baseline threshold experiment and 168 objects were selected for use in the main classification experiment. All images represented simple, common objects from various semantic categories (for

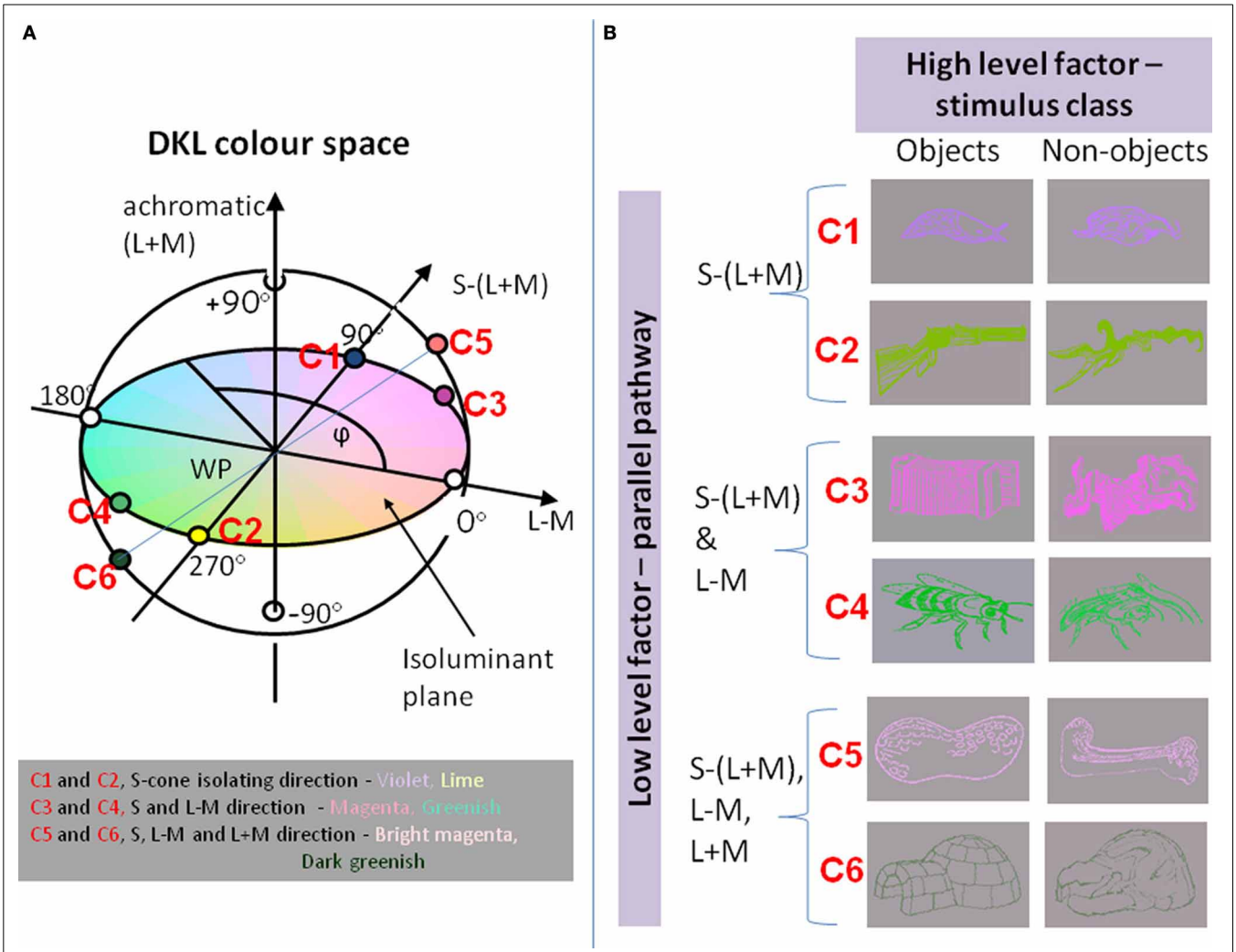


FIGURE 1 | (A) The chromaticities of stimuli in the DKL color space. Along the achromatic axis, cone contrasts in all three cone classes vary (L + M + S). Along the L − M axis, only the difference between L- and M-cone varies, keeping L + M constant. Along the S − (L + M) axis, the difference between S cones and the sum of L and M cones varies. Colors along the S-cone-isolating line range from violet to lime; intermediate isoluminant colors range from magenta to greenish; addition of an achromatic component to the magenta and greenish stimuli results in bright magenta to dark greenish. **(B)** *Examples of stimuli*: objects and non-objects, represented in colors that excite different directions in color space.

example, ship, stapler, harmonica, grasshopper, etc.; see Appendix A in (Martinovic et al., 2011) for a detailed list). Non-objects were produced by manipulating images of objects using the image distorting functions of the freely-distributed GNU Image Manipulation Programme (GIMP). After scrambling, we checked whether the resulting image adequately approximated the aspect ratio of the object it was derived from and whether it maintained the closed line structure that characterizes real objects. If not, it was edited by hand to better approximate these characteristics. Afterwards the images were converted to JPEGs and their file sizes compared. JPEG file size provides an objective estimate of visual complexity for line drawings that has been used in picture naming studies (Szekely and Bates, 2000), including the normative set provided by Bates et al. (2003). Where big discrepancies in size were present, the larger of the images were edited by hand to reduce the number of inner contours while maintaining an object-like structure. In the final stimulus set, there were no differences in visual complexity between objects and non-objects [$t_{(167)} = 1.63$, n.s.]. We also assessed low-level differences in object and non-object images by running a permutation analysis of their Fourier spectra. This analysis, using 1000 permutations, revealed that although images of objects contained more cardinally oriented lines than images of non-objects, these differences were not significant.

In the experiments, object and non-object contours were defined along three directions in DKL color space: (1) S-cone-isolating [S − (L + M)], or (2) intermediate isoluminant [S − (L + M) and L − M], or 3) a full-color stimulus with an additional achromatic component [S − (L + M); L − M; L + M], providing a luminance signal (see **Figure 1**). For each direction, both increments and decrements were used in order to obtain a signal that was representative for the whole direction (see

Figure 1; data were collapsed across increments and decrements in the final analysis, as they did not differ significantly between each other). Thus, the stimuli either involved processing predominantly in the koniocellular pathway (S-cone-isolating contours), in both pathways capable of chromatic processing (konio- and parvocellular), or in all three visual pathways (full color images including chromatic and achromatic information: konio-, parvo- and magnocellular). The majority of the stimuli subtended a visual angle of approx. 5° × 2° (the smallest stimulus was around 3° × 1°; the biggest stimulus was around 9° × 3.5°) and were shown on a gray background. Stimulus onset was synchronized to the vertical retrace of the monitor. Stimulus presentation was balanced across the sample to control for item-specific effects: thus, across the sample, each item was presented equally often with contours defined along each of the three directions of the DKL color space.

Static random luminance noise was superimposed over the stimulus display area in the form of 3 × 3 pixel elements modulated at an RMS noise contrast of 19.5% (Ruppertsberg et al., 2003). The noise was added to each trial starting with the fixation cross preceding the stimulus presentation. The purpose of the noise was to reduce luminance-related artifactual activity which would be inevitable for isoluminant stimuli with high-frequency edges. In Martinovic et al. (2011) the same approach was used and both behavioral and ERP findings were not consistent with a luminance artifact account.

OBSERVER ISOLUMINANCE

Individual differences in luminous efficiency may result in a small luminance artifact in the nominally isoluminant L-M signal (Wyszecki and Stiles, 2000). To control for this, prior to the experiment heterochromatic flicker photometry (HCFP; Walsh, 1958) was used to adjust the point of isoluminance for each participant.

The display alternated between two polarities of a chromatic stimulus (bluish/yellowish, magenta/greenish) at a frequency of 20 Hz. The participants adjusted the luminance of the colored stimuli in order to find a point at which the flicker was minimized. The rationale for this technique is that the chromatic system is too slow to follow fast temporal changes (flickering), while the luminance system is able to detect fast changing luminance differences. Therefore, if the perception of flicker is minimal, the difference in luminance is also minimized. Objects from the 225 threshold item set were randomly chosen as stimuli during HCFP. The procedure was repeated ten times. The lowest and highest values were then eliminated, and the mean of the remaining values taken.

PROCEDURE

Baseline experiment: threshold measurements

An initial session consisting of control measurements (Cambridge Colour Test and Heterochromatic Flicker Photometry) and the baseline psychophysical experiment was conducted with each participant, lasting one and a half hours in the eye movement experiment and 2 h in the EEG experiment.

The baseline experiment was conducted to define a common contrast metric for chromatic and luminance stimuli, as comparing responses to isoluminant and achromatic stimuli is not straightforward (Shevell and Kingdom, 2008). This difficulty can be overcome by matching the stimuli in terms of threshold units, thereby using a behavioral measure that is independent of the actual physical contrast. Such stimuli can be then used to address specific research questions regarding the role of chromatic and luminance signals in the human visual system. We took measurements of object discrimination contrast thresholds prior to the main experiment. The task required discrimination of object and non-object images taken from the same stimulus pool and was thus closely matched to the task in the main classification experiment. The reason behind this was to attempt to match effective stimulus strength (i.e., salience) for the object classification task as closely as possible. For this, a discrimination threshold with a similar task and with stimuli of similar spatio-temporal properties is much more suitable than a detection threshold, a contrast-matching threshold or a less similar discrimination threshold procedure. Cole et al. (1993) discuss the differences in neuronal populations involved in stimulus detection and in the processing of stimuli above detection threshold, with stimuli above detection threshold being encoded by a significantly larger pool of units. Zele et al. (2007) and Vassilev et al. (2009) discuss more extensively the suitability of detection threshold units for equating stimuli in terms of reaction times for rod and cone stimuli respectively.

Stimuli in the main experiment were matched in discrimination threshold units individually for each participant so that maximum possible contrast was achieved within the available gamut. This procedure was intended to ensure that any differences that emerge at suprathreshold cannot be accounted for by simple stimulus salience differences between different directions in color space. For example, a simple effect of salience would result in performance between directions in color space differing uniformly for both objects and non-objects. This was not observed in the previous study by Martinovic et al. (2011), as accuracies for non-objects remained similar across the low-level conditions, while accuracies for objects were significantly lower in the S-cone isolating condition. Due to the properties of the S − (L + M) mechanism, reductions in performance for S-cone isolating stimuli are to be expected even when attempts are made to closely match stimuli in terms of contrast (for a discussion, see O'Donell et al., 2010).

Stimulus contrast in the main experiment was adjusted toward the maximal monitor's gamut relative to discrimination thresholds in order to ensure that all stimuli were as high in contrast as possible while remaining approximately iso-salient for each individual participant. This was achieved by using multiple-of-threshold contrasts within the monitor gamut where the scaling factor was the same in all color directions. The following procedure was used to scale the stimuli: DKL radius in the direction in which the threshold was closest to the monitor's gamut was set to the value just below gamut and all the other contrasts were adjusted upwards from threshold using the scale factor calculated on the basis of this, closest-to-gamut direction. This procedure was intended to allow for an adequate signal-to-noise ratio in the EEG while maintaining equal salience along different color directions. It also allowed us to assess if behavioral measures, saccades, eGBA and tGBA relate to contrast, as different contrast level (in terms of multiple-of-threshold) was used for each participant.

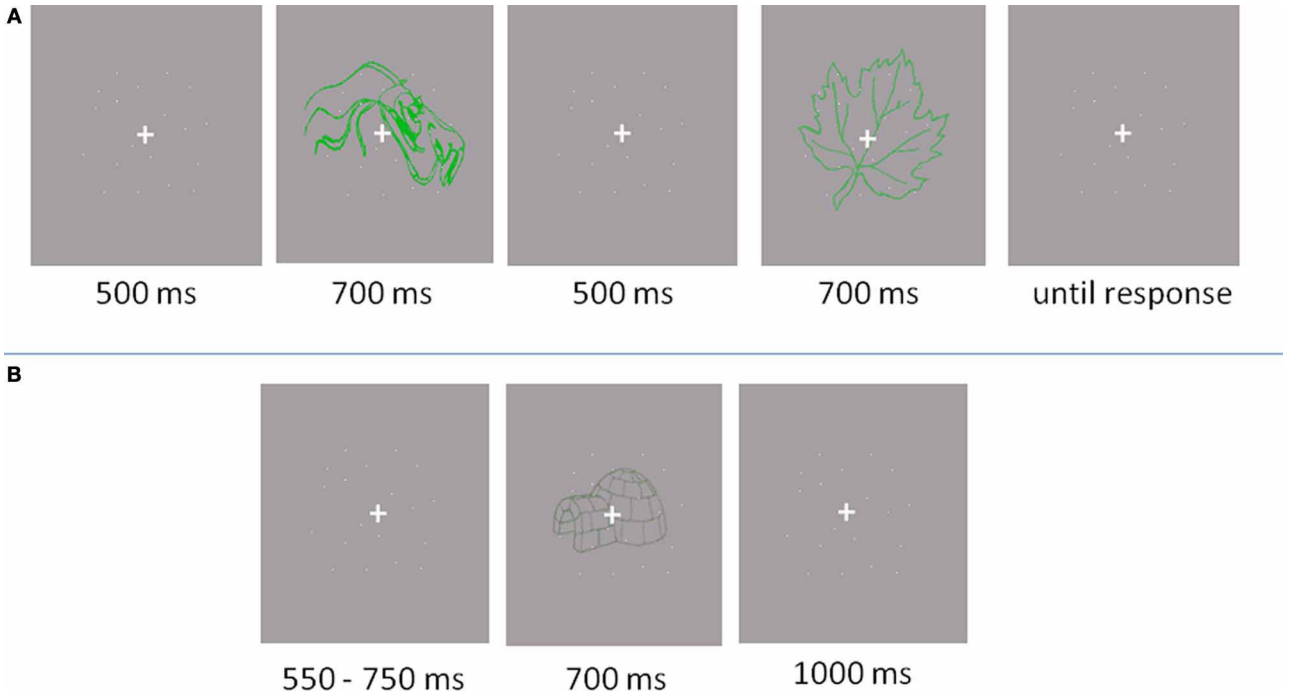


FIGURE 2 | Trail outlooks. (A) Baseline experiment: Participants responded if the object was located in the first or the second interval. **(B)** Main experiment: Participants responded whether the presented item was an object or a non-object.

In the baseline experiment, a two-interval forced choice paradigm (2IFC) was implemented (see **Figure 2A**). A fixation cross (0.46 by 0.46° of visual angle) appeared in the centre of the screen for 500 ms, followed by the first item displayed for 700 ms. Subsequently, another fixation cross appeared for 500 ms, followed by the second item for another 700 ms. After the second item, participants indicated by pressing a button which of the two items represented an object. The next trial started after the response. Participants were told to give a correct answer, rather than a fast answer. Acoustic feedback was provided, indicating incorrect responses with a beep.

The participant's responses guided an adaptive QUEST procedure that controlled stimulus contrast (Watson and Pelli, 1983). To estimate the color contrast threshold from the relative frequency of a correct response, defined as the 81% correct point on the psychometric function, a Weibull function was fitted. In the EEG experiment, thresholds in each of the tested directions (S-cone isolating, intermediate isoluminant, full color) were measured three times for every participant; in the eye movement experiment, chromatic thresholds were measured three times while a luminance threshold was measured once and then combined with a fixed-contrast, intermediate isoluminant signal prior to scaling (see **Figures 3**, **4** for more detail) to create a full-colour stimulus. Differences between increment and decrement thresholds were assessed using paired *t*-tests.

The main experiments: EEG and eye movements

The main experiment was conducted in a separate session and lasted one and a half hours for eye movement recording and one hour for the EEG recording (see **Figure 2B**). First, a practice block of 20 trials was performed. The items used in the practice were not used in experimental trials. Participants were required to discriminate between drawings of familiar, nameable objects and unfamiliar, unnamable objects (non-objects). Participants were instructed to fixate the cross throughout the experiment and not to scan the presented images with their eyes. In the EEG experiment, there were four 84 trial blocks while in the eye movement experiment there were 12 blocks of 28 trials (336 trails in total). A trial started with a variable baseline period (550–750 ms) of fixation. The stimulus was then displayed for 700 ms, followed by a fixation cross displayed for 1000 ms. The participants were required to indicate whether the presented item belonged to an object or non-object category by pressing a button. Button-to-response allocation was balanced across participants. After each trial, an "X" appeared on the screen for 900 ms. The participants were advised to refrain from blinking unless the "X" was displayed.

BEHAVIORAL DATA ANALYSIS

A few thresholds in the eye movement experiment were typed incorrectly into the script that computed the scale factors: for participant 2, these were the magenta and luminance decrement thresholds, for participant 4 the lime threshold and for participant 5 the lime and luminance increment thresholds. These data were left out in all subsequent analyses (behavioral and saccade rate).

The accuracies and RTs from the main experiment were analyzed. Only correct trials with RTs between 300 and 1700 ms

(the maximum time allowed for responses) were used in further analyses. Median RTs for correct items were computed for each participant. Differences in accuracies and RTs between the conditions were analyzed with a 3 × 2 × 2 mixed ANOVA with the within-subject factors *direction in color space* (S- cone isolating, intermediate isoluminant, full color) and *object class* (object, non- object) and a between-subject factor of *experiment* (EEG or eye movement). Greenhouse-Geisser correction was used when necessary. *Post-hoc* paired *t*-tests with Bonferroni correction for multiple comparisons were used. Bonferroni-corrected *p*-values were adjusted by multiplying the *p* value with the number of comparisons in order to make it easier to compare them with classically used significance levels (0.05, 0.01, 0.005, 0.001).

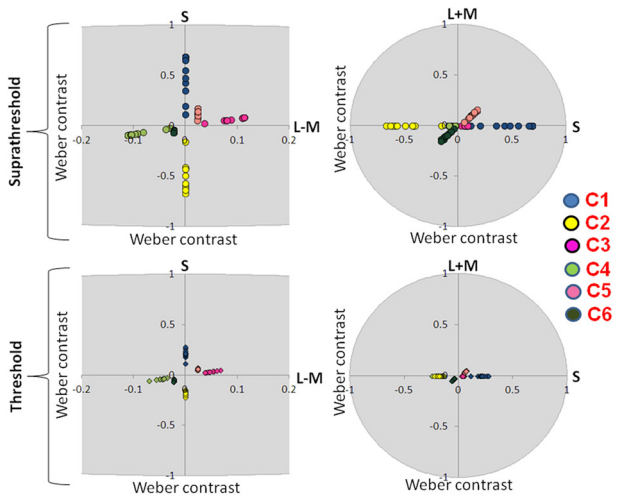


FIGURE 3 | Suprathreshold and threshold contrasts for the eye movement experiment. Left side of the figure shows chromatic contrasts (S and L − M) while right side of the figure shows the luminance contrast in relation to S-cone contrast (S and L + M). Contrasts for each participant are represented with a single dot. C1: S-cone increment; C2: S-cone decrement; C3: intermediate isoluminant increment; C4: intermediate isoluminant decrement; C5: full-colour increment; C6: full-colour decrement.

EYE MOVEMENT RECORDING AND ANALYSIS

Recordings were performed at a sampling rate of 500 Hz. The Eyelink camera was placed on the desktop below the monitor. Participants had their head stabilized with a chin rest. The system was calibrated using the Eyelink's inbuilt 9-point calibration system. Calibration was performed at the start of the experiment and repeated between blocks if the in-built calibration check indicated that this was necessary.

Eye movements were analyzed for all correct trials using custom scripts for Matlab. Trials with saccades already detected by the Eyelink algorithm were not discarded in light of Mergenthaler and Engbert's (2010) findings; we wanted to capture not only the miniature saccades but also the somewhat larger inspection saccades. Data were segmented into epochs that included the time 500 ms before and 1500 ms after stimulus onset. Miniature saccades were detected using the Engbert and Kliegl (2003) algorithm (accessible at http://www.agnld.uni-potsdam.de/~ralf/MS). Only binocular movements were taken into further analysis. To test if the saccades in the stimulus display period (0–700 ms

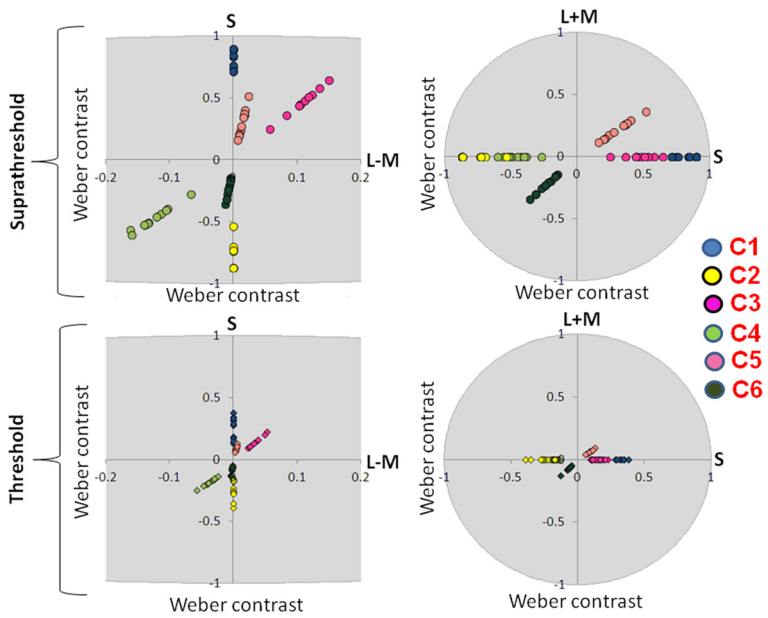


FIGURE 4 | Suprathreshold and threshold contrasts for the EEG experiment. Left side of the figure shows chromatic contrasts (S and L − M) while right side of the figure shows the luminance contrast in relation to S-cone contrast (S and L + M). Contrasts for each participant are represented with a single dot. C1: S-cone increment; C2: S-cone decrement; C3: intermediate isoluminant increment; C4: intermediate isoluminant decrement; C5: full-colour increment; C6: full-colour decrement.

after stimulus onset) revealed a bimodal amplitude distribution which was found in the free-viewing study by Mergenthaler and Engbert (2010) we conducted Hartigan's unimodality test (Hartigan and Hartigan, 1985). Saccade frequencies were compared in the time window between 200 ms and 700 ms after stimulus onset for trials with correct responses. This is the time window in which the tGBA was also analyzed (see below).

Differences in fixational saccade rates between conditions were analyzed with a 3 × 2 repeated measures ANOVA with the factors *direction in color space* (S- cone isolating, intermediate isoluminant, full color) and *object class* (object, non-object). Greenhouse-Geisser correction was used when necessary. *Post-hoc* tests were performed using paired *t*-tests, with Bonferroni correction for multiple comparisons.

EEG DATA ACQUISITION AND ANALYSIS

In the EEG experiment, continuous EEG was recorded from 32 locations using active Ag–AgCl electrodes (Biosemi ActiveTwo amplifier system) placed in an elastic cap. Standard locations of the international 10–20 system (Jasper, 1958) were used. In the Biosemi system the typically used "ground" electrodes in other EEG amplifiers are replaced through the use of two additional active electrodes. In the 32-electrode montage these electrodes are positioned in close proximity to the electrode Cz of the international 10–20 system: Common Mode Sense (CMS) acts as a recording reference and Driven Right Leg (DRL) serves as ground (Metting Van Rijn et al., 1990, 1991). Vertical and horizontal electrooculograms were recorded in order to exclude trials with large eye movements and blinks. EEG data processing was performed using the EEGlab toolbox (Delorme and Makeig, 2004) combined with self-written procedures running under Matlab. EEG signal was sampled at a rate of 512 Hz and epochs lasting 2000 ms were extracted, starting from 500 ms before stimulus onset and incorporating the 1500 ms after stimulus presentation. Removal of epochs with artifacts was performed using the FASTER (Fully Automated Statistical Thresholding for EEG artifact Rejection) plug-in for EEGlab (Nolan et al., 2010). The average rejection rate for artifact-contaminated trials was 22%. Trials with incorrect responses were excluded from the analysis. This left an average of 44 trials per condition. While FASTER-based artifact rejection was performed with Fz as reference, all other procedures were performed using the average reference.

The saccadic artifact was removed from the EEG using the procedure established by Keren et al. (2010). These authors derived a saccadic potential filter on the basis of data from five participants who performed an object/non-object classification task while eye movements and EEG were co-recorded. Based on Keren et al.'s (2010) suggested procedure, the eye channels were combined into a single channel referenced to the electrode Pz (radial EOG; rEOG) and data were convolved with the saccadic filter. Local peaks greater than 3.5 times the root mean square of the rEOG were identified as saccades. This threshold was selected because it produced the most similar distribution of saccades from EEG data to that observed in the actual eye movement experiment (see **Figure 7A**). Epochs lasting 100 ms before and after each miniature saccade were cut out. This resulted in datasets with an average of 275 epochs. Independent component analysis (ICA) was performed on these datasets using EEGlab' s extended infomax algorithm (Lee et al., 1999). High-density EEG data can be considered to represent linear mixtures of activity from multiple independent generators, so ICA is intended to "unmix" them into minimally dependent source signals. When conducted on artifact-free data, ICA can reveal specific aspects of neural activity (e.g., occipital alpha-band sources; Makeig et al., 2004). It is more often used to remove ocular or muscular artifacts from EEG data since such artifacts are considered to be independent from neurally-generated activity (for a review focused on microsaccadic artifacts, see Schwartzman and Kranczioch, 2011). The major components resulting from an ICA on peri-saccadic epochs are thus likely to be those originating in the spike potential artifact. These ICAs were copied over to the complete datasets for each participant. Components that reflected typical fixational saccade activity patterns (see Keren et al., 2010) were subtracted. This resulted in a subtraction of 3 components on average (range: 0–7). Subsequently, FASTER was used again, to interpolate globally and locally contaminated channels.

Oscillatory activity in the gamma band (30–120 Hz in 4 Hz steps) was estimated using multitapers (Mitra and Pesaran, 1999) as implemented in the Fieldtrip toolbox for Matlab (Oostenveld et al., 2011). We used a fixed time window of 250 ms moved in 20 ms steps and 5 orthogonal Slepian tapers yielding a frequency smoothing of ∼12 Hz. This method gives a time-varying magnitude of the signal in each frequency band leading to a time-by-frequency (TF) representation of the signal. We verified if the artifactual ocular activity was successfully removed by inspecting the time-frequency plots at all electrodes to see if the tGBA activity at frontal and eye channels was close to baseline. This led to the removal of 6 participants, with 11 participants remaining in the sample. Total GBA was analyzed in the 200–700 ms window. In order to identify the electrodes, time window and frequency range of the tGBA, mean baseline-corrected spectral activity (baseline: 200 ms prior to stimulus onset) was collapsed for all conditions together and represented in TF-plots in the 30–120 Hz range for all electrodes. Electrode sites were then selected on the basis of grand mean topographies, with maximal activity in artifact-corrected data expected at posterior sites (Keren et al., 2010; Hassler et al., 2011). Due to inter-individual differences in the induced gamma peak in the frequency domain, a maximal frequency for each participant was chosen on the basis of an average across the conditions. We used a frequency band of ±4 Hz around this peak frequency for statistical analysis. Differences in tGBA between conditions were analyzed with a 3 × 2 repeated measures ANOVA with factors *direction in color space* (S- cone isolating, intermediate isoluminant, full color) and *object class* (object, non-object). Greenhouse-Geisser correction was used when necessary. *Post-hoc* tests were performed using paired t-tests, with Bonferroni correction for multiple comparisons.

RESULTS

PSYCHOPHYSICS: THRESHOLD MEASUREMENTS

Figure 3 presents scaled, suprathreshold contrasts as well as contrasts at threshold for the eye movement experiment, while **Figure 4** presents these contrasts for the EEG experiment. On the left side, contrasts are plotted in the isoluminant plane

(S vs. L − M); on the right side, the y-axis is the achromatic axis (L + M) and the x-axis the S-cone axis.

The scale factors in the EEG experiment ranged from 2.24 to 5.56, with the average factor being 3.46. The scale factors in the eye experiment ranged from 0.85 to 3.23, with the average factor being 2.20. These scale factors reflect the ratio of the contrast used in the experiment to that participant's threshold. The scale factors were significantly larger in the EEG experiment [$t_{(16.29)} = 3.08$, $p = 0.007$].

There were no significant differences between the threshold contrasts for increments and decrements [S − (L + M): $t_{(20)} = 0.79$, $p = 0.44$; S − (L + M) & L − M: $t_{(21)} = -1.58$, $p = 0.13$; S − (L + M) & L − M & L + M: $t_{(20)} = -0.22$, $p = 0.83$). This justified the collapsing of data across increments and decrements.

BEHAVIORAL DATA: ACCURACY AND REACTION TIMES

Figure 5A shows the accuracies while **Figure 5B** shows reaction times. The data was analyzed with a mixed ANOVA, as described in the behavioral data analysis section.

In accuracy, there was no overall difference between classifying objects and non-objects [$F_{(1,\ 21)} = 0.06$, $p = 0.81$], but there was an interaction with experiment [$F_{(1,\ 21)} = 6.80$, $p = 0.02$, $\eta_p^2 = 0.25$]. *Post-hoc* paired *t*-tests determined that while objects were classified less successfully than non-objects in the EEG experiment [$t_{(10)} = -3.15$, $p = 0.02$], classification accuracy did not differ in the eye movement experiment [$t_{(11)} = 1.37$, $p = 0.80$]. Independent sample *t*-tests showed that accuracy for both objects [$t_{(21)} = 2.70$, $p = 0.013$] and non-objects [$t_{(12.08)} = 4.39$, $p = 0.001$] was significantly better in the EEG experiment. There was also a main effect of direction in color space [$F_{(2,\ 42)} = 7.03$, $p = 0.002$, $\eta_p^2 = 0.25$], with *post-hoc* paired *t*-tests revealing worse classification of S-cone isolating stimuli than full-colour stimuli [$t_{(22)} = -5.20$, $p = 0.0001$]. On the other hand, there was no difference between the intermediate isoluminant and full-colour stimuli [$t_{(22)} = -2.07$, $p = 0.15$] and intermediate isoluminant and S-cone isolating stimuli [$t_{(22)} = 1.40$, $p = 0.54$]. This effect of direction in color space was the same for both experiments [$F_{(2,\ 42)} = 0.69$, $p = 0.51$]. Finally, there was an interaction between the two factors of object class and direction in color space [$F_{(1.59,\ 33.26)} = 9.25$, $p = 0.001$, $\eta_p^2 = 0.31$] which did not differ across experiments [$F_{(1.59,\ 33.26)} = 1.75$, $p = 0.19$]. Paired *t*-tests indicated that the differences between directions in color space were driven by superior performance for objects that did not contain solely S-cone signals [S-cone isolating objects vs. intermediate isoluminant objects: $t_{(22)} = -5.22$, $p = 0.0002$; S-cone isolating objects vs. full-colour objects: $t_{(22)} = -4.70$, $p = 0.0009$] with performance for intermediate isoluminant and full-colour objects and all non-objects being at a relatively similar level (**Figure 5A**; all *p*s > 0.1).

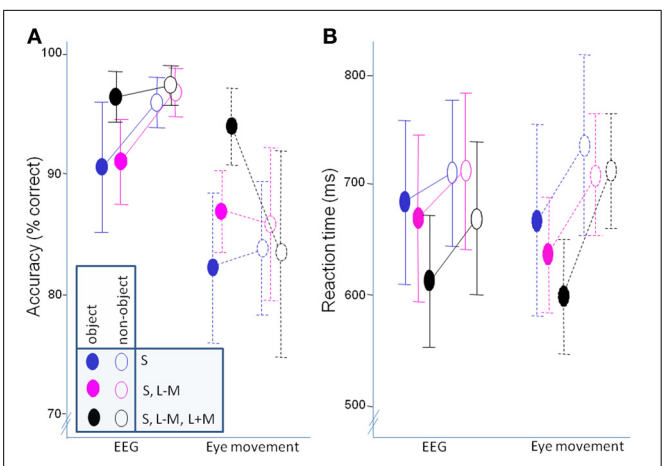


FIGURE 5 | Behavioral data. (A) accuracy; **(B)** mean of median response times. Error bars represent 95% confidence intervals.

Reaction times were faster for objects than for non-objects [$F_{(1,\ 21)} = 59.17$, $p < 0.000001$, $\eta_p^2 = 0.74$], with differences between the two experiments [$F_{(2,\ 21)} = 6.14$, $p = 0.02$, $\eta_p^2 = 0.23$]. While there were no differences between experiments in speed of responses to objects [$t_{(21)} = 0.49$, $p = 0.63$] and non-objects [$t_{(21)} - 0.52$, $p = 0.61$], the difference between the two classes seemed to be less pronounced in the EEG experiment [$t_{(10)} = 3.07$, $p = 0.05$] than in the eye movement experiment [$t_{(11)} = 9.09$, $p = 0.00001$; see **Figure 5B**]. The effect of direction in color space [$F_{(2,\ 42)} = 10.19$, $p = 0.0002$, $\eta_p^2 = 0.33$] did not differ across experiments [$F_{(2,\ 42)} = 1.11$, $p = 0.34$]. The effect was somewhat different to that observed for accuracy, as *post-hoc* tests revealed that it was the speed of classification for full-colour stimuli that was most important in driving the difference, offering an advantage both when compared to S-cone isolating [$t_{(22)} = 4.31$, $p = 0.0009$] and intermediate isoluminant stimuli [$t_{(22)} = 2.86$, $p = 0.03$]. There was no difference between the two types of isoluminant stimuli [$t_{(22)} = 1.55$, $p = 0.40$]. Finally, there was also an interaction between object class and direction in color space [$F_{(2,\ 42)} = 4.70$, $p = 0.01$, $\eta_p^2 = 0.18$] which did not differ across experiments [$F_{(2,\ 42)} = 0.62$, $p = 0.54$]. The interaction was caused by the fact that the differences in RT between directions in color space occurred for full-colour vs. intermediate isoluminant objects [$t_{(22)} = 3.51$, $p = 0.02$] and full-colour vs. S-cone isolating objects [$t_{(22)} = 4.89$, $p = 0.0006$], while the speed for intermediate isoluminant vs. S-cone isolating objects and all non-objects remained similar (*p*s > 0.1).

Additionally, a Pearson correlation analysis was performed in order to examine potential relationships between behavioral responses (accuracies and mean RTs) and contrast ratios used in the experiment. A total of 12 comparisons were made and Bonferroni correction was used to correct for multiple comparisons. There was a significant correlation between contrast ratio and accuracy for S-cone isolating non-objects [$r_{(23)} = 0.60$, $p = 0.05$] and full-colour non-objects [$r_{(23)} = 0.62$, $p = 0.02$]. Other correlations were not significant: (accuracies: *r* ranging from 0.40 to 0.47; RTs: r ranging from −0.12 to −0.32; all *p*s > 0.1).

FIXATIONAL SACCADES

As shown in **Figure 6A**, fixational saccades during picture presentation (0–700 ms after stimulus onset) included a broad range of differently-sized saccades. On the contrary, very small saccades were dominant during periods when the fixation cross was displayed. In our analysis, the fixation cross period involved 500 ms of fixation prior to the stimulus onset and 800 ms after stimulus

offset. **Figure 6B** indicates that fixational saccades during picture presentation showed a linear relation between size and speed (also known as the main sequence). Hartigan's unimodality test showed that the distribution of saccades during picture presentation was not multi-modal ($p = 0.59$). Therefore, we analyzed the frequencies of saccades in this period irrespective of their size.

Figure 7 shows the plot of fixational saccade rates across time. Fixational saccades drop substantially 100–150 ms after picture presentation, peaking from approx. 200 to 500ms. There was a main effect of object class [$F_{(1, 11)} = 4.78$, $p = 0.05$, $\eta_p^2 = 0.30$], with more fixational saccades for objects ($M = 22.17$, $SE = 5.46$) than for non-objects ($M = 18.36$, $SE = 4.49$). There was a main effect of direction in color space [$F_{(2, 22)} = 6.77$, $p = 0.005$, $\eta_p^2 =$ 0.38), indicating that fixational saccade rates differed across the three color directions, while there was no significant interaction between the factors direction in color space and objecthood [$F_{(2, 22)} = 1.84$, $p = 0.18$]. *Post-hoc* tests revealed that the difference between color directions was driven by higher saccadic rates for full stimuli ($M = 25.00$, $SE = 6.00$) than for S-cone isolating stimuli ($M = 15.29$, $SE = 3.83$; $p = 0.03$), with intermediate isoluminant stimuli ($M = 20.50$, $SE = 5.39$) not being different from full stimuli ($p = 0.13$) or from S-cone isolating stimuli ($p = 0.25$).

Again, we performed a Pearson correlation analysis to examine the relationship between behavioral responses (accuracies and mean RTs), contrast ratios, and rates of fixational saccades in the period between 200 and 700 ms. A total of 18 comparisons were made and Bonferroni correction was used to correct for multiple comparisons. No significant correlations were found: (accuracies: r ranging from 0.03 to 0.43; RTs: r ranging from −0.55 to −0.35; contrast ratios: $r =$ ranging from 0.32 to 0.63; all ps > 0.1).

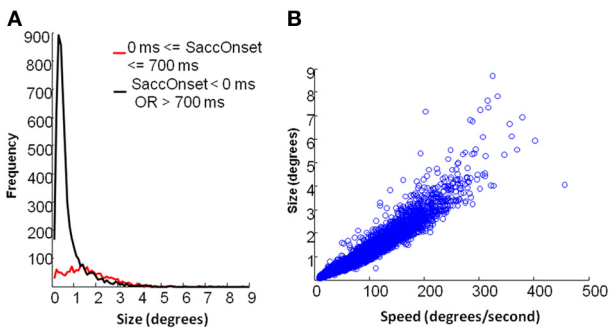


FIGURE 6 | Eye movement data: saccade properties. (A) Distribution of saccades by size. Black line depicts saccades during fixation cross and red line depicts saccades during picture presentation. **(B)** main sequence relation between speed and size of saccades for the period of picture presentation.

GAMMA-BAND ACTIVITY

Successful removal of miniature saccade artifacts using the saccadic potential filter (Keren et al., 2010) was possible in 11 out of 17 participants. Visual inspection revealed that the remaining 6 participants still had relatively high tGBA at ocular and frontal channels after artifact removal. The relatively low efficiency of artifact removal could be due to the reduced rate of fixational saccades (see fixational saccade results) in our study when compared to Yuval-Greenberg et al. (2008) and Keren et al. (2010). A lower saccade rate reduces the amount of data that is fed into the ICA which adversely impacts the quality of the artifact removal. In our eye movement experiment, the number of

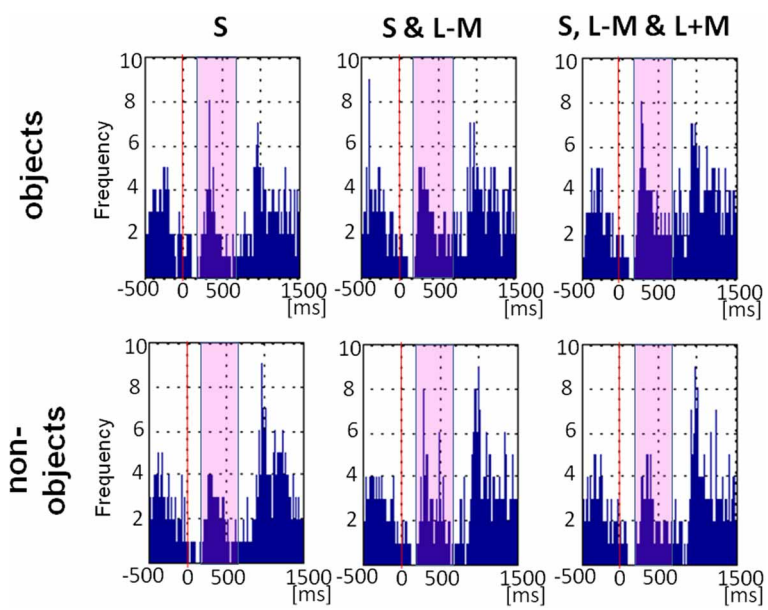


FIGURE 7 | Eye movement data: saccade rates across time. Frequency plot of all fixational saccades in the period including −500 ms before picture onset and 1500 ms after picture onset. Solid red line indicates stimulus onset and the magenta rectangle highlights the period 200–700 ms post-stimulus which was the main focus of our analysis.

fixational saccades was found to vary vastly between participants, with 6 out of 12 participants having a total of 80 or less fixational saccades during the 200–700 ms period after picture presentation while other participants had between 123 and 291 saccades in this period (large individual differences in fixational saccade rates were also reported by Makin et al., 2011). The number of participants with relatively low saccade rates approximately corresponds to the number of participants in the EEG study (6 out of 18) in whom artifact removal was not successful. An independent t-test revealed that the number of 'saccades' detected with the saccadic potential filter was lower in the 6 rejected participants ($M_{\text{reject}} = 262$, $SD_{\text{reject}} = 25$; $M_{\text{sample}} = 290$, $SD_{\text{sample}} = 28$; $t_{(15)} = 2.09$, $p = 0.05$), indicating that it could indeed be that lower saccade rates in those participants may have led to an artifact which could not be effectively removed with the ICA procedure. It is important to note that the one participant in whom there were no components that appeared to correspond to the known topographical and temporal properties of the artifact was not removed from the sample, since tGBA did not show the typical artificial pattern. Therefore, we assume that he maintained fixation successfully, while the rejected participants probably made fewer and/or smaller fixational saccades that did not allow their proper identification with the Keren et al. (2010) method.

Figure 8A shows the grand-mean time-course of the eGBA at posterior electrodes, **Figure 8B** shows the topography and **Figure 8C** shows the relative change in signal power from baseline in the analyzed time-frequency window. There was no significant effect of object class on eGBA relative power [$F_{(1,\ 10)} = 2.76$, $p = 0.1$). There was a significant effect of direction in color space [$F_{(2,\ 20)} = 5.00$, $p = 0.02$, $\eta_p^2 = 0.33$). *Post-hoc* t-tests showed that the eGBA relative power was significantly lower for intermediate isoluminant stimuli than for full-colour stimuli ($p = 0.04$); no other comparisons were significant (all p-values > 0.1). There was no interaction between object class and direction in color space [$F_{(2,\ 20)} = 1.55$, $p = 0.2$]. Evoked GBA was significant compared to baseline only in the S-cone isolating non-object condition ($p = 0.002$; all other ps > 0.1).

Figure 9A shows the grand-mean time-course of the tGBA at posterior electrodes, **Figure 9B** shows the topography while **Figure 9C** shows the relative change in power from baseline in the analyzed time-frequency window. There was no significant effect of object class [$F_{(1,\ 10)} = 1.38$, $p = 0.3$] or direction in color space [$F_{(2,\ 20)} = 0.11$, $p = 0.9$] on tGBA relative power. There was a significant interaction between object class and direction in color space [$F_{(2,\ 20)} = 3.77$, $p = 0.04$, $\eta_p^2 = 0.27$]. While *post-hoc* comparisons were not significant, it would appear from the graph (**Figure 9C**) that relative power is higher for intermediate isoluminant objects than for intermediate isoluminant non-objects, while for S-cone isolating and full-colour stimuli the relative powers are roughly similar for objects and non-objects. Total GBA was significant compared to baseline in the S-cone isolating non-object condition ($p = 0.006$) and the intermediate isoluminant object condition ($p = 0.01$), with a trend toward significance for the full-colour object condition ($p = 0.06$; all other ps > 0.1).

As before, we performed a Pearson correlation analysis in order to establish whether there are relations between behavioral responses (accuracies and mean RTs) and contrast ratios used in the experiment, on one hand, and tGBA in the period between 200 and 700 ms, on the other hand. As a total of 18 comparisons were made, Bonferroni correction was used. There was a trend for total GBA for intermediate isoluminant non-objects to correlate with speed of responding to these non-objects [$r_{(11)} = -0.79$, $p = 0.07$]. Other correlations were not significant: (accuracies: r ranging from -0.37 to 0.77; RTs: r ranging from −0.48 to 0.71; contrast ratios: r ranging from −0.21 to 0.34; all ps > 0.1).

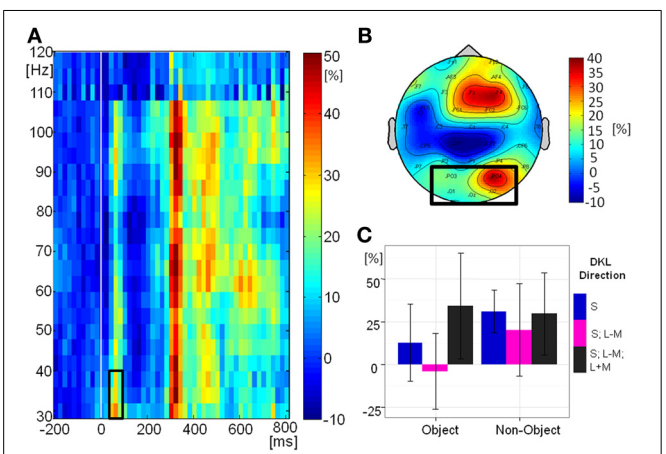


FIGURE 8 | Evoked GBA. (A) Grand mean baseline-corrected TF-plot averaged at the regional mean sites (see **panel B**) across all conditions. Box indicates the time window for statistical analysis. **(B)** Grand mean amplitude-map (average across all conditions) for activity within the black box in Panel **A**). Box indicates electrode sites included in the regional mean. **(C)** Bar plot of amplitudes of evoked GBA for each condition at the regional mean during the selected time window, with 95% confidence interval bars.

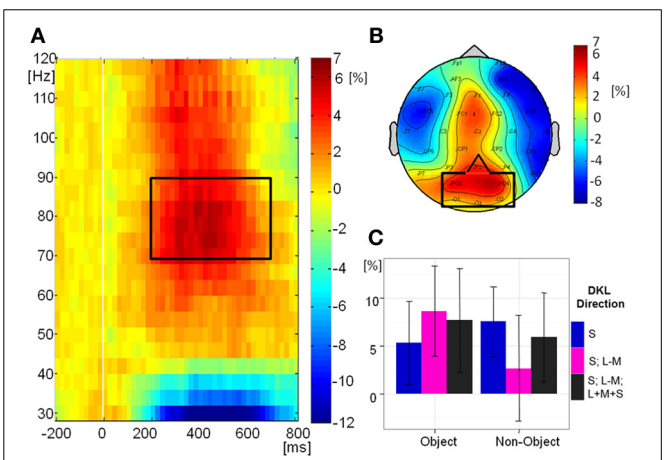


FIGURE 9 | Total GBA. (A) Grand mean baseline-corrected TF-plot averaged at the regional mean sites (see **panel B**) across all conditions. Box indicates the time window for statistical analysis. **(B)** Grand mean amplitude-map (average across all conditions) for activity within the black box in **panel A**). Box indicates electrode sites included in the regional mean. **(C)** Bar plot of amplitudes of total GBA for each condition at the regional mean during the selected time window, with 95% confidence interval bars.

The same type of analysis was performed for eGBA but no significant correlations were found (accuracies: r ranging from −0.59 to 0.73; RTs: r ranging from −0.54 to −0.02; contrast ratios: r ranging from −0.53 to 0.47; all ps > 0.1).

DISCUSSION

We investigated modulations of behavioral responses, fixational saccades and gamma-band activity by low- and high-level factors in an object classification task. Stimuli were defined along different directions in cardinal color space so that they differentially excited distinct post-receptoral mechanisms, with contrasts matched in terms of discrimination thresholds. This provided a controlled low-level manipulation, while stimulus class (object or non-object) provided a high-level manipulation. In both the eye movement and the EEG experiments, behavioral performance was the fastest for full-colour objects and least accurate for S-cone isolating objects, with performance for non-objects remaining similar across all directions in color space. The stimulus contrasts were somewhat higher in the EEG experiment, but in the analysis, the experiment factor only interacted with object class, with an accuracy advantage for classifying objects in the EEG experiment but not in the eye movement experiment, and a less pronounced reaction time advantage for objects in the EEG experiment. Performance for S-cone isolating and full-colour non-objects was also correlated with contrast. Therefore, lower contrast seems to have a more adverse effect on performance for non-objects. Fixational saccade rates 200–700 ms after stimulus onset depended on low and high-level factors independently, being higher for full-colour stimuli and for objects. Evoked GBA was fairly low and its amplitude was modulated by low-level factors only. In contrast, artifact-free, low-amplitude sustained tGBA that lasted approximately 200–700 ms was dependent on both low and high-level factors.

The behavioral results extend the pattern from the previously conducted EEG experiment (Martinovic et al., 2011): performance for objects differs across the directions in color space, while performance for non-objects remains steady. Differences between the two experiments were observed only in terms of responses to stimulus class, with performance in the EEG experiment being more accurate overall, with less pronounced differences between the two stimulus classes in terms of reaction times. The most substantial difference between the two experiments was in terms of maximal achievable contrasts, which resulted in significantly higher contrast ratios in the EEG experiment. As accuracy was related to contrast ratios for two out of three non-object conditions, this would imply that non-object performance is more driven by contrast. This finding emphasizes the importance of low-level signals in driving task performance: although the contrasts were set to various multiple-of-threshold levels, these levels may have been close enough to threshold to still enact an influence on accuracy rates. Ceiling effects that are commonly observed in object classification experiments (e.g., Gruber and Müller, 2005; Busch et al., 2006) were not reached, except perhaps for the full combination objects and non-objects in the EEG experiment.

Mergenthaler and Engbert (2010) demonstrated that in a free viewing task saccades are distributed bimodally, with those below 0.4° less numerous and predominantly around 0.1° in size, and those above 0.4° much more numerous and mostly around 10° in size (their stimulus was presented full screen). On the contrary, in their fixational task, saccades were distributed unimodally with a peak around 0.5° and the vast majority of saccades smaller than 1°. In our study, fixational saccades observed before and after stimulus presentation match the distribution of saccades in Mergenthaler and Engbert's fixation task. However, we find that saccades during picture presentation contained a significant proportion of larger saccades (>1°) when compared to saccades made during periods when only the fixation cross was presented. We did not observe a bi-modal distribution. In fact, with saccades over 1° prominent in our data, it could be that an onset of a complex stimulus within the fixation area preferentially elicits inspection saccades and perhaps even voluntary, exploratory saccades. This suggestion is in line with a recent study by Otero-Millan et al. (2013), which suggests that fixation and exploration behaviors are not in fact different, opposing phenomena, but can rather be placed on the extremes of the same continuum. In their study, Otero-Milan et al. presented observers with scenes of varying sizes and found that as the scenes decreased in size, so did the size of produced saccades. Otero-Milan et al. report that in a free-viewing task the saccade magnitude distribution ranged from 0.1 to 10 deg for stimuli sized between 4 and 8 deg in width, with less saccades for the blank scenes than for natural scenes. In line with this finding, it is perhaps not surprising to observe more inspection saccades in our experiment, as participants are asked to classify images containing relatively low-contrast, task-relevant visual content—however, this suggestion warrants further investigation.

Otero-Millan et al. (2008) demonstrated that high-level modulations of microsaccades can occur. In our study, fixational saccade rates 200–700 ms after stimulus onset were enhanced for objects as opposed to non-objects, in line with Hassler et al. (2011) and Yuval-Greenberg et al. (2008). Modulations of microsaccades by low-level factors observed in our experiment extend previous findings. Valsecchi and Turatto (2007) demonstrated that the characteristic microsaccadic signature rate was observable for isoluminant red-green stimuli and did not differ significantly from the saccades elicited by stimuli defined with a further luminance component. If the superior colliculus is "color blind", as Marrocco and Li's (1977) findings are often taken to suggest, then Valsecchi and Turatto's (2007) results suggest that cortical areas responsive to color are involved in microsaccades. Here we demonstrate that S-cone isolating contours result in fewer fixational saccades compared to full-colour stimuli, without finding a significant difference for the intermediate isoluminant stimulus. While S-cones do not project directly to the superior layers of the superior colliculus, S-cone elicited neural responses have been reported to be as fast as L − M elicited responses at the level of its intermediate layers (White et al., 2009), indicating cortico-tectal loops of similar timing (but see also Tailby et al., 2012). Fixational behavior is related to foveating the target of interest, and our findings support the suggestion that fixational saccades are highly related to the acquisition of fine spatial details during foveal processing (Ko et al., 2010) and play a very important part in edge detection (Kuang et al., 2012). This is also in line

with Otero-Millan et al. (2008), who reported increases in fixational saccades rates in a visual search task in those parts of the image that contained the targets.

As mentioned in the introduction, the central part of the fovea (approx. 0.3°–0.4°) does not contain S-cones. Thus, S-cones could perhaps play a less important role in driving exploratory saccades that are coupled with foveal processing strategies. Our results on fixational oculomotor behavior complement findings on voluntary saccades driven by S-cone isolating stimuli, with absence of overt but not covert inhibition of return (Sumner et al., 2002, 2004) already reported. Further, visual search is less efficient for stimuli that differ from other elements in the search array only in S-cone increment contrast (Lindsey et al., 2010). The low-level and high-level influences on fixational saccades were independent of each other, implying two separate control systems. Fixational saccade rates were reduced for S-cone isolating contours compared to full color contours which parallels the effect observed for accuracy. However, they did not correlate with contrast or performance measures, which suggests that they did not make a particularly strong contribution to efficient task performance.

An alternative account of our behavioral and fixational saccade findings would be that the multiple-of-discrimination-threshold approach did not appropriately match contrasts between different directions in color space, S − (L + M) stimuli being particularly adversely affected. This would have led to a reduction in both performance and fixational saccades. There are several arguments against this interpretation. A contrast mismatch would have led to general differences in saliency, thus similar patterns of results should be expected for objects and non-objects. However, we observed an interaction between the two factors in the analysis of accuracy rates and reaction times, with performance differences between directions in color space emerging for objects but not for non-objects. The overall levels of accuracy were, however, relatively low. Although stimuli were displayed at on average 2–3.5 times threshold, performance in the majority of conditions did not reach ceiling, ranging from around 83% correct to around 97% correct (see **Figure 5A**). Thresholds were measured for discriminating objects from non-objects in a 2IFC paradigm, while the main experiments use single-trial discrimination of images. Transition to a one-interval forced choice (1IFC) would lead to a decrease of performance equivalent to $\sqrt{2}$ times 2IFC threshold (Kingdom and Prins, 2010). While the performance decrease for S-cone isolating stimuli in the eye movement experiment can be approximated in this fashion on the basis of units-of-threshold, this is not the case for full-colour stimuli, in which performance for objects is far superior than what would be predicted simply on the basis of 2IFC-to-1IFC performance transition (see **Figure 5A**). As discussed previously, differences between object and non-object performance and their relations to suprathreshold contrast are an important result of this study. There is, however, one more potential issue that could emerge due to the transition between 2IFC and 1IFC: the single-trial task has the problem of being "criterion-dependent" (for a detailed elaboration, see Kingdom and Prins, 2010). There is a risk that the criterion-free 2IFC is not suitable for equating contrasts for single-trial yes/no tasks if the transition to a single trial also introduces a large bias. This can cause differences in accuracy, as the biased category would receive near-ceiling accuracy while the opposite category would have much lower accuracy rates. In a recent study, we have found that single-trial classification of line drawing objects and non-objects, such as those used in this study, does not introduce biases and results in similar sensitivity across different mechanisms and their combinations for stimuli at threshold (Martinovic et al., 2013). In addition to that, inspection of **Figure 5A** demonstrates that ceiling effects were not consistently reached for objects or non-objects, which is another argument against a large bias for any of the two categories in our multiple-of-threshold stimuli.

In the EEG, we observed low levels of gamma-band activity, with above baseline eGBA in 1 of 6 conditions and above baseline tGBA in 2 out of 6 conditions. Evoked GBA was related to iGBA in a causal fashion by Herrmann et al. (2004b) and in the S-cone isolating non-object condition in our study both responses are indeed above baseline. However, this is not the case for the other condition with significant tGBA. All previous studies with visual objects resulted in a robust, high-amplitude eGBA response, followed by a small-amplitude iGBA (see e.g., Busch et al., 2006; Fründ et al., 2008; Martinovic et al., 2008a,b). Although our data provides some support that the two responses are likely to occur together, it also partly runs contrary to Herrmann et al.'s (2004b) memory match and utilization model, since eGBA does not always precede tGBA. The modulations of evoked and total GBA in our study also dissociate, with eGBA being influenced by low-level factors and tGBA showing a combined low and high-level modulation. The tGBA effect seems to be driven by the difference between intermediate isoluminant objects and non-objects (see **Figure 9C**). Larger tGBA relative power for intermediate isoluminant non-objects also showed a tendency to be associated with faster responses, which indicates that tGBA 200–700 ms post-stimulus onset might relate to task performance. However, the fact that the signals are weak and thus likely to be noisy makes these effects very difficult to interpret and necessitates a replication.

Furthermore, around one third of participants (6 out of 17) were rejected due to inadequate ocular artifact removal from tGBA. It can be argued that this was because tGBA and fixational saccades are intrinsically coupled, and therefore it is problematic to remove ocular artifacts without removing cortical-only signal. However, Craddock et al. (2013) have already used the Keren et al. (2010) approach successfully to remove ocular artifacts and reveal underlying tGBA. Therefore, we presume that artifact rejection has failed on those participants due to the fact that they made smaller numbers of fixational saccades. The ICA approach relies predominantly on the quality and amount of the initial input (Groppe et al., 2009)—in other words, if there were not enough fixational movements to successfully train the algorithm, this would adversely affect the artifact removal procedure. We did indeed have fewer peri-saccadic trials to subject to the ICA for these rejected participants than for the rest of the sample. We consider this to be due to the relatively low levels of saccades elicited by our stimuli. Poletti and Rucci (2010) suggested that the required precision of fixation has a great contribution to the

miniature saccade rate's modulation and our experiments had a fixation cross superimposed over the stimulus, unlike Hassler et al. (2011) and Yuval-Greenberg et al. (2008). The number of miniature saccades decreases as the fixation target gets bigger (McCamy et al., 2013), but we used a relatively small fixation cross. Perhaps even more importantly, the stimuli were of low contrast when compared to those usually used in object recognition studies, which is likely to result in fewer microsaccades (Cui et al., 2009). When considering artifact removal efficiency in terms of the fixational saccade findings of Cui et al. (2009), one should also consider the difference in suprathreshold contrast between experiments. Since contrast was lower for participants in the eye movement experiment, the number of saccades in the EEG experiment was likely to have been larger, but in spite of that artifact removal proved to be problematic in a large number of participants.

The number of analyzed trials per condition is also important in achieving adequate signal-to-noise ratio when studying small amplitude EEG components. In our study, the number of analyzed trials does not differ much to studies on GBA prior to the publication of Yuval-Greenberg et al.'s paper in 2008. For example, an average of 44 trials in this experiment compares to 47 trials in Martinovic et al. (2007). However, since the removal of ocular artifact reduces overall amplitude, the number of trials could have posed an additional problem for obtaining adequate signal-to-noise ratio in the tGBA window (Jerbi et al., 2009). Insufficient number of trials would have had adverse effect on the gamma-activity levels. However, there are inherent limitations when working with meaningful, nameable stimulus sets. We used 225 images for threshold 2IFC measurements and 168 images for the single-trial main experiments, compiled from a range of existing stimulus sets. It is difficult to include more images without having pictures of familiar objects that look overly similar, introducing undesirable memory effects, or including images of relatively unfamiliar objects or objects from non-canonical views which pose their own recognition challenges. Recent studies with meaningful, nameable object stimuli used 100 stimuli per condition (Hassler et al., 2013) and 74 stimuli per condition (Craddock et al., 2013), which is higher than the 56 stimuli per condition in this study. A study with a larger number of stimuli, utilizing matched-contrast isoluminant conditions, would be needed before a firm conclusion could be made that isoluminant line-drawing stimuli are not suitable for eliciting GBA in general.

Comparison of fixational saccade findings and GBA findings is complicated by the fact that they were conducted on two samples which differed in contrast levels at which the stimuli were displayed. However, in terms of performance, between-experiment differences concerned only object-class, indicating that lower contrast has a more adverse effect on performance for non-objects. The important finding that performance for line-drawings of objects is more contrast-invariant will need to be replicated with other stimulus materials (e.g., outlines, line fragmented stimuli, Gaborised stimuli). The main importance of this study is that it shows for the first time that peaks in saccade rate around 200–700 ms after stimulus onset are attenuated for S-cone isolating stimuli when compared to full-colour stimuli and that fixational saccades exhibit independent low and high-level effects, in line with Engbert's (2012) recent model. No relations with behavioral performance or contrast were found. On the other hand, eGBA 50–150 ms after stimulus onset depends on low-level factors and tGBA 200–700 ms after stimulus onset depends on both low and high-level factors, although both are of very low amplitude in this particular paradigm. Both fixational saccades and GBA therefore appear to be useful markers of visual processes involved in object recognition and classification, although studies with isoluminant and/or low contrast luminance stimuli may not be ideal for eliciting robust GBA. We conclude that cortical loops involved in the processing of objects are preferentially excited by stimuli that contain achromatic information. Their activation can lead to relatively early exploratory eye movements even for foveally-presented stimuli.

ACKNOWLEDGMENTS

The research was supported by a BBSRC New Investigator grant BB/H019731/1 to Jasna Martinovic and by a DFG project grant to Matthias M. Mueller and Jasna Martinovic. We would like to thank Justyna Mordal for her involvement in collecting the initial twelve EEG datasets, Alon Keren for advice on how to set up the artifact removal procedures, Joe MacInnes and Hannah Krueger for advice on how to use the eye tracking system and Frouke Hermens for advice on setting up the fixational saccade experiment.

AUTHOR CONTRIBUTIONS

Jasna Martinovic and Sophie M. Wuerger designed the experiment; Maciej Kosilo and Jasna Martinovic collected the data; Maciej Kosilo, Sophie M. Wuerger, Matt Craddock, Ben J. Jennings and Jasna Martinovic analyzed the data; Maciej Kosilo, Sophie M. Wuerger, Matt Craddock, Ben J. Jennings, Amelia R. Hunt and Jasna Martinovic wrote the manuscript.

REFERENCES

Alario, F. X., and Ferrand, L. (1999). A set of 400 pictures standardized for French: norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behav. Res. Methods Instrum. Comput.* 31, 531–552. doi: 10.3758/BF03200732

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/0898929033216 62976

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychon. Bull. Rev.* 10, 344–380. doi: 10.3758/BF03196494

Betta, E., and Turatto, M. (2006). Are you ready? I can tell by looking at your microsaccades. *Neuroreport* 17, 1001–1004. doi: 10.1097/01.wnr.0000223392.82198.6d

Brainard, D. (1996). "Cone Contrast and Opponent Modulation Color Spaces," in *Human Color Vision,* eds P. Kaiser and R. M. Boynton. (Washington, DC: Optical Society of America), 563–579.

Bumsted, K., and Hendrickson, A. (1999). Distribution and development of short-wavelength cones differ between Macaca monkey and human fovea. *J. Comp. Neurol.* 403, 502–516. doi: 10.1002/(SICI)1096-9861(19990125)403: 4<502::AID-CNE6>3.3.CO;2-E

Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., and Herrmann, C. S. (2004). Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clin. Neurophysiol.* 115, 1810–1820. doi: 10.1016/j.clinph.2004.03.015

Busch, N. A., Herrmann, C. S., Müller, M. M., Lenz, D., and Gruber, T. (2006). A cross-laboratory study of event-related gamma band activity in a standard object recognition paradigm. *Neuroimage* 33, 1169–1177. doi: 10.1016/j.neuroimage.2006.07.034

Cole, G. R., Hine, T., and Mcilhagga, W. (1993). Detection mechanisms in L-cone, M-cone and S-cone contrast space. *J. Opt. Soc. Am. A Opt. Image Sci. Vision* 10, 38–51. doi: 10.1364/JOSAA.10.000038

Cornsweet, T. N. (1956). Determination of the stimuli for involuntary drifts and saccadic eye movements. *J. Opt. Soc. Am.* 46, 987–993. doi: 10.1364/JOSA.46.000987

Craddock, M., Martinovic, J., and Müller, M. M. (2013). Task and spatial frequency modulations of object processing: an EEG study. *PLoS ONE* 8:e70293. doi: 10.1371/journal.pone.0070293

Cui, J., Wilke, M., Logothetis, N. K., Leopold, D. A., and Liang, H. L. (2009). Visibility states modulate microsaccade rate and direction. *Vision Res.* 49, 228–236. doi: 10.1016/j.visres.2008.10.015

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Derrington, A. M., Krauskopf, J., and Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of Macaque. *J. Physiol.* 357, 241–265.

Engbert, R. (2012). Computational modeling of collicular integration of perceptual responses and attention in microsaccades. *J. Neurosci.* 32, 8035–8039. doi: 10.1523/JNEUROSCI.0808-12.2012

Engbert, R., and Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Res.* 43, 1035–1045. doi: 10.1016/S0042-6989(03)00084-1

Engbert, R., and Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychol. Sci.* 15, 431–436. doi: 10.1111/j.0956-7976.2004.00697.x

Eskew, R. T., McLellan, J., and Guilianini, F. (1999). "Chromatic detection and discrimination," in *Color vision: from genes to perception,* eds K. Gegenfurtner and L. Sharpe. (Cambridge, UK: Cambridge University Press), 345–368.

Fründ, I., Busch, N. A., Korner, U., Schadow, J., and Herrmann, C. S. (2007). EEG oscillations in the gamma and alpha range respond differently to spatial frequency. *Vision Res.* 47, 2086–2098. doi: 10.1016/j.visres.2007.03.022

Fründ, I., Busch, N. A., Schadow, J., Gruber, T., Körner, U., and Herrmann, C. S. (2008). Time pressure modulates electrophysiological correlates of early visual processing. *PLoS Biol.* 3:e1675. doi: 10.1371/journal.pone.0001675

Groppe, D. M., Makeig, S., and Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *Neuroimage* 45, 1199–1211. doi: 10.1016/j.neuroimage.2008.12.038

Gruber, T., and Müller, M. M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb. Cortex* 15, 109–116. doi: 10.1093/cercor/bhh113

Hafed, Z. M., Goffart, L., and Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *Science* 323, 940–943. doi: 10.1126/science.1166112

Hamm, J. P., and McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 413–426. doi: 10.1037/0096-1523.24.2.413

Hartigan, J. A., and Hartigan, P. M. (1985). The dip test of unimodality. *Ann. Stat.* 13, 70–84. doi: 10.1214/aos/1176346577

Hassler, U., Barreto, N. T., and Gruber, T. (2011). Induced gamma band responses in human EEG after the control of miniature saccadic artifacts. *Neuroimage* 57, 1411–1421. doi: 10.1016/j.neuroimage.2011.05.062

Hassler, U., Friese, U., Martens, U., Trujillo-Barreto, N., and Gruber, T. (2013). Repetition priming effects dissociate between miniature eye movements and induced gamma-band responses in the human electroencephalogram. *Eur. J. Neurosci.* 38, 2425–2433. doi: 10.1111/ejn.12244

Hendry, S. H.C., and Reid, R. C. (2000). The koniocellular pathway in primate vision. *Annu. Rev. Neurosci.* 23, 127–153. doi: 10.1146/annurev.neuro.23.1.127

Herrmann, C. S., Frund, I., and Lenz, D. (2010). Human gamma-band activity: a review on cognitive and behavioral correlates and network models. *Neurosci. Biobehav. Rev.* 34, 981–992. doi: 10.1016/j.neubiorev.2009.09.001

Herrmann, C. S., Lenz, D., Junge, S., Busch, N. A., and Maess, B. (2004a). Memory-matches evoke human gamma-responses. *BMC Neurosci.* 5:13. doi: 10.1186/1471-2202-5-13

Herrmann, C. S., Munk, M. H.J., and Engel, A. K. (2004b). Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn. Sci.* 8, 347–355. doi: 10.1016/j.tics.2004.06.006

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)* 195, 215–243.

Jasper, H. H. (1958). The 10-20 electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375.

Jerbi, K., Ossandon, T., Hamame, C. M., Senova, S., Dalal, S. S., Jung, J. L., et al. (2009). Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. *Hum. Brain Mapp.* 30, 1758–1771. doi: 10.1002/hbm.20750

Kaiser, P. K., and Boyton, R. M. (1996). *Human Color Vision.* Washington, DC: Optical Society of America.

Keren, A. S., Yuval-Greenberg, S., and Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: Characterization, detection and suppression. *Neuroimage* 49, 2248–2263. doi: 10.1016/j.neuroimage.2009.10.057

Kingdom, F. A. A., and Prins, N. (2010). *Psychophysics: A Practical Introduction*. London: Academic Press; Elsevier.

Ko, H. K., Poletti, M., and Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nat. Neurosci.* 13, 1549–1553. doi: 10.1038/nn.2663

Kuang, X. T., Poletti, M., Victor, J. D., and Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Curr. Biol.* 22, 510–514. doi: 10.1016/j.cub.2012.01.050

Kulikowski, J. J. (2003). "Neural basis of fundamental filters in vision," in *Modulation of Neuronal Signalling: Implications for Active Vision,* eds G. T. Buracas, O. Ruksenas, G. M. Boynton, and T. D. Albright (Dordrecht: NATO Science Series, Life Sciences), 3–68. Available online at: https://www.escholar.manchester.ac.uk/uk-ac-man-scw:3c198

Kveraga, K., Boshyan, J., and Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240. doi: 10.1523/JNEUROSCI.3481-07.2007

Laubrock, J., Kliegl, R., Rolfs, M., and Engbert, R. (2010). When do microsaccades follow spatial attention? *Atten. Percept. Psychophys.* 72, 683–694. doi: 10.3758/APP.72.3.683

Lee, T. W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* 11, 417–441. doi: 10.1162/089976699300016719

Lindsey, D. T., Brown, A. M., Reijnen, E., Rich, A. N., Kuzmova, Y. I., and Wolfe, J. M. (2010). Color channels, not color appearance or color categories, guide visual search for desaturated color targets. *Psychol. Sci.* 21, 1208–1214. doi: 10.1177/0956797610379861

Macleod, D. I. A., and Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *J. Opt. Soc. Am.* 69, 1183–1186. doi: 10.1364/JOSA.69.001183

Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci.* 8, 204–210. doi: 10.1016/j.tics.2004.03.008

Makin, A. D. J., Ackerley, R., Wild, K., Poliakoff, E., Gowen, E., and El-Deredy, W. (2011). Coherent illusory contours reduce microsaccade frequency. *Neuropsychologia* 49, 2798–2801. doi: 10.1016/j.neuropsychologia.2011.06.001

Marrocco, R. T., and Li, R. H. (1977). Monkey superior colliculus - properties of single cells and their afferent inputs. *J. Neurophysiol.* 40, 844–860.

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., and Hubel, D. H. (2009). Microsaccades: a neurophysiological analysis. *Trends Neurosci.* 32, 463–475. doi: 10.1016/j.tins.2009.05.006

Martinovic, J., and Busch, N. A. (2011). High frequency oscillations as a correlate of visual perception. *Int. J. Psychophysiol.* 79, 32–38. doi: 10.1016/j.ijpsycho.2010.07.004

Martinovic, J., Gruber, T., and Müller, M. M. (2007). Induced gamma-band responses predict recognition delays during object identification. *J. Cogn. Neurosci.* 19, 921–934. doi: 10.1162/jocn.2007.19.6.921

Martinovic, J., Gruber, T., Hantsch, A., and Müller, M. M. (2008a). Induced gamma-band activity is related to the time point of object identification. *Brain Res.* 1198, 93–106. doi: 10.1016/j.brainres.2007.12.050

Martinovic, J., Gruber, T., and Müller, M. (2008b). Coding of visual object features and feature conjunctions in the human brain. *PLoS ONE* 3:e3781. doi: 10.1371/journal.pone.0003781

Martinovic, J., Milne, M., Semple, E., and Jennings, B. J. (2013). "Comparison of object classification from luminance, chromatic and mixed-contrast line drawings and their Gaborised outlines," in *22nd Symposium of the International*

Colour Vision Society, eds. V. Bonnardel, J. L. Barbur, and M. Rodriguez-Carmona (Winchester: The Colour Group (Great Britain)).

Martinovic, J., Mordal, J., and Wuerger, S. M. (2011). Event-related potentials reveal an early advantage for luminance contours in the processing of objects. *J. Vis.* 11, 1. doi: 10.1167/11.7.1

McCamy, M. B., Najafian Jazi, A., Otero-Millan, J., Macknik, S. L., and Martinez-Conde, S. (2013). The effects of fixation target size and luminance on microsaccades and square-wave jerks. *PeerJ.* 1, e9. doi: 10.7717/peerj.9

Melloni, L., Schwiedrzik, C. M., Rodriguez, E., and Singer, W. (2009). (Micro)Saccades, corollary activity and cortical oscillations. *Trends Cogn. Sci.* 13, 239–245. doi: 10.1016/j.tics.2009.03.007

Mergenthaler, K., and Engbert, R. (2010). Microsaccades are different from saccades in scene perception. *Exp. Brain Res.* 203, 753–757. doi: 10.1007/s00221-010-2272-9

Metting Van Rijn, A. C., Peper, A., and Grimbergen, C. A. (1990). High quality recording of bioelectric events: I: interference reduction, theory and practice. *Med. Biol. Eng. Comput.* 28, 389–397. doi: 10.1007/BF02441961

Metting Van Rijn, A. C., Peper, A., and Grimbergen, C. A. (1991). High quality recording of bioelectric events. II: a low noise low-power multichannel amplifier design. *Med. Biol. Eng. Comput.* 29, 433–440. doi: 10.1007/BF02441666

Mitra, P. P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophys. J.* 76, 691–708. doi: 10.1016/S0006-3495(99)77236-X

Nolan, H., Whelan, R., and Reilly, R. B. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *J. Neurosci. Methods* 192, 152–162. doi: 10.1016/j.jneumeth.2010.07.015

O'Donell, B. M., Barraza, J. F., and Colombo, E. M. (2010). The effect of chromatic and luminance information on reaction times. *Vis. Neurosci.* 27, 119–29. doi: 10.1017/S0952523810000143

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869. doi: 10.1155/2011/156869

Otero-Millan, J., Macknik, S. L., Langston, R. E., and Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6175–6180. doi: 10.1073/pnas.1222715110

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., and Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *J. Vis.* 8. doi: 10.1167/8.14.21

Poletti, M., and Rucci, M. (2010). Eye movements under various conditions of image fading. *J. Vis.* 10, 6. doi: 10.1167/10.3.6

Regan, B. C., Reffin, J. P., and Mollon, J. D. (1994). Luminance noise and the rapid determination of discrimination ellipses in color deficiency. *Vision Res.* 34, 1279–1299. doi: 10.1016/0042-6989(94)90203-8

Reid, R. C., and Shapley, R. M. (2002). Space and time maps of cone photoreceptor signals in macaque lateral geniculate nucleus. *J. Neurosci.* 22, 6158–6175.

Rieder, M. K., Rahm, B., Williams, J. D., and Kaiser, J. (2011). Human gamma-band activity and behavior. *Int. J. Psychophysiol.* 79, 39–48. doi: 10.1016/j.ijpsycho.2010.08.010

Ripamonti, C., Woo, W. L., Crowther, E., and Stockman, A. (2009). The S-cone contribution to luminance depends on the M- and L-cone adaptation levels: Silent surrounds? *J. Vis.* 9, 10. doi: 10.1167/9.3.10

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Res.* 49, 2415–2441. doi: 10.1016/j.visres.2009.08.010

Ruppertsberg, A., Wuerger, S. M., and Bertamini, M. (2003). The chromatic input of global motion perception. *Vis. Neurosci.* 20, 421–428. doi: 10.1017/S0952523803204077

Schadow, J., Lenz, D., Thaerig, S., Busch, N. A., Frund, I., Rieger, J. W., et al. (2007). Stimulus intensity affects early sensory processing: visual contrast modulates evoked gamma-band activity in human EEG. *Int. J. Psychophysiol.* 66, 28–36. doi: 10.1016/j.ijpsycho.2007.05.010

Schwartzman, D. J., and Kranczioch, C. (2011). In the blink of an eye: the contribution of microsaccadic activity to the induced gamma band response. *Int. J. Psychophysiol.* 79, 73–82. doi: 10.1016/j.ijpsycho.2010.10.006

Shevell, S. K., and Kingdom, F. A. A. (2008). Color in complex scenes. *Annu. Rev. Psychol.* 59, 143–166. doi: 10.1146/annurev.psych.59.103006.093619

Sumner, P., Adamjee, T., and Mollon, J. D. (2002). Signals invisible to the collicular and magnocellular pathways can capture visual attention. *Curr. Biol.* 12, 1312. doi: 10.1016/S0960-9822(02)01020-5

Sumner, P., Nachev, P., Vora, N., Husain, M., and Kennard, C. (2004). Distinct cortical and collicular mechanisms of inhibition of return—revealed with S cone stimuli. *Curr. Biol.* 14, 2259–2263. doi: 10.1016/j.cub.2004.12.021

Szekely, A., and Bates, E. (2000). *Objective Visual Complexity as a Variable in Studies of Picture Naming*. Center for Research in Language Newsletter, 12. Available online at: http://www.crl.ucsd.edu/newsletter/12-2/article.html

Tailby, C., Cheong, S. K., Pietersen, A. N., Solomon, S. G., and Martin, P. R. (2012). Colour and pattern selectivity of receptive fields in superior colliculus of marmoset monkeys. *J. Physiol. Lond.* 590, 4061–4077. doi: 10.1113/jphysiol.2012.230409

Tailby, C., Solomon, S. G., and Lennie, P. (2008). Functional asymmetries in visual pathways carrying S-cone signals in Macaque. *J. Neurosci.* 28, 4078–4087. doi: 10.1523/JNEUROSCI.5338-07.2008

Tallon-Baudry, C. (2009). The roles of gamma-band oscillatory synchrony in human visual cognition. *Front. Biosci.* 14, 321–332. doi: 10.2741/3246

Truccolo, W. A., Ding, M. Z., Knuth, K. H., Nakamura, R., and Bressler, S. L. (2002). Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clin. Neurophysiol.* 113, 206–226. doi: 10.1016/S1388-2457(01)00739-8

Valsecchi, M., Dimigen, O., Kliegl, R., Sommer, W., and Turatto, M. (2009). Microsaccadic inhibition and P300 enhancement in a visual oddball task. *Psychophysiology* 46, 635–644. doi: 10.1111/j.1469-8986.2009.00791.x

Valsecchi, M., and Turatto, M. (2007). Microsaccadic response to visual events that are invisible to the superior colliculus. *Behav. Neurosci.* 121, 786–793. doi: 10.1037/0735-7044.121.4.786

Vassilev, A., Murzac, A., Zlatkova, M. B., and Anderson, R. S. (2009). On the search for an appropriate metric for reaction time to suprathreshold increments and decrements. *Vision Res.* 49, 524–529. doi: 10.1016/j.visres.2008.12.010

Walsh, J. W. T. (1958). *Photometry, 3rd Edn.* London, UK: Constable & Co. Ltd.

Watson, A. B., and Pelli, D. (1983). QUEST: a Bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120. doi: 10.3758/BF03202828

White, B. J., Boehnke, S. E., Marino, R. A., Itti, L., and Munoz, D. P. (2009). Color-related signals in the primate superior colliculus. *J. Neurosci.* 29, 12159–12166. doi: 10.1523/JNEUROSCI.1986-09.2009

Wuerger, S. M., Ruppertsberg, A., Malek, S., Bertamini, M., and Martinovic, J. (2011). The integration of local chromatic motion signals is sensitive to contrast polarity. *Vis. Neurosci.* 28, 239–246. doi: 10.1017/S0952523811000058

Wuerger, S. M., Watson, A. B., Ahumada, A. J. Jr. (2002). "Towards a spatio-chromatic standard observer for detection," in *Human Vision and Electronic Imaging VII, Proceedings of SPIE*, Vol. 4662, eds B. E. Rogowitz and T. N. Pappas (Bellingham, WA: SPIE), 159–172. doi: 10.1117/12.469512

Wyszecki, G., and Stiles, W. S. (2000). *Color Science: Concepts and Methods, Quantitative Data and Formulae.* New York, NY: John Wiley & Sons.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., and Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron* 58, 429–411. doi: 10.1016/j.neuron.2008.03.027

Zele, A. J., Cao, D. C., and Pokorny, J. (2007). Threshold units: A correct metric for reaction time? *Vision Res.* 47, 608–611. doi: 10.1016/j.visres.2006.12.003

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 August 2013; accepted: 30 November 2013; published online: 18 December 2013.

*Citation: Kosilo M, Wuerger SM, Craddock M, Jennings BJ, Hunt AR and Martinovic J (2013) Low-level and high-level modulations of fixational saccades and high frequency oscillatory brain activity in a visual object classification task. Front. Psychol. **4**:948. doi: 10.3389/fpsyg.2013.00948*

This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.

Copyright © 2013 Kosilo, Wuerger, Craddock, Jennings, Hunt and Martinovic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Psychology

ORIGINAL RESEARCH
published: 16 September 2015
doi: 10.3389/fpsyg.2015.01289

Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions

Haline E. Schendan[1]* *and Giorgio Ganis*[1, 2, 3]

[1] School of Psychology, Cognition Institute, University of Plymouth, Plymouth, UK, [2] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA, [3] Department of Radiology, Harvard Medical School, Boston, MA, USA

OPEN ACCESS

Edited by:
Chris Fields,
Independent Researcher, USA

Reviewed by:
Alex D. Clarke,
University of California, Davis, UK
Caterina Gratton,
University of California, Berkeley, USA

***Correspondence:**
Haline E. Schendan,
School of Psychology, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK
haline.schendan@plymouth.ac.uk

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *23 December 2013*
Accepted: *12 August 2015*
Published: *16 September 2015*

Citation:
Schendan HE and Ganis G (2015) Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions. Front. Psychol. 6:1289. doi: 10.3389/fpsyg.2015.01289

People categorize objects more slowly when visual input is highly impoverished instead of optimal. While bottom-up models may explain a decision with optimal input, perceptual hypothesis testing (PHT) theories implicate top-down processes with impoverished input. Brain mechanisms and the time course of PHT are largely unknown. This event-related potential study used a neuroimaging paradigm that implicated prefrontal cortex in top-down modulation of occipitotemporal cortex. Subjects categorized more impoverished and less impoverished real and pseudo objects. PHT theories predict larger impoverishment effects for real than pseudo objects because top-down processes modulate knowledge only for real objects, but different PHT variants predict different timing. Consistent with parietal-prefrontal PHT variants, around 250 ms, the earliest impoverished real object interaction started on an N3 complex, which reflects interactive cortical activity for object cognition. N3 impoverishment effects localized to both prefrontal and occipitotemporal cortex for real objects only. The N3 also showed knowledge effects by 230 ms that localized to occipitotemporal cortex. Later effects reflected (a) word meaning in temporal cortex during the N400, (b) internal evaluation of prior decision and memory processes and secondary higher-order memory involving anterotemporal parts of a default mode network during posterior positivity (P600), and (c) response related activity in posterior cingulate during an anterior slow wave (SW) after 700 ms. Finally, response activity in supplementary motor area during a posterior SW after 900 ms showed impoverishment effects that correlated with RTs. Convergent evidence from studies of vision, memory, and mental imagery which reflects purely top-down inputs, indicates that the N3 reflects the critical top-down processes of PHT. A hybrid multiple-state interactive, PHT and decision theory best explains the visual constancy of object cognition.

Keywords: category decision, categorization, identification, recognition, object constancy, visual perception, event-related potentials, knowledge memory

Introduction

People categorize objects accurately (e.g., car, dog, hat) even when visual input is impoverished, for example, due to fog, poor lighting, or unusual viewing angles. They show remarkable visual *constancy* of categorization: People maintain high accuracy despite suboptimal viewing conditions, though performance is slower with impoverished than optimal visual stimuli (Palmer et al., 1981; Tarr et al., 1998). Hierarchical bottom-up processing along the ventral visual stream and frontoparietal decision-making processes have well-established, necessary roles in the visual constancy of category decisions (Tanaka, 2003; Grill-Spector and Malach, 2004; Philiastides and Sajda, 2007). However, recent evidence implicates additional top-down feedback modulations onto posterior information processing areas in order to explain human performance fully, especially under more impoverished conditions (Kosslyn et al., 1994), in which case bottom-up models underperform people (Serre et al., 2007a).

This study aimed to address a critical unanswered issue of when and how bottom-up processes and top-down feedback contribute to visual category decisions. Most prior work focused on functional anatomy using slow hemodynamic measures with a time scale of seconds (Grill-Spector et al., 1999; Lerner et al., 2001), but few used electromagnetic techniques with high time resolution within the range of neural processing (i.e., milliseconds), such as event-related potentials (ERPs), as used here. Also, most studies and theories focus on object cognition under optimal visual input. Consequently, the time when the visual constancy of object cognition is achieved under non-optimal conditions in humans has received relatively little attention.

Timing is important because theories can be grouped into two major classes based on time course, early or late: Early theories propose an early time course within 130–215 ms via bottom-up (Thorpe et al., 1996) and/or top-down processes (Bar, 2003), and late theories propose a later time course and a key role for decision-making (Philiastides and Sajda, 2007) or top-down processes for attention (Stuss et al., 1992; Ganis et al., 2007; Schendan and Lucia, 2010; Clarke et al., 2011). Most vision theories, accounts, or models posit an early time course. Bottom-up models are based on the initial bottom-up pass through the ventral visual hierarchical pathway (Riesenhuber and Poggio, 1999) and posit early time courses (**Figure 1A**). However, a bottom-up model cannot fully explain the visual constancy of human object cognition (Serre et al., 2007a). For example, on ultra rapid category detection tasks, a name cues the target category before a masked image appears briefly (∼20 ms) (Delorme et al., 2000). When masking reduces feedback processing (Di Lollo et al., 2000), the initial fast feedforward sweep along the ventral stream dominates performance, consistent with computational models (Serre et al., 2007a). Critically, however, such bottom-up models cannot match human performance (a) when the mask is removed and so feedback inputs are involved, or (b) when people see the image longer before the mask appears (e.g., 80 vs. 50 ms) because then feedback inputs come into play long enough to boost performance. Bottom-up models also perform poorly when objects are impoverished (as by distance, i.e., farther away). Such limitations led to the suggestion that the bottom-up pathway could provide the initial input and object hypothesis to test using top-down processes (Serre et al., 2007b).

Consequently, other early and late theories posit an important role for feedback inputs. Most of these are *perceptual hypothesis testing* (PHT) theories that propose iterative top-down processes to achieve the visual constancy of object categorization. These top-down processes include *prediction* of a tentative object hypothesis based on prior information (e.g., memory) and *testing* of these predictions using ongoing perceptual input. Top-down processes are important when the stimulus input is ambiguous or impoverished. This is because stimulus ambiguity and impoverishment (e.g., due to rotation, deformation, and illumination changes from one experience to the next) cause the memory and currently perceived object to differ substantially in appearance (Ullman, 1996; Humphreys et al., 1997). This can result in an initial mismatch to stored memory and consequent failure of decision-making processes to categorize the object based on initial bottom-up computations. Temporal lobe, parietal, and prefrontal variants of PHT theories propose different mechanisms.

Temporal lobe variants (**Figure 1B**) capitalize on reciprocal connections among ventral visual areas in which bottom-up inputs automatically and reflexively trigger feedback from higher-level areas down to lower areas (Bullier, 2001; Ganis and Kosslyn, 2007). In such computational models (Ullman, 1996; Edelman, 1999), higher areas use stored knowledge to reach a fast initial, broad classification that feeds back to lower areas. This first top-down process interacts dynamically with bottom-up perceptual information to refine this classification. A second top-down process uses knowledge about the current context, such as the surrounding scene (e.g., kitchen) or task goal (e.g., find the car), to further select the most appropriate object model to feedback to lower level areas (Ullman, 1996). In addition, reverse hierarchy theory (Hochstein and Ahissar, 2002) proposes further that, once the initial bottom-up pass reaches advanced ventral visual areas, top-down processes for selective attention bind sensory features, and conscious visual perception begins (Treisman, 2006). Consequently, perceptual hypotheses are generated that project back along the visual hierarchy in reverse order to lower-level areas, which provide the detailed information needed to test the hypotheses. Interactive activation and competition theory (Humphreys et al., 1995, 1999) proposes further that these processes are task-dependent (e.g., most important for object naming) and involve multiple knowledge stores, which are themselves connected recurrently within and between each other (Price et al., 1996): A structural description system in left posterior inferotemporal cortex stores knowledge about shape and interacts with a semantic memory system, which, in turn, interacts with knowledge systems that store the names and semantic classes (e.g., animal, vehicle, tool).

Parietal and parietal-prefrontal variants propose that the ventral stream can support decisions about an object from known views, but, when viewing an object from an angle that impoverishes the image, additional spatial transformations must

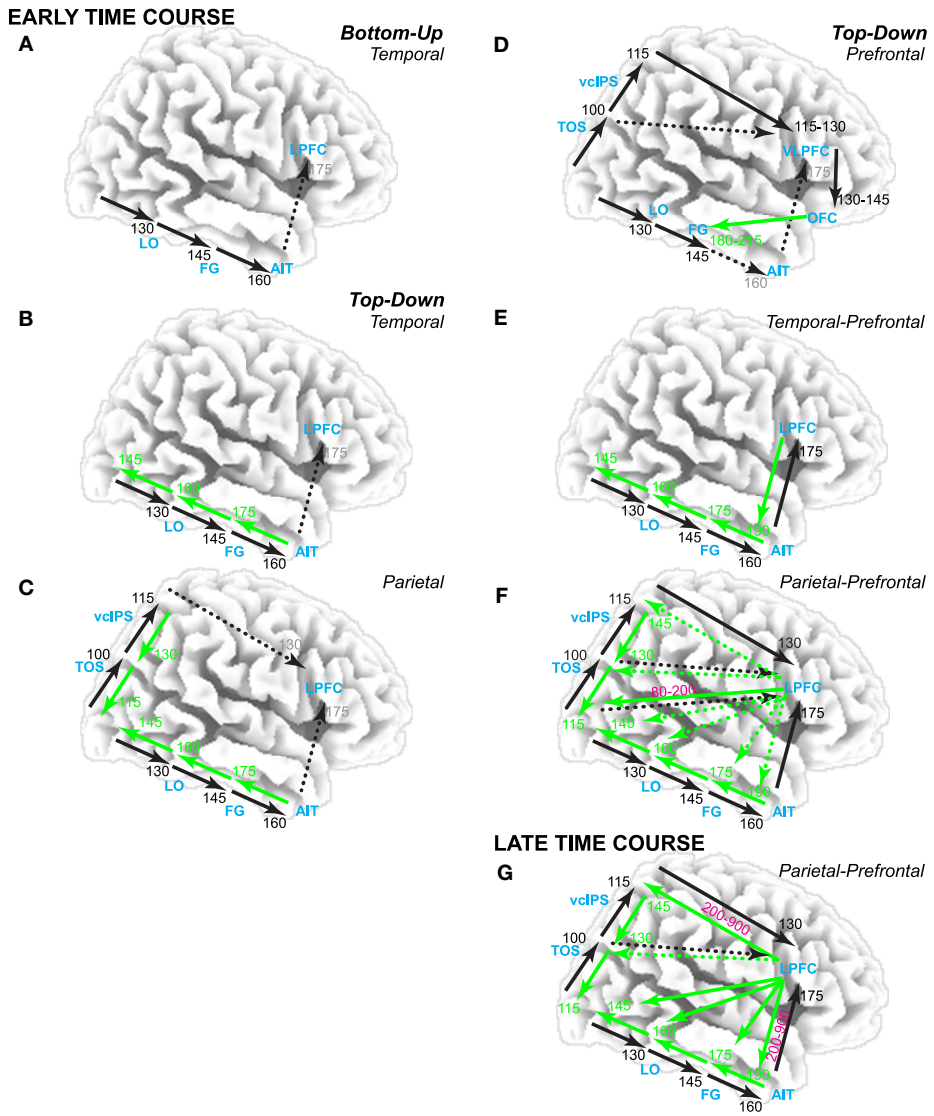


FIGURE 1 | Theories of visual category decisions. Timing estimates based on human brain electromagnetic potential data. Black arrows are bottom-up. Green arrows are top-down. Dotted arrows are implied but not specified. Times in black are earliest time of bottom-up input to that region. Times in green are earliest time of feedback input from nearest higher order area to that region. Times in magenta are when prefrontal top-down inputs interact with bottom-up and/or feedback interactions along the visual pathways. Times in gray are associated with implied activity. Theories posit an early time course before 200 ms **(A–F)** or a later time course **(G)**. **(A)** Bottom-up theories posit that the initial feedforward pass through the ventral visual pathway supports object cognition. According to decision theory, this supports a category decision in lateral prefrontal cortex (LPFC). In contrast, *perceptual hypothesis testing* (PHT) theories **(B–G)** emphasize top-down contributions: **(B)** Temporal lobe variants assume bottom-up inputs along the ventral visual hierarchy trigger feedback along the pathway, which consequently modifies bottom-up processing. **(C)** Parietal variants emphasize that the dorsal stream is necessary for complete object constancy. **(D)** One prefrontal variant posits a role for top-down input from ventral LPFC (VLPFC) and orbitofrontal cortex (OFC). **(E)** Temporal-prefrontal variants emphasize bottom-up and feedback processes from visual areas along the ventral pathway through prefrontal cortex. **(F)** Parietal-prefrontal variants emphasize parietal-prefrontal processes of selective attention to locations and features associated with an object category that have been cued by a search template prior to stimulus onset; this modulates visual processing early in time from 80 to 200 ms. **(G)** Late parietal-prefrontal variants emphasize parietal-prefrontal processes of selective attention that contribute model prediction and testing processes when the category is not cued before stimulus onset; note, fMRI tests of parietal-prefrontal PHT variants implicate VLPFC in model prediction and testing (Ganis et al., 2007; Schendan and Stern, 2008).

be computed, such as those implicated in mental rotation (Tarr and Pinker, 1989; Turnbull et al., 1997; Gauthier et al., 2002). These transforms align the percept and stored object knowledge spatially (Bülthoff et al., 1995) and may be implemented in occipitoparietal areas along the dorsal visual stream. A parietal variant predicts dorsal transforms are rapid, happening within 200 ms (**Figure 1C**), because the dorsal stream processes visual information faster than the ventral stream (Bullier, 2001). A parietal-prefrontal variant involves mental imagery processes implicated in mental rotation, which are slow because they

involve top-down processes from prefrontal cortex after 200 or 500 ms that are implicated in selective attention and model verification (see parietal-prefrontal theories below, **Figure 1G**) (Schendan and Kutas, 2003; Schendan and Lucia, 2009).

While temporal and parietal variants imply a role for prefrontal cortex, prefrontal variants specify such a role. One prefrontal variant (**Figure 1D**) assumes that people routinely accomplish object cognition within about 200 ms using low spatial frequency information from V2/V4 to compute a coarse scene representation along the dorsal pathway (Bar, 2003). This representation is sent forward rapidly into Brodmann's area (BA) 45 of ventral lateral prefrontal cortex (VLPFC) and then orbitofrontal cortex, which uses this information to predict possible categories within 130 ms after visual stimulation and feeds these back to fusiform cortex in the ventral stream within 180–215 ms (Bar et al., 2006). Other prefrontal PHT variants can be summarized within a free-energy type framework (Friston, 2010). Of these, temporal-prefrontal variants focus on ventral stream and prefrontal interactions (**Figure 1E**). For example, in hierarchical Bayesian models (Lee and Mumford, 2003), bottom-up processes (e.g., ventral stream) can yield a perceptual hypothesis that serves as a predictive code to test using information coming in from the stimulus (e.g., to prefrontal cortex). In contrast, parietal-prefrontal variants implicate top-down selective attention processes, which involve interactions between parietal and prefrontal cortex (Spreng et al., 2013). For example, in one such variant (**Figure 1F**), dorsolateral prefrontal area 46 feeds back a signal to visual areas that competitively biases processing of features at the attended location that match the search template for the object (Deco and Rolls, 2004). Spatial biases feedback via the dorsal pathway, and object biases feedback via the ventral pathway. This model aims to explain cognition when the location or object is cued before the stimulus and so attention can modulate early visual processing within 200 ms (Di Russo et al., 2003). In contrast, other models explain category decisions without cueing and implicate processes primarily after the initial bottom-up activation of the ventral stream, that is, after 200 ms (**Figure 1G**). For example, model verification theory (Lowe, 2000) proposes that, for a slightly impoverished image, the bottom-up pass can suffice to match the percept to the correct model, whereas for a more impoverished image (e.g., degraded picture), the bottom-up pass may only find a weak match to knowledge (or initial classification Ullman, 1996) that is insufficient for an accurate decision. Consequently, top-down processes implicated in selective attention perform model verification to determine the knowledge in posterior cortex that best explains the percept. A prediction process selects the locations of salient features, evaluates their match to knowledge, and generates a prediction about a candidate object model (e.g., a category). A testing process, which may involve parietal spatial transformation and mental rotation processes (e.g., as in some parietal vision theories), evaluates the predicted model for its fit with the percept. An adaptive resonance variant provides important computational solutions for how such processes may operate (Fazl et al., 2009), such as a mismatch reset signal from prefrontal cortex that controls prediction and testing cycles until enough evidence accumulates for a decision.

While vision and decision theories have evolved separately, both explain category decisions under uncertainty due to impoverished sensory input, and decision theories specify roles for prefrontal and parietal cortex. Evidence accumulation is a core process in decision-making theories (Ratcliff, 1978), which offer mathematical solutions for how frontoparietal areas accumulate and evaluate evidence for a decision (Gold and Shadlen, 2007). As perceptual impoverishment increases, decision certainty decreases, and decision processes are recruited more. Decision theories explain decision processes based on information from perception (Gold and Shadlen, 2007), category knowledge (Philiastides and Sajda, 2007), and recognition memory (Ratcliff, 1978). Decision accounts propose that prefrontal and parietal cortices accumulate evidence from ventral areas via bottom-up inputs (Philiastides and Sajda, 2007), making them bottom-up theories (like **Figure 1A** or the bottom-up pathways in **Figure 1F**). Critically, the brain regions and event-related potentials (ERPs) associated with category decisions and impoverishment effects on visual cognition are similar (e.g., Schendan and Kutas, 2003; Ganis et al., 2007; Jiang et al., 2007; Schendan and Stern, 2008; Wheeler et al., 2008). Findings from the present study favor a hybrid decision and parietal-prefrontal PHT theory in which both bottom-up and top-down interactions occur between prefrontal decision and posterior evidence components of the brain's decision network (**Figure 1G**).

In summary (**Figure 1**), vision and decision theories differ in involvement of parietal and prefrontal cortex and various top-down processes, which predicts different time courses. All propose object constancy of category decisions within 200 ms, except for parietal-prefrontal PHT theories that propose that, when the category is unknown before stimulus onset, interactive bottom-up and feedback processes from the visual pathways into lateral prefrontal cortex between 200 and 900 ms support object constancy.

The present study aimed to define the time course of category decisions under uncertainty due to impoverished visual input. To do so, ERPs were recorded using the paradigm from an fMRI study (Ganis et al., 2007) that uniquely manipulated both visual impoverishment and knowledge and found evidence favoring parietal-prefrontal PHT and decision theories (Philiastides and Sajda, 2007). Subjects decided whether they could categorize more (MI) and less (LI) impoverished drawings of real objects and pseudo versions of them, which differ in knowledge activation. FMRI activation is greater for MI than LI images, and more so for real than pseudo objects in the VLPFC (BA 45 and 47/12), occipitoparietal, and occipitotemporal object processing areas implicated in selective attention, spatial transformation, and category decisions. Critically, this *impoverished-real-object effect* implicates not only perceptual processing but also the knowledge activation needed for PHT and a category decision. After all, by design, real objects activate knowledge, whereas the novel shapes of pseudo objects do so minimally if at all (Kroll and Potter, 1984). Thus, impoverishment effects for both object types reveal perceptual processing, whereas those for real more than pseudo objects reflect knowledge processing, thereby distinguishing between the contributions of sensory-perceptual

vs. knowledge (i.e., memory) evidence used for PHT and a category decision. Critically, the fMRI pattern for impoverished real objects refutes a purely bottom-up account of object constancy, which predicts the opposite impoverishment effect (i.e., greater activation for LI images, regardless of object type, because LI images have more perceptual features). Moreover, when top-down processes for visuospatial working memory cannot be engaged fully in a category decision, performance is impaired with MI (but not LI) objects (Ganis et al., 2007). Thus, altogether, convergent evidence indicates that impoverished-real-object effects reflect top-down contributions, not only bottom-up input, to PHT and category decisions.

This design improves upon electromagnetic brain potential studies on object constancy, decisions, and category knowledge in four ways as follows. (1) It manipulates both impoverishment and object type (i.e., knowledge). Previously, either impoverishment of real objects in fragmented drawings (Viggiano and Kutas, 2000; Schendan and Kutas, 2002, 2007a; Schendan and Maher, 2009) and rotated views varied (Schendan and Kutas, 2003) or categorization success (knowledge) varied between stimuli (Holcomb and McPherson, 1994; Schendan et al., 1998; McPherson and Holcomb, 1999; Gruber and Müller, 2005, 2006; Gruber et al., 2006; Sehatpour et al., 2006, 2008; Schendan and Maher, 2009; Voss et al., 2010). (2) Pseudo objects here had been constructed from the real objects to equate them on low-level features, perceptual properties, and coherent object structure, and, in work with these intact versions, ERPs differ only after 175 ms when initial bottom-up processing is largely complete, confirming matched low-level sensory attributes between types (Schendan et al., 1998). Other studies compared real objects relative to either pseudo objects chosen from a different set of real objects that were unknown to subjects (Holcomb and McPherson, 1994; McPherson and Holcomb, 1999) or distorted or scrambled versions that are unknown (Gruber and Müller, 2005, 2006; Busch et al., 2006; Gruber et al., 2006; Sehatpour et al., 2006, 2008), or compared objects with less than more novel or meaningful visual structures (Daffner et al., 2000a; Folstein and van Petten, 2008; Voss et al., 2010). Notably, despite these visual differences, all these studies confirm ERP effects only after 175 or 215 ms, suggesting that knowledge is the primary factor distinguishing real and pseudo objects. (3) This experiment assessed many categories, whereas ERP work on category decisions focused on face-selective activity with cars as the comparison category (Philiastides et al., 2006; Philiastides and Sajda, 2006, 2007). (4) There is no repetition confound. Here, subjects categorize each object once, instead of repeatedly at multiple levels of impoverishment (Stuss et al., 1986; Doniger et al., 2000; Viggiano and Kutas, 2000; Schendan and Kutas, 2002; Philiastides and Sajda, 2006; Ratcliff et al., 2009). This is important because repetition affects behavior (i.e., priming) and ERPs, making them more positive after 200 ms (Schendan and Kutas, 2003, 2007a; Henson et al., 2004; Schendan and Maher, 2009), and these effects are larger for meaningful than meaningless objects (e.g., real vs. pseudo) (Snodgrass and Feenan, 1990; Schendan and Kutas, 2002; Schendan and Maher, 2009; Voss et al., 2010). Further, repetition effects differ between impoverishment levels, being largest at moderate levels (Snodgrass and Feenan, 1990) and when objects repeat from LI to MI than MI to LI (Schendan and Kutas, 2003).

The time when ERPs show the impoverished-real-object effect defines when PHT and decision processes contribute to the visual constancy of category decisions based on knowledge, not just sensory evidence. To infer the timing of cortical sources, ERP results were integrated with fMRI location information by both estimating the ERP sources and relating similar functional patterns between methods (Luck, 1999). To use vision and decision theories to predict the ERP effects, this report capitalizes on the multiple-state interactive (MUSI) account of the brain basis of visual object cognition to define the times and scalp sites to analyze (Schendan and Kutas, 2003, 2007a; Schendan and Maher, 2009; Schendan and Ganis, 2012). This framework proposes that posterior object processing areas activate at multiple times in brain "states" serving distinct functions. This account extends the principle that different brain areas can perform different functions for cognition at different points in time because bottom-up, feedback, and recurrent activity alters neuronal computations, as demonstrated, for example, in visual area V1 (Lamme and Roelfsema, 2000). Likewise, object-sensitive areas perform different functions in perception and cognition due to different neural computations associated with bottom-up, feedback, and recurrent activity (Schendan and Lucia, 2010).

State 1: Initial activity in object processing areas feeds forward from occipital to temporal cortex between ~120 and ~200 ms when a visual object is broadly perceptually categorized (e.g., as a face instead of nonface object) (Schendan et al., 1998; Schendan and Ganis, 2013), as described for ventral visual hierarchy processing (**Figures 1A,B**, **11**). This state is indexed by early ERPs reflecting activity in object-sensitive areas related to categorical perception: the vertex positive potential and its occipitotemporal N170 counterpart (VPP/N170) (Schendan and Lucia, 2010). When input is optimal, this predominantly bottom-up activation of knowledge should be sufficient (Serre et al., 2007a) to enable object cognition (i.e., entry level categorization) and phenomenological awareness of this knowledge in State 2 with little or no need for additional top-down processing from prefrontal cortex (Schendan and Kutas, 2007a).

State 2: Object processing areas activate again interactively due primarily to top-down processing among these areas and VLPFC as well as other areas such as parietal cortex (Schendan and Lucia, 2009). This is indexed by mid-latency negative ERPs between 200 and 500 ms: an N3 complex (including components known as template matching N2[00], N300, N350, frontal N400). The N3 is the first ERP in response to pictures that modulates according to cognitive factors affecting posterior object processing cortex and VLPFC similarly (Barrett and Rugg, 1990; Zhang et al., 1995; McPherson and Holcomb, 1999; Doniger et al., 2000, 2001; Curran et al., 2002; Schendan and Kutas, 2002, 2003, 2007a; Folstein and van Petten, 2004, 2008; Philiastides and

Sajda, 2006, 2007; Philiastides et al., 2006; Sehatpour et al., 2006; Gratton et al., 2009; Schendan and Lucia, 2009, 2010) and that localizes to these brain areas (David et al., 2005, 2006; Sehatpour et al., 2008; Schendan and Maher, 2009; Schendan and Lucia, 2010; Clarke et al., 2011; Bastin et al., 2013). States 1 and 2 are thus described in the time course for late parietal-prefrontal PHT theories (**Figure 1G**) and are consistent with these ideas for the first 500 ms of visual processing.

State 3: Top-down interactive processes, including conscious, effortful, cognitive control functions, perform internal evaluation, and verification after about 400 to 500 ms. For example, (a) a parietal P600 (or P3[00]) component reflects later strategic evaluation or verification of earlier category decision processes, being more positive for correct decisions, and strategic, effortful mental rotation of objects, being larger when more mental rotation is needed, and (b) a parietal late positive complex (LPC) complex is associated with higher-order semantic analysis, being larger when semantic integration is more challenging (i.e., contextually incongruous) (Schendan and Lucia, 2009; Schendan and Maher, 2009; Sitnikova et al., 2010).

For each theory, **Table 1** summarizes the predictions for the pattern of ERP effects, and the MUSI framework specifies the ERPs, effects, and their direction. Posterior cortex theories (**Figures 1A–C**) predict only early effects. See **Table 1** (VPP/N170 predictions i): All vision theories in **Figure 1** predict the same impoverishment and type effects between 130 and 215 ms. This is explained by the bottom-up processes in these theories. Bottom-up processing (e.g., **Figure 1A**) predicts overall less neural activity for MI than LI objects and for pseudo than real objects (i.e., independent impoverishment and type effects) during the initial bottom-up pass through the ventral stream in state 1. The impoverishment effect happens because MI objects show fewer visual features and so they activate fewer neurons and/or activate each neuron less, relative to LI objects. The type effect happens because the initial pass categorizes by activating knowledge, which is less successful for pseudo than real objects, by design. Altogether, this predicts that the VPP/N170 will be larger for LI than MI and for real than pseudo objects (see **Table 1** Bottom-up).

See **Table 1** (predictions ii): Temporal, parietal, and prefrontal variants of top-down PHT theories (**Figures 1B–D**, respectively) predict, in addition, early impoverished-real-object effects (see **Table 1** Temporal and Parietal, and Prefrontal) due to feedback at this time; note, for one prefrontal variant (Bar, 2003), this interaction effect will be found as long as MI stimuli contain sufficient low spatial frequency information to compute a coarse object representation along the dorsal stream.

See **Table 1** (Prefrontal; predictions iii): Prefrontal PHT variants can accommodate (**Figures 1D–F**) or predict (**Figure 1G**) later type and impoverishment effects. For example, one early prefrontal PHT variant can accommodate additional late type and impoverishment effects (see bottom-up inputs to AIT and VLPFC in **Figure 1D**). Type effects occur at later times when meaning is activated after categorization

TABLE 1 | Predicted pattern of impoverishment (I) and type (T) effects according to vision and decision theories and summary of ERP results.

Predictions				Theory					Results
	ERP	Effect	Direction	Bottom–up	Top–down perceptual hypothesis testing/decision				
				Early	Early		Late		
				Temporal	Temporal and Parietal	Prefrontal	Parietal–prefrontal	MUSI and Decision	
i	VPP/N170	I	LI Larger	X	X	X	X	–	–
i	145–160 ms	T	Real larger	X	X	X	X	–	X?
ii		I × T	I Larger for real	–	X	X	–	–	–
iii	N3	I	MI larger	–	–	?	X	X	X
iii	200–500 ms	T	Real larger	–	–	?	X	X	X
iv		I × T	I larger for real	–	–	–	X	?	X
iii	P600	I	LI larger	–	–	?	X	X	X
iii	500–900 ms	T	Real larger	–	–	?	X	X	X
iv		I × T	I Larger for real	–	–	–	X	?	X
	SW	I	LI larger						X
	700–900 ms	T	Real larger						X
		I × T	–						X
				Figure 1A	Figures 1B,C	Figures 1D–F	Figure 1G	Figures 1G, 11	Figure 11

*LI, less impoverished; MI, More impoverished; MUSI, Multiple State Interactive account of visual object cognition; X, predicted effect; ?, consistent but not specifically predicted; X?, Spurious effect due to low level sensory differences. N400 (300–500 ms) predictions and results are the same as for the N3. LPC (late positive complex) predictions and results are the same as for the P600. Early prefrontal theories include early prefrontal, temporal-prefrontal, and parietal-prefrontal as in **Figures 1D–F**.*

(Bar et al., 2006). Also later during post-categorization times, high spatial frequencies have a role (Bar, 2003), predicting impoverishment effects at later times due to less power at high spatial frequencies in MI than LI pictures. Early temporal-prefrontal and parietal-prefrontal PHT variants (**Figures 1E,F**) can likewise accommodate late type and impoverishment effects based on post-categorization processes. However, as categorization is already done, none of these predict late impoverished-real-object effects. Only late parietal-prefrontal PHT theories predict late type and impoverishment effects, as these propose that knowledge activation for the category decision with MI objects continues to be attempted after the initial bottom-up pass, that is, after 200 ms. The MUSI framework (**Table 1**) predicts the direction of these late ERP effects. Late ERPs will be more negative for MI than LI stimuli (impoverishment effect) and for real than pseudo objects (type effect); in other words, the N3 will be larger for MI stimuli and pseudo objects, whereas the P600/LPC will be larger for LI stimuli and real objects. This is due to stronger activation of memory for real than pseudo objects and LI than MI stimuli. This direction of impoverishment effects on the P600/LPC is also predicted by the slow mental rotation process in some parietal-prefrontal PHT variants (**Figure 1G**) because negativity is greater for more than less rotated objects (i.e., impoverished regarding match to memory) during mental rotation (Schendan and Lucia, 2009).

See **Table 1** (predictions iv): Late parietal-prefrontal PHT variants (**Figure 1G**) assume that bottom-up processing before 200 ms (as in **Figure 1A**) provides the front-end to later top-down processes, which predict later impoverished-real-object effects after 200 ms. The interaction effect would happen when prefrontal cortex biases attention (Deco and Rolls, 2004) or uses attention processes to control prediction and testing cycles (Lowe, 2000; Fazl et al., 2009). A later time course is consistent with ERP evidence for feature search along the ventral stream between 150–200 and 300–450 ms (Luck, 2006). By some accounts, the interaction happens when late mental rotation processes in frontoparietal cortex are recruited (Tarr and Pinker, 1989; Schendan and Stern, 2008). This predicts the interaction after 200 ms in state 2 during the N3 when parietal feedback interactions compute spatial relations among object parts and, especially after ∼500 ms in state 3 during the P600/LPC when spatial transformations implicated in mental imagery of object rotation happen (Schendan and Lucia, 2009). Note, some temporal-prefrontal PHT variants (**Figure 1E**, Humphreys et al., 1997; Hochstein and Ahissar, 2002) and decision theories can suggest an add-on of later selective attention processes that would essentially be the same mechanism described in parietal-prefrontal PHT theories (**Figure 1G**) and so could accommodate late type and impoverishment effects and their interaction. In addition, because these theories use a bottom-up model as the front end to hypothesis testing (e.g., model verification) or decision processes, they predict the same pattern of early effects as bottom-up models: Early impoverishment and type effects. They also predict no early interaction effects because frontoparietal contributions happen later.

The MUSI framework and decision theories predict type and impoverishment effects only during later ERPs. MUSI predicts this because category decision processes happen after the initial bottom-up pass after 200 ms (Schendan and Maher, 2009). Decision theories predict this due to bottom-up accumulation of evidence in frontoparietal areas implicated in decision-making and task difficulty between 200 and 450 ms during the D220 and late component (Philiastides and Sajda, 2007), which correspond to components of the N3 complex. MUSI and decision theories do not predict but can accommodate late impoverished-real-object effects, as both posit late prefrontal activity, and MUSI posits further that prefrontal top-down processes are critical for category decisions. Finally, note, most vision theories, other than parietal and parietal-prefrontal PHT theories, were created to explain cognition with optimal input so are problematic for predicting effects with MI stimuli and pseudo objects, but it is important to attempt to make explicit predictions in order to test the strengths and limitations of these theories.

For completeness, we assessed two other late ERPs that modulate during category decisions. Later in state 2, the centroparietal N400 between 300 and 500 ms reflects interactive activation of semantic memory, especially meaningful knowledge associated with linguistic stimuli (e.g., a name), in anterior temporal cortex and VLPFC (Marinkovic et al., 2003; Lau et al., 2008; Kutas and Federmeier, 2011). Only parietal-prefrontal PHT and decision theories posit a role for word meaning, which is knowledge that can contribute to category decisions and prediction. Hence, the N400 will be more negative for MI than LI and for real than pseudo objects and show impoverished-real-object effects, like the N3 and P600/LPC (**Table 1**). Also, a broad slow wave (SW) starting around 700 ms has been associated with response planning for category decisions, including naming, being more positive for named than unnamed objects (Schendan and Kutas, 2002, 2003; Folstein et al., 2008; Schendan and Lucia, 2009; Schendan and Maher, 2009; Sitnikova et al., 2010). This predicts greater SW positivity for LI than MI and for real than pseudo objects, but no interaction, as the SW reflects processes after the category decision.

Materials and Methods

Methods were the same as for the event-related fMRI version (Ganis et al., 2007) except for modifications needed for ERPs.

Materials

Fragmented drawings from the Snodgrass and Vanderwart (1980) set depicted 128 real objects and 64 pseudo versions of them. For a prior ERP study (Schendan et al., 1998), we created pseudo objects by rearranging parts of the real objects into perceptually closed objects that could exist in a Euclidean 3-dimensional world but not be categorized. Findings show processing differences between the matched sets of the intact real and pseudo objects only after 175 ms during the N3 complex, confirming that, as designed, real, and pseudo objects are well-matched for low-level visual feature processing. All drawings were impoverished by deleting random squares of pixels across 8 *fragmentation levels* in a series using the algorithm of Snodgrass

et al. (1987). Levels 1 (intact) to 6 (most fragmented) were used here. Such random impoverishment methods have the following advantages. First, fragmentation is not determined by a theory that could bias the features and properties in the stimuli, it does not depend on subjective judgments, and it produces stimuli that are challenging to categorize. Second, the stimuli do not depend upon uncontrolled variations in individual perceptual processing, as when visual input is impoverished by short presentation duration (Snodgrass et al., 1987; Snodgrass and Corwin, 1988a). Third, no masking is used that could limit top-down processes (Di Lollo et al., 2000). Of 260 fragmentation series for real objects, Snodgrass and Corwin (1988a) produced 150, and the first author produced 110 using the same software for a prior study (Schendan and Kutas, 2002). Two hundred of these series were chosen for the behavioral study that accompanied the fMRI version and generated normative data (Ganis et al., 2007) that were then used to choose 128 series, each of which had 2 fragmentation levels (low vs. high) that met two criteria: (1) At least 75% of people named each object correctly at both levels based on naming norms. (2) For each object, response times (RTs) were faster numerically for the low than high fragmentation level. Of these 128, 96 were from the Snodgrass and Corwin (1988a) set. Low fragmentation was intended for the LI condition; high fragmentation was intended for the MI condition. For pseudo objects, the same software fragmented these images to the same level as their corresponding real objects. These methods produced list I and its three orders used for fMRI (Ganis et al., 2007), and, for this ERP version, we added a second list (II): An object (real or pseudo) depicted at a higher fragmentation level in one list was presented instead at a lower fragmentation level in the other list, and *vice versa* (i.e., level 1, 2, 3, 4, 5, or 6 in list I became level 6, 5, 4, 3, 2, or 1 in list II, respectively). Each list was shown in 3 pseudo-random orders of intermixed, real, and pseudo objects counterbalanced across subjects. Based on normative data (Snodgrass and Vanderwart, 1980), stimuli chosen for the MI and LI real object conditions, respectively, did not differ in visual complexity (2.9 vs. 2.9), name agreement (86 vs. 87%), image agreement (3.7 vs. 3.6), familiarity (3.4 vs. 3.2), name frequency (18 vs. 15), and acquisition age (2.6 vs. 2.8).

Pseudo objects served two goals. First, they enable an *impoverished-real-object effect* to be revealed. By design (Schendan et al., 1998), these pseudo-objects match real object versions in low-level features, perceptual properties, and coherent object structure but, unlike real objects, activate knowledge weakly, if at all. Second, they served as catch trials to ensure that people categorized the real objects. Pseudo objects cannot be categorized by design, enabling subjects who do not reliably discriminate real and pseudo objects to be excluded. Catch trials validate the key press reports objectively and independently. While overt naming unambiguously reveals categorization accuracy (Schendan and Maher, 2009), it has the disadvantages of (a) demanding additional lexical retrieval not required for categorization *per se* (Damasio et al., 1996) and (b) introducing movement artifacts. Importantly, key press reports of categorization are reliable (Snodgrass and Yuditsky, 1996), and ERP effects are similar for key press and naming measures of categorization (Schendan and Maher, 2009). The design aimed to equate numbers of categorized and uncategorized trials so as not to discourage people from trying to categorize. While this necessitated using half the number of trials for pseudo relative to real objects, ample trials remained for valid ERPs in all conditions, as confirmed by visual inspection to ensure reliable waveforms from each subject. However, real and pseudo versions therefore also could not be presented in matched yoked pairs, as in our prior work showing no ERP effects before 175 ms (Schendan et al., 1998). Therefore, while, for completeness, the present study assesses ERP type effects before 175 ms, these likely reflect low-level feature differences, not just knowledge. Consequently, we focus conclusions on type effects after 175 ms that replicate those with the fully matched set (Schendan et al., 1998) and any impoverished-real-object effects (i.e., impoverishment by type interaction). Further any such interactions will be interpreted with this caveat in mind.

Design and Procedure

A 2 × 2 repeated measures factorial design (**Figure 2A**) included factors of impoverishment (LI, MI) and object type (real, pseudo). General health history and Edinburgh Handedness (Oldfield, 1971) questionnaires were administered before each session. The ERP session started with instructions on the computer screen that subjects paraphrased aloud, and any misconceptions were corrected. They were instructed on the task, to maintain eye gaze on the fixation mark at the center of the screen, and blink only in the fixation period. They then received 10 practice trials using the experiment methods but different stimuli. On each experiment trial, a fixation period of 5400–5700 ms preceded each picture, which was presented for 1000 ms while subjects decided whether they could categorize each object. They pressed "1" as soon as they knew what the object was, or "2" if they did not know, as quickly as possible without sacrificing accuracy. Participants were informed that categorization would be challenging by design because the images were degraded. They were not informed that some objects were impossible to categorize (i.e., pseudo objects) and so, from the subjects' perspective, pseudo objects were just images that they could not categorize (i.e., possible "real" objects that they failed to categorize).

Electroencephalography (EEG)

The *ERP System* software (Holcomb, 2003) presented stimuli and recorded and analyzed data on PCs running Windows XP. A *Belkin* Nostromo game pad detected responses. EEG data were recorded at 200 Hz (bandpass 0.01 to 100 Hz; SA Instrumentation Company) from 60 Ag/AgCl electrodes attached to a plastic cap (**Figure 2B**). Cap, nose, and right mastoid electrodes and one below the right eye (monitoring eye blinks) were referenced to the left mastoid. Bilateral eye electrodes (monitoring eye movements) were referenced to each other. Using *ERP System* software and standard methods (Luck, 2005), 27% of EEG trials were excluded from analysis that contained above threshold blinks (determined for each individual participant, and based on polarity inversion between the lower eye and right frontopolar electrode 4), eye and other movement artifacts (based on peak to peak amplitude for the bilateral

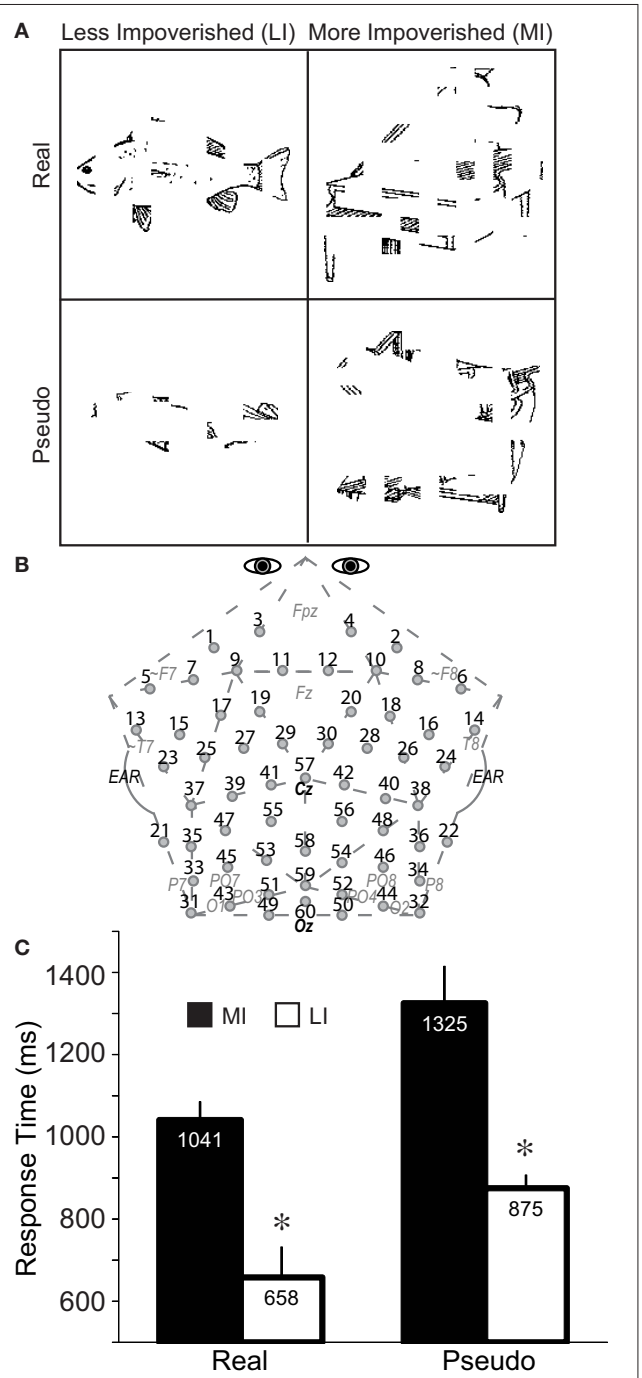


FIGURE 2 | Method and performance. (A) A 2 × 2 repeated measures design was used with impoverishment (less, more) and object type (real, pseudo) as factors. Fragmented line drawings of real and pseudo objects were shown. Pseudo objects had been created by re-locating the local parts of each real object to create a closed, perceptually coherent but unknown more global shape that could exist in a Euclidean 3-dimensional world but cannot be categorized (Schendan et al., 1998). Subjects pressed "1" to report that they categorized the object or, if not, they pressed "2," as soon as possible after the picture appeared. A median split of the RTs to real and pseudo objects, separately, for correct responses (i.e., 1 for real objects, 2 for pseudo objects) separated these conditions into more (MI) and less (LI) impoverished

(Continued)

FIGURE 2 | Continued

conditions. Shown are real objects of an LI fish at fragmentation level 3, and MI piano at level 4, and an LI pseudo-fish at level 5, and MI pseudo-piano at level 4; note, sample stimuli reflect the consistent finding that more fragmented real objects are related to slower RTs, whereas more fragmented pseudo objects are related to faster RTs. Stimuli subtended 6 by 6 degrees of visual angle, on average, with a visual contrast of approximately 30% (dark pixels against a brighter background). **(B)** Custom 60-channel geodesic montage for EEG recording (Electrocap International). Circles show electrode locations. Numbers label each electrode. Approximate locations of 10–20 sites are shown in gray italics; site 57 is at Cz, site 60 is Oz; pairs 31–32, and 49–50 are 1 cm below the inion. **(C)** Response times to MI and LI real and pseudo objects. Error bars show the 95% confidence interval (Morey, 2008). *Significant impoverishment effect.

eye electrodes and individual electrodes, respectively), muscle activity (based on high frequency local peaks within a time period). ERPs were calculated offline by averaging artifact-free EEG in each condition, time-locking to object onset with a 100 ms pre-stimulus baseline, and re-referencing to the mean of both mastoids. To compare with some prior studies, ERPs were also re-referenced to the common average of all electrodes, except bilateral eyes, and plotted positive up, which highlights the resemblance between frontopolar N3 effects with the mastoid reference (e.g., site 3) and occipitotemporal positivity ("P3") effects with the common average reference (e.g., site 22).

Analyses

Accuracy and the RTs and ERPs on correct trials were analyzed. "Correct trials" for real objects corresponded to "categorized" responses (i.e., hits). "Correct trials" for pseudo objects corresponded to "not categorized" responses (i.e., correct rejections). For each subject, the RT median for real and pseudo objects, separately, split trials into MI (slower) and LI (faster) conditions, which was the main analysis in the fMRI version and found to be most valid way to subdivide the trials to reveal impoverishment effects (Ganis et al., 2007). For the fMRI version, data were also re-analyzed using fragmentation level to define MI and LI conditions, revealing the same results as for the median RT split, though slightly less significant, consistent with the known performance variability among fragmentation series (Snodgrass and Corwin, 1988a). Consequently, categorization performance (i.e., median RT split), as opposed to fragmentation level, best captures the full set of image characteristics that defines each stimulus' goodness (i.e., impoverishment) for a category decision: Individual RT captures all factors that impoverish each picture and affect the category decision, and the results define the full range of processes that contribute to the visual constancy of object cognition. Thus, for completeness, as for fMRI, data were analyzed in two additional ways: (a) over fragmentation levels and (b) RT median split for only levels 3, 4, and 5 for which average visual complexity was equated between the MI and LI sets. For the latter (b), median RTs were re-computed for levels 3–5 and trials split into MI and LI conditions, accordingly: 98 of 128 real objects in list I; 75 of them in list II (fewer due to the level switch); for correct trials after artifact rejection, about 52 real and 39 pseudo object trials were analyzed from each subject

on average. For the former (a), to assess whether results would change if fragmentation defined MI and LI levels, ERP data were re-analyzed using fragmentation level to define MI (levels 4–5) and LI (levels 2–3) conditions; these levels yielded similar trial numbers in each condition, while also minimizing perceptual differences between MI and LI trials. Indeed, as for fMRI, the ERP results defined using fragmentation replicated those using the RT definition (both all trials and levels 3–5). In sum, regardless of how impoverishment is defined, results remained the same. As results of all analyses did not differ, the best controlled analysis that yielded the largest effects (i.e., RTs for levels 3–5) is reported.

Mean ERP amplitudes, time windows and electrodes were chosen based on prior ERP studies of vision and categorization; all components analyzed here have known scalp distributions (Picton et al., 2000; Luck, 2005): (a) From 145 to 160 ms assessed the VPP/N170 (Schendan and Lucia, 2010). (b) The N3 complex is a negative-going ERP over frontal locations that can sometimes invert polarity over occipitotemporal locations between 200 and 700 ms with a peak typically around 350 ms. As the N3 complex has subcomponents that can differ over time, the frontal N3 and its occipitotemporal counterparts were assessed from 200 to 299, 300 to 399, and 400 to 499 ms; note, the 300 to 499 ms times also assessed the centroparietal N400 (Schendan and Maher, 2009). (c) From 500 to 699 ms assessed the P600, (d) 700 to 899 ms assessed the SW, and both these time periods after 500 ms also assessed the LPC. Focal spatiotemporal planned contrast ANOVAs isolated effects (*df* s[1, 18]) to lateral pairs or midline sites and times when an ERP was maximal and overlapped least with others: (a) 145 to 160 ms for the VPP at pair 29–30, and its polarity inverted N170 at occipitotemporal pair 33–34; (b) 200 to 299, 300 to 399, and 400 to 499 ms for frontopolar ERPs at pair 3–4 and occipitotemporal polarity inverted counterparts at pair 21–22, and 300 to 399 and 400 to 499 ms for frontocentral negativities at pair 29–30; (c) pair 47–48 from 300 to 399 and 400 to 499 ms for the centroparietal N400; (e) pair 53–54 from 500 to 699 and 700–899 ms for the parietal P600 and broad LPC; (d) 500 to 699 and 700 and 899 ms for the SW at frontocentral pair 11–12 and broad LPC. The Bonferroni method corrected for planned comparison of multiple sites within a time period by dividing the alpha of 0.05 for each time period by the number of sites tested (**Table 3**).

Mixed ANOVAs included 2 Impoverishment (MI, LI) $\times$ 2 object Type (real, pseudo) within-subjects factors and between-subject nuisance variables of list (I, II) and order (A, B, C) of no interest and not reported. For ERP ANOVAs, a within-subjects factor of electrode was added, and midline (labeled as such) and lateral electrodes (unlabeled) were analyzed separately to assess hemispheric asymmetries with an added within-subject factor of hemisphere in lateral ANOVAs, and, in midline ANOVAs, lobe (parietal [sites 57, 58], occipital [59, 60]). The Huynh–Feldt correction was applied for violations of the sphericity assumption. For brevity, only results for critical factors of impoverishment and type, and their interactions are reported, as scalp location effects alone are not of theoretical interest. Degrees of freedom (*df* s) are listed with the first report of each effect. Planned simple effects tests assessed the impoverishment by type interaction for focal results, which target specific ERP components.

Source Estimates

Theoretically, the inverse problem of localizing the cortical sources of electromagnetic data recorded from the scalp has no unique solution. Standardized low resolution brain electromagnetic tomography (s*LORETA*) estimates the sources (Pascual-Marqui, 2002). The sLORETA software computes the three-dimensional (3D) distribution of current density using a standardized, discrete, 3D distributed, linear, minimum norm inverse solution. Localization is data-driven, unbiased (even with noisy data), and exact but has low spatial precision due to smoothing assumptions resulting in highly correlated adjacent cortical volume units. A realistic head model constrains the solution anatomically using the structure of cortical gray matter from the Montreal Neurological Institute (MNI) average of 152 human brains as determined using the probabilistic Talairach atlas. Images plot the exact magnitude of the estimated current density based on the standardized electrical activity in each of 6239 voxels of 5 mm^3 size. The sLORETA software computed the sources of the grand average ERPs over all sites, except nose, and eyes (Pascual-Marqui, 2002). Electrode coordinates were digitized using an infrared digitization system, and imported into *LORETA-Key* software. This coordinate file was then converted using the sLORETA electrode coordinate conversion tools. The transformation matrix was calculated with a regularization parameter (smoothness) corresponding to a signal-to-noise ratio of 50. We localized the difference waves of each of the 4 effects (**Figure 7**). The ERP difference data are akin to the signal differences between fMRI conditions and so limit sources to those that could reflect fMRI activation, and difference waves may reveal weaker sources better (Luck, 2005).

Subjects

Ethical approval granted through the Institutional Review Board of Tufts University. Participants were 39 healthy Tufts University students or people from the greater Boston community. 1 person was excluded due to a data recording error and another due to strabismus. Data were analyzed from 24 of the 37 subjects remaining who met the following inclusion criteria: (a) The d'-value was 1.0 or better ($\mu = 2.35$) based on the hit rate for real objects, and false alarm rate to pseudo objects out of the total trials eliciting a response (i.e., excluding ambiguous no responses). (b) Two-thirds or more of real and pseudo object trials were correct to ensure valid RTs and ERPs following artifact rejection. (c) Visual inspection of each subject and condition confirmed each ERP was valid ($\mu = 28$ and 26 trials, respectively, at levels 3–5) (Picton et al., 2000). The analyzed group was half female, aged $\mu = 21.2$ years (range 18.0–29.8), had education $\mu = 14.4$ years (range 12–20), and handedness score $\mu = 97.8$ (right-handed).

Results

Performance

Performance replicated the fMRI version (Ganis et al., 2007). Results of signal detection theory (SDT) analyses with logistic distributions (Snodgrass and Corwin, 1988b) validated category decision accuracy. Subjects reliably decided that real objects were categorized and pseudo objects were not. The average

discrimination index (d'_L) was 4.13 (corrected rates: 73.6% hits, 6.9% false alarms), demonstrating very high detection of knowledge conveyed by real objects. The average criterion (C_L) was 0.97, which was above the neutral 0 level [$t_{(23)} = 7.80$, $p < 0.001$], indicating subjects were slightly biased to be conservative in reporting detection of knowledge. Subjective probability that each picture could be categorized can affect ERPs, such as P300-like potentials (e.g., P600, LPC) (Johnson, 1986), so, to assess this, response rates were computed collapsed across both object types (real, pseudo). Results showed that subjects decided that they could categorize about half of the pictures: 50.0% categorized vs. 49.0% uncategorized [levels 3–5, $F_{(1, 18)} = 0.13$, $p = 0.72$]. This 50:50 decision rate demonstrates that subjective probability of response type (and picture categorizability) cannot explain ERP effects.

RTs (**Figure 2C**) were faster in LI than MI conditions, by design, $F_{(1, 18)} = 182.83$, and for real than pseudo objects, $F_{(1, 18)} = 25.14$ (ps < 0.0001). LI were faster than MI, but more so for pseudo than real objects, resulting in an Impoverishment by type interaction, $F_{(1, 18)} = 9.25$, $p = 0.007$. Since this could be due to the overall slower RTs for pseudo than real objects, normalized RT scores (MI-LI/MI) were analyzed, demonstrating that impoverishment effects were actually greater for real (score = 0.36) than pseudo objects (score = 0.33), $F_{(1, 18)} = 6.09$, $p = 0.024$. Results do not reflect speed-accuracy trade-offs, because RTs and accuracy for real objects did not correlate across subjects ($r = 0.14$, $p > 0.5$). Analyses of the relation between fragmentation level and RT confirmed that, as designed, RT correlated with fragmentation level for real objects, $r = 0.61$, $p < 0.001$.

ERPs

The aim was to determine when impoverishment and object type interact such that the impoverishment effect is larger for real than pseudo objects. **Table 1** summarizes ERP results, which were most consistent with late parietal-prefrontal PHT, MUSI, and decision theories. After 200 ms, impoverishment affected knowledge activation, modulating the N3 complex, N400, P600, and SW (**Figures 3**, **4**); note, as results suggested no distinct LPC effects, henceforth, we refer only to the P600 and the SW.

N170/VPP

From 145 to 160 ms, omnibus results showed that object type interacted significantly with lateral and midline electrode sites (**Table 2**). Focal spatiotemporal analyses showed a marginal type effect at frontocentral pair 29–30 (**Table 3**) where positivity was slightly greater for real than pseudo objects.

N3 Complex and N400

Omnibus results at N3 and N400 times from 200 to 500 ms (**Table 2**) showed significant effects of type and impoverishment. Most important, impoverishment by type interactions were significant at lateral sites the entire time from 200 to 500 ms and at the midline from 400 to 500 ms.

N3 complex (200–500 ms)

Focal spatiotemporal results demonstrated that the frontal N3 was more negative for (a) MI than LI stimuli for real objects only (**Figures 3**, **5A**) and (b) pseudo than real objects on LI more than MI trials (**Figures 3**, **5B**). Occipitotemporal counterparts showed the same but with opposite polarity (i.e., more positive). Specifically, the results (**Table 3**) showed main effects of type were significant the entire time from 200 to 500 ms at frontopolar, frontocentral, and occipitotemporal sites. Main effects of impoverishment were significant at frontopolar sites the entire time, frontocentral sites from 400 to 500 ms, and occipitotemporal sites from 200 to 300 ms. The critical impoverishment by type interactions were significant at frontopolar sites from 300 to 400 ms; note, interactions were marginal at other times frontally and occipitotemporally from 200 to 300 ms. Planned contrasts (**Table 3**) showed that only real objects had significant impoverishment effects during the entire frontopolar N3 (200 to 500 ms) and later frontocentral N3 (300 to 500 ms); note, this effect was marginal on the occipitotemporal N250 from 200 to 300 ms. Further, type effects were significant, for LI, at all times and N3 sites and, for MI, from 200 to 400 ms at frontopolar sites and all times at occipitotemporal sites; note, for MI, type was marginal at frontocentral sites. With a common average reference, N3 effects split about evenly between frontal and occipitotemporal sites (**Figures 4**, **5C,D**).

N400 (300–500 ms)

Focal results demonstrated that the N400 was less negative for LI real objects than all other stimuli, demonstrating impoverished-real-object effects (**Figures 3**, **4**, **6**). Specifically, the results (**Table 3**) showed significant impoverishment effects at centroparietal pair 47–48 from 400 to 500 ms, though type effects and the impoverishment by type interaction were marginal. Planned contrasts (**Table 3**) supported the critical interaction, as impoverishment was significant for real objects only, and type was significant for LI stimuli only. Notably, while the earlier frontal N3 showed type effects for both MI and LI stimuli, type effects between 400 and 700 ms at the parietal N400 and P600 sites, occurred only for LI objects, dissociating the frontal and parietal ERPs.

P600/LPC (500–700 ms)

Around 500 ms, N3 complex effects ended, and the parietal P600 showed impoverished-real-object effects, as the impoverishment effect was larger for real than pseudo objects. Positivity was greater for LI than MI stimuli and for real than pseudo objects, and the impoverishment effect was larger for real than pseudo objects (**Figures 3**, **6**). With a common average reference, a left mid-parietal P600 inverted polarity to an N600 at right frontal sites (**Figure 4**). Accordingly, omnibus results from 500 to 700 ms resembled those from 400 to 500 ms, demonstrating type and impoverishment effects and their interaction (**Table 2**).

Focal results (**Table 3**) at parietal pair 53–54 showed significant effects of impoverishment and type, though their interaction was marginal. Planned contrasts (**Table 3**) showed impoverishment was significant for both object types for the first time between 500 and 700 ms, as earlier ERPs showed impoverishment effects only for real objects. Further, type was significant for LI stimuli only. These results confirm the

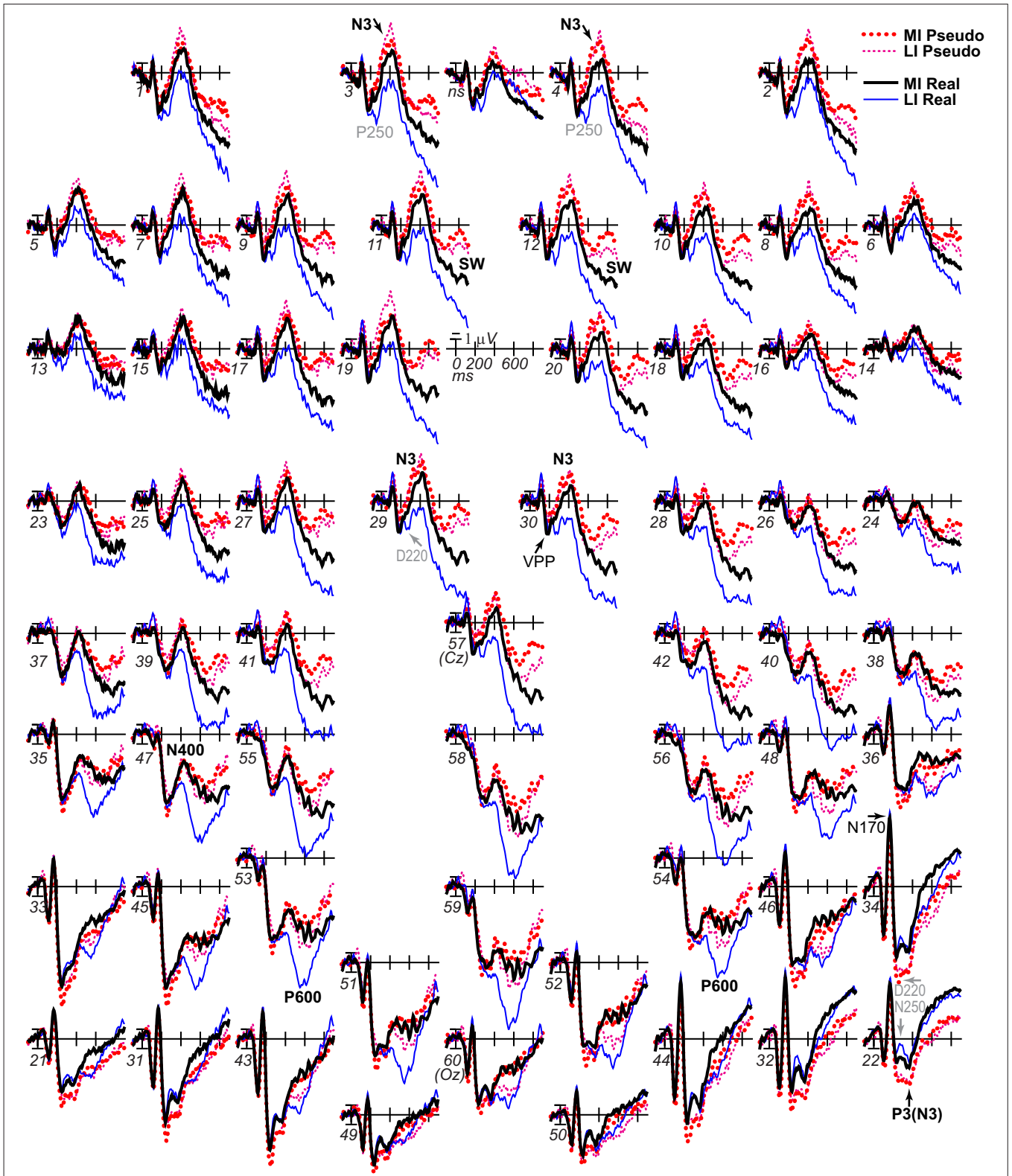


FIGURE 3 | ERP effects of impoverishment and object type. Grand average ERPs at all channels show effects of impoverishment (more [MI], less [LI] impoverished) and object type (real, pseudo). Unless otherwise specified, ERPs in this and following figures were low-pass filtered at 30 Hz and were referenced to the average of left and right mastoids. Numerals label electrode locations; ns, nose. Impoverishment and object type modulated the N3 complex (including P250/N250 and D220 components; components inverted polarity between frontal and occipitotemporal sites), N400, P600, and slow wave (SW) components after 200 ms, but not the earlier VPP/N170.

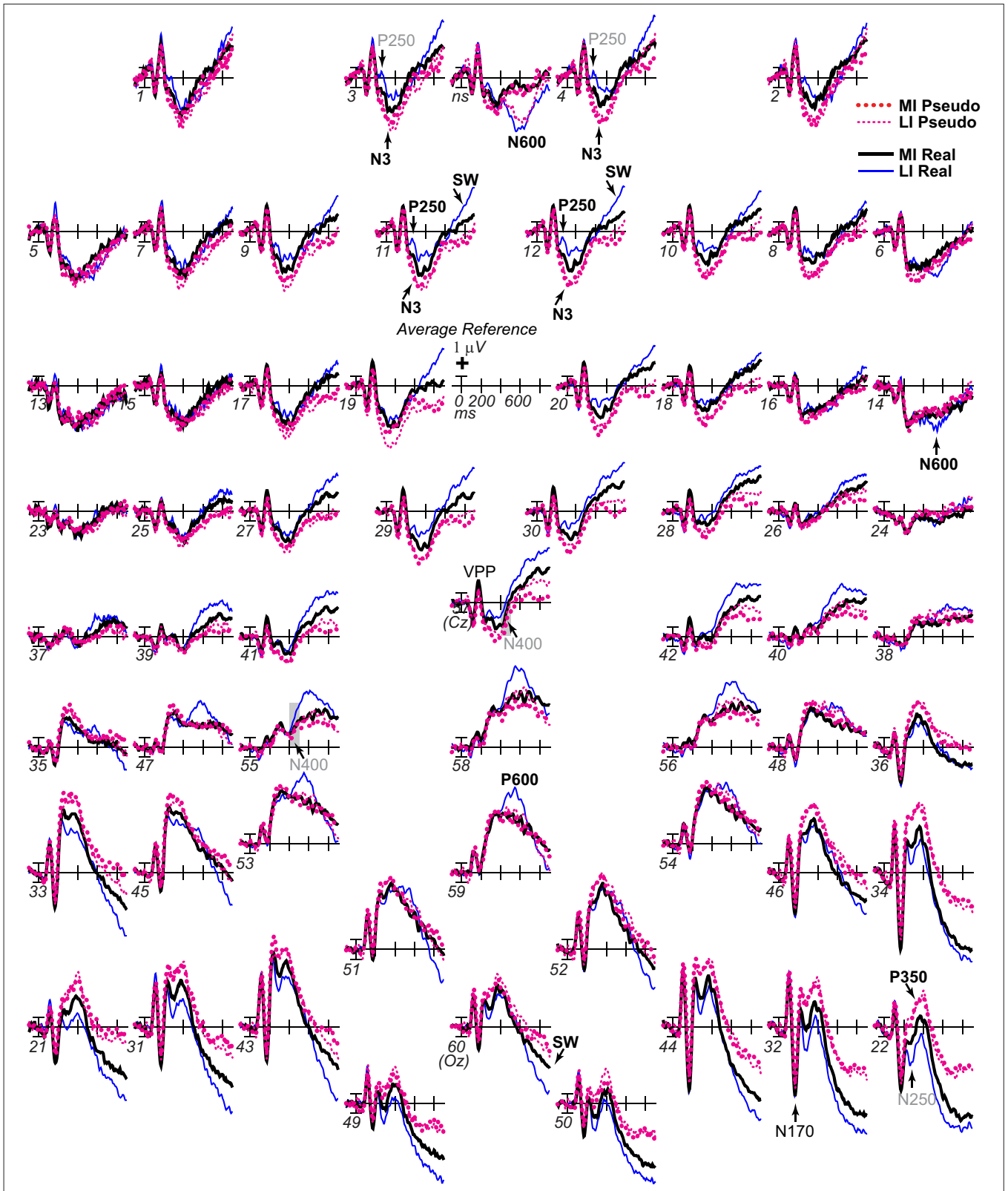


FIGURE 4 | ERP effects of impoverishment and object type with common average reference. As in **Figure 3**, grand average ERPs are plotted at all sites, but, in order to compare with other work, the reference was computed using the average of all scalp sites (i.e., the "average reference"), and ERPs were instead plotted positive up. ERP effects of impoverishment (more [MI], less [LI] impoverished) and object type (real, pseudo) are shown. Compared with **Figure 3**, with the

(Continued)

FIGURE 4 | Continued
average reference, here, the parietal P600 inverts polarity over lateral frontal and frontopolar sites to an N600, especially at the right. The late SW from 500 to 900 ms has an occipital distribution that inverts polarity over frontocentral sites near the midline, and is larger over the left hemisphere. Note, with the common average reference, the N400 pattern (gray shadow) cannot be discerned from the overlapping N3 and P600 times, highlighting the importance of using the same reference sites across studies to identify components and draw conclusions; studies analyzing data using the common average reference may misattribute N3 and/or P600 effects to the N400.

TABLE 2 | *F*–values for significant effects in omnibus lateral (Lat) and midline (Mid) ANOVAs with impoverishment and object type factors at each time period after 200 ms.

ERP	VPP/N 170		N3				N3 and N400		P600		SW	
Time (ms)	145–160		200–300		300–400		400–500		500–700		700–900	
Source	Lat	Mid	Lat	Mid	Lat	Mid	Lat	Mid	Lat	Mid	Lat	Mid
Impoverishment (I)	–	–	–	–	–	–	15.0**	19.8**	38.1**	49.7**	11.9**	–
Type (T)	–	–	–	–	6.3*	4.4*	14.2**	17.8**	23.0**	30.0**	112.3**	46.8**
I × Hemisphere (H)	–	–	–		–		–		–		5.8*	
I × Electrode (E)	–	–	5.1**	–	–	–	2.8*	–	4.8**	–	5.2*	13.2**
I × Lobe (L)	–	–		–		–		–		–		16.1**
I × L × E	–	–		–		–		7.4*		14.3**		7.0*
T × E	4.57**	8.03*	35.7**	28.4**	48.6**	52.7**	22.3**	20.9**	32.8**	32.5**	54.4**	81.3**
T × E × H	–	–	3.0**		2.5**		3.4**		4.2**		2.1*	
T × L	–	10.2**		40.1**		45.4**		14.2**		50.0**		81.9**
T × L × E	–	–		–		–		11.1**		39.8*		5.7*
I × T	–	–	10.6**	–	5.5*	–	9.1**	5.3*	5.4*	5.1*	–	–
I × T × E	–	–	–	–	–	–	–	–	–	5.1*	–	7.6*
I × T × L × E	–	–		–		–		5.4*		8.6**		–

*df (1, 18) for lateral effects of I, T, I × T, I × H, and all Midline effects and their interactions; df (27, 486) for Lateral Electrode effects and their interactions. Times rounded to nearest 5 ms. Epsilon values for lateral sites at each time period (start-end time in ms): 145–160 (T × E = 0.132), 200–300 (T × E = 0.16; I × E = 0.16; T × E × H = 0.27), 300–400 (T × E = 0.17; T × E × H = 0.35), 400–500 (T × E = 0.20; I × E = 0.16; T × E × H = 0.22), 500–700(T × E = 0.21; I × E = 0.12; T × E × H = 0.25), 700–900(T × E = 0.15; I × E = 0.10; I × T × E = 0.12; T × E × H = 0.32). –p > 0.05; *p < 0.05; **p < 0.01.*

impoverished-real-object effect on the P600 and dissociate it from other ERPs.

SW/LPC (500–900 ms)

Around 700 ms, positivity on a broad anterior SW was greater for LI real objects than MI ones, which was greater than for LI pseudo objects than MI ones, and type effects continued (**Figure 3**). With a common average reference, the SW was a negativity at occipital sites that inverted polarity to positivity over mid-frontal sites (**Figure 4**). Omnibus results from 700 to 900 ms (**Table 2**) showed impoverishment and type effects continued, but the impoverished-real-object effect was only at the midline where the impoverishment by type by electrode interaction was significant due to impoverishment effects for real but not pseudo objects at central more than posterior midline sites.

Focal results at frontocentral pair 11–12 (**Table 3**) showed effects of type and impoverishment from 500 to 900 ms, and impoverishment and type interacted marginally from 700 to 900 ms. Planned contrasts (**Table 3**) showed impoverishment was significant for real objects from 500 to 900 ms and marginal for pseudo objects from 500 to 700 ms (LPC time only). Further, unlike the N400 and P600, the N3 and SW showed type effects for both LI and MI stimuli. Thus, no distinct LPC effects were observed, and the anterior SW from 700 to 900 ms showed impoverishment effects for real objects only.

N3 Onset

To define precisely when the impoverished-real-object effect starts, the onset of N3 effects was defined as the time when 15 consecutive points first become significant in a series of point-by-point *F*-tests (Picton et al., 2000) at focal frontopolar pair 3–4 and right occipitotemporal site 22, as frontal N3 effects were bilateral and occipitotemporal N250 effects were larger on the right. The criterion was met for the onset of type effects with LI stimuli by 230 ms. However, omnibus and focal results confirmed type and impoverishment effects during the N3 so it is informative to consider fewer consecutive times. The results thereby also suggested an onset around 250 ms for the impoverished-real-object effect when the most consecutive significant points showing this interaction were at frontopolar site 3 (7 points, $ps < 0.05$, plus 1, $p = 0.084$). Simple effects tests defined the start of impoverishment effects for real objects likewise as 255 ms at frontopolar site 4 (site 3 onset at 245 ms, 13 points, $ps < 0.05$, plus 2, $ps < 0.064$). Type effects started

TABLE 3 | *F*–values for significant effects of improvements (I) and type (T) in focal ANOVAs at specific lateral electrode pairs and times.

ERP	VPP	N3 complex									N400	P600	SW	
Site	FC 29–30	Fp 3–4			FC 29–30			OT 21–22			CP 47–48	P 53–54	FC 11–12	
Time (ms)	145–160	200–300	300–400	400–500	200–300	300–400	400–500	200–300	300–400	400–500	400–500	500–700	500–700	700–900
FOCAL OMNIBUS														
T	4.6^{+}	44.52**	90.5**	48.52**	14.92**	22.88**	23.22**	26.76**	46.99**	20.8**	5.54^{+}	30.65**	24.1**	141.8**
I	–	7.21 b*	14.12**	9.11**	–	6.13^{+}	11.56**	7.7 b*	–	–	15.05**	47.26**	8.1a*	12.91**
T × H	–	–	–	–	–	–	–	–	–	–	–	4.95^{+}	–	–
I × T	–	5.88^{+}	9.31**	5.55^{+}	5.78^{+}	–	7.12^{+}	–	–	–	6.78^{+}	5.78^{+}	–	4.9^{+}
I × T × H	–	–	–	–	–	–	–	4.47^{+}	–	–	–	–	–	5.44^{+}
FOCAL PAIRWISE CONTRASTS														
Real														
I	–	12.76**	21.28**	16.66**	–	9.96**	15.46**	4.55^{+}	–	–	17.07**	33.65**	10.3**	21.87**
Pseudo														
I	–	–	–	–	–	–	–	–	–	–	–	11.67**	–	–
I × H	–	–	–	–	–	–	–	–	–	–	–	–	5.86^{+}	–
Less Improvished (LI)														
T	–	26.86**	45.88**	27.58**	16.68**	18.65**	18.52**	11.44**	20.55**	9.06**	8.85**	37.33**	14.1**	78.14**
T × H	–	–	–	–	–	–	–	5.46^{+}	–	–	–	–	–	–
More Improvished (MI)														
T	–	17.03**	15.17**	–	4.8^{+}	4.73^{+}	7.27^{+}	37.88**	39.8**	24.27**	–	–	8.98**	55.92**

*Hemisphere (H). df(I, 18). Times rounded to nearest 5 ms. Fp, Frontopolar; FC, Frontocentral; OT, Occipitotemporal; CP, Centroparieta1; P, Parietal. –$p > 0.05$; $^{+}p < 0.05$; **$p < 0.01$; correction for multiple comparisons: alpha was 0.025 for 145–160 ms at pairs 29–30 and 33–34 and 500–700 ms at pairs 53–54 and 11–12 (a* $p < 0.0250$); alpha was 0.0167 for 200–300 and 300–400 ms at pairs 3–4, 29–30, and 21–22 (b* $p < 0.0167$); alpha was 0.0125 for 400–500 ms at pairs 3–4, 29–30, 21–22, and 47–48 (c* $p < .0125$).*

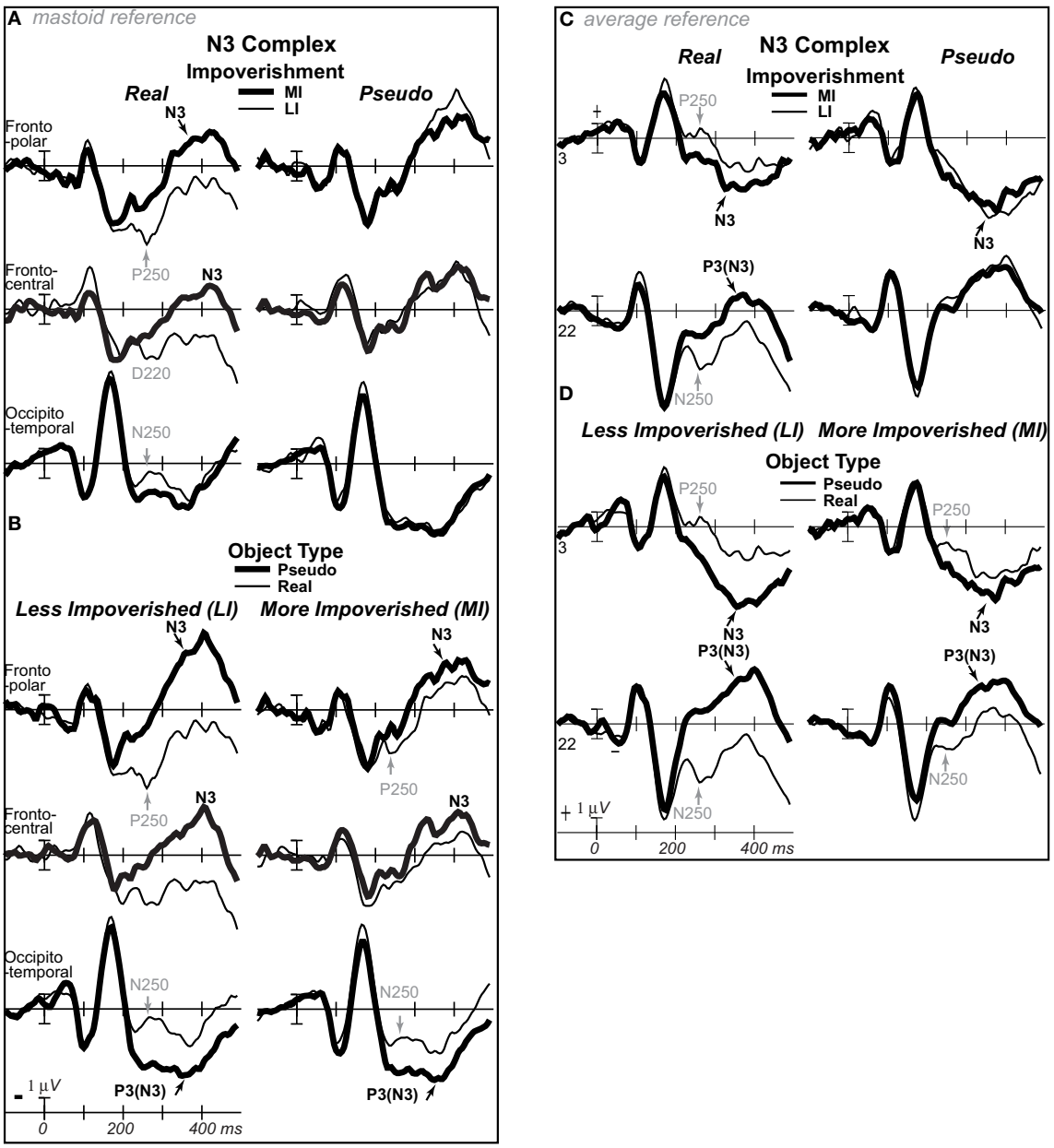


FIGURE 5 | ERP effects of impoverishment and object type on the N3 complex. (A,B) Shown are sites of the N3 complex maxima (left frontopolar site 3, right frontocentral site 30, right occipitotemporal site 22). Frontal effects inverted polarity to positivity at occipitotemporal sites, especially on the right ("P3[N3]" maximal at site 22), including an N250; note, a D220 index of task difficulty for decisions also inverted polarity between frontocentral and occipitotemporal sites. **(A)** N3 effects of impoverishment shown for real objects and pseudo objects. The frontal N3 showed an impoverished-real-object effect, including a frontopolar P250 component: The frontal N3 components were more negative for MI than LI real objects but not pseudo objects; note, the N3 showed no such effect for pseudo objects, but, in contrast, briefly at the peak, the N3 was instead slightly more negative for LI than MI pseudo objects. The occipitotemporal N250 but not later posterior N3 counterparts showed impoverishments effects for real objects. **(B)** N3 effects of object type shown on LI and MI trials. The N3 complex was larger for real than pseudo objects, and this type effect was larger on LI than MI trials. **(C,D)** To compare with other publications, the reference was computed using the average of all scalp sites (i.e., "common average reference"), and ERPs were plotted positive up. Shown are left frontopolar site 3 and occipitotemporal site 22. **(C)** N3 effects of impoverishment shown, for real and pseudo objects. **(D)** N3 effects of object type shown on LI and MI trials. Here, with the average reference, the effects over occipitotemporal sites become larger than when the bilateral mastoid reference is used instead (see **A,B**): Notice the similarity of effects between frontopolar site 3 in **(A,B)** and occipitotemporal site 22 here [also site 22 in **(A,B)** is more like site 3 here]. Crucially, the frontopolar ERPs with a mastoid reference [e.g., P250, N3 in **(A,B)**] correspond, with the average reference shown here, to the occipitotemporal ERPs (e.g., N250, P3(N3) at site 22 here). This demonstrated a clear link between the present and prior research on the frontocentral N3 complex and its subcomponents, and prior research on the occipitotemporal N250 and Ncl, which were defined using the nose or average reference,

(Continued)

FIGURE 5 | Continued
as shown here; note scalp distribution shapes with nose and average reference are similar. Like the frontopolar P250/N3 with the mastoid reference (see **A,B**), here with an average reference, the occipitotemporal N250 and P3(N3) show the impoverished-real-object effect, being more positive for MI than LI real objects but not pseudo objects, and this effect inverts polarity over frontopolar sites to P250 and N3 effects. Further, like the frontopolar P250 and N3 with the mastoid reference (see **A,B**), here with an average reference, the occipitotemporal N250 and P3(N3) show object type effects, being more positive for pseudo than real objects on LI and MI trials, and these effects invert polarity over frontopolar sites. The whole head ERPs in **Figure 4** demonstrate that this polarity inversion of effects occurs between frontal sites toward the midline (3–4, 11–12, 19–20, 29–30) and more lateral occipitotemporal sites with a right hemisphere maximum (22, 32, 34), especially for the N250, consistent with the known right lateralization of the N250 (i.e., N250r).

around the same time posteriorly regardless of impoverishment but ∼50 ms later on the frontopolar N3 for MI relative to LI stimuli: It started for LI stimuli between 230 and 250 ms (all sites) and, for MI stimuli, from 215 to 220 ms at occipitotemporal site 22 and later at 270 ms at frontopolar site 4 (14 consecutive points) and 280 ms at frontopolar site 3 (7 points, ps < 0.019, plus 1, $p = 0.051$). Altogether, these onsets suggest that impoverishment starts to modulate knowledge around the time when knowledge starts to contribute to the category decision: ∼250 ms.

Cortical Sources

For the four difference waves (**Figure 7**), cortical sources were estimated. The main focus was the time of the N3 peak from 300 to 400 ms (**Figures 8A–D**). Sources of this impoverishment effect (MI vs. LI) for real objects localized to occipitotemporal and lateral prefrontal areas found with fMRI (Ganis et al., 2007), whereas, for pseudo objects, impoverishment differences localized only to prefrontal areas. Sources of the object type effects (real vs. pseudo) on both LI and MI trials were in occipitotemporal areas. Sources at other times were also estimated. At all times after 200 ms, type effects continued in the same occipitotemporal areas (**Figures 8C,D,G,H**). Impoverishment sources varied over time and with object type (**Figures 8A,B,E,F**). The 200 to 300 ms time during the P250/N250 component showed the same impoverished-real-object pattern of sources as the peak N3 time period. Later, from 400 to 500 ms when the N3 ends and the N400 peaks, impoverishment effects for real objects showed only the occipitotemporal source (see intracranial ERP in **Figure 8A**). Around 450 ms, the maximum source shifted to anterotemporal cortex for both real and pseudo objects, suggesting an additional contribution from this region to the N400. From 500 to 700 ms, the estimated intracranial ERP for the anterotemporal source resembled the scalp P600 impoverishment waveform, which is maximal at this time, and more mediotemporal sources also contributed (**Figures 8E,F**). From 700 to 900 ms when the late SW dominates, anterotemporal impoverishment activity continued only for real objects. In addition, for both object types, impoverishment effects now appeared in the posterior cingulate cortex (PCC; **Figures 8E,F**).

Later ERPs Related to RTs

For completeness and because RTs occurred after the SW, cortical dynamics closer to the motor response were also assessed. EEG was re-analyzed to reject artifacts both between 900 and 1400 ms post-stimulus and during a pre-stimulus baseline of −100 to 100 ms. Analysis times from 900 to 1099 ms captured most MI real object RTs, and 1100 to 1400 ms captured most MI pseudo object RTs. Results showed anterior SW effects of impoverishment continued until 1099 ms and type until 1400 ms. Greater positivity was also found on a left mid-occipital-parietal slow wave (pSW) for MI than LI real objects from 900 to 1400 ms, which inverted polarity anteriorly, and the pSW showed type effects for MI trials until 1099 ms (**Figures 9A,B**). Critically, no impoverishment by type interactions were found after 900 ms. Both times showed main effects of type and impoverishment laterally, and type at midline sites (Fs > 10.70, ps < 0.005), and type and impoverishment each interacted with lateral electrode (Fs > 4.33), type with midline electrode and with lobe (Fs > 29.33), and impoverishment with midline electrode by lobe (Fs > 28.76), ps < 0.003. From 900 to 1099 ms, results also showed interactions of impoverishment by hemisphere, by midline electrode, by lobe (Fs > 5.4), by electrode by hemisphere ($F = 2.19$), ps < 0.04, and by Type by midline electrode ($F = 9.74$, $p = 0.006$). Focal simple effects tests on frontal SW pair 11–12 showed all impoverishment and type effects were significant from 900 to 1099 ms and both type effects from 1100 to 1400 ms (Fs > 4.51, ps < 0.05). Parietal pair 51–52, where the pSW was large, showed impoverishment by hemisphere for real objects from 900 to 1400 ms (Fs > 5.22), and type on MI trials from 900 to 1099 ms ($F = 4.68$), ps < 0.05.

A correlation analysis across subjects explored the relationship between RTs and impoverishment effects at pSW parietal pair 51–52 from 900 1400 ms. Results showed that RT and ERP impoverishment effects from 900 to 1099 ms for real objects correlated significantly for the pSW effect at both 51 and 52 (rs > 0.43, ps < 0.035). From 1100 to 1400 ms, RT and ERP impoverishment effects for pseudo objects correlated at site 52 ($r = 0.473$, $p = 0.02$). As the pSW became more positive, RTs became slower (**Figure 9C**).

sLORETA on this data revealed brain sources from 900 ms onwards (**Figure 9D**) in supplementary motor area (SMA), which was activated in fMRI (Ganis et al., 2007).

Fragmentation Level ERPs

The results so far used the median split of RTs to define MI and LI conditions. In a separate analysis of ERPs until 900 ms, fragmentation levels 4–5 defined the MI condition and levels 2–3 defined the LI condition (**Figure 10**). Results of the fragmentation level analyses replicated all results from the RT split analyses. It may be noted that impoverishment effects for real objects were slightly smaller with fragmentation level defining impoverishment, but this would be expected. After all, the most and least fragmented images were excluded

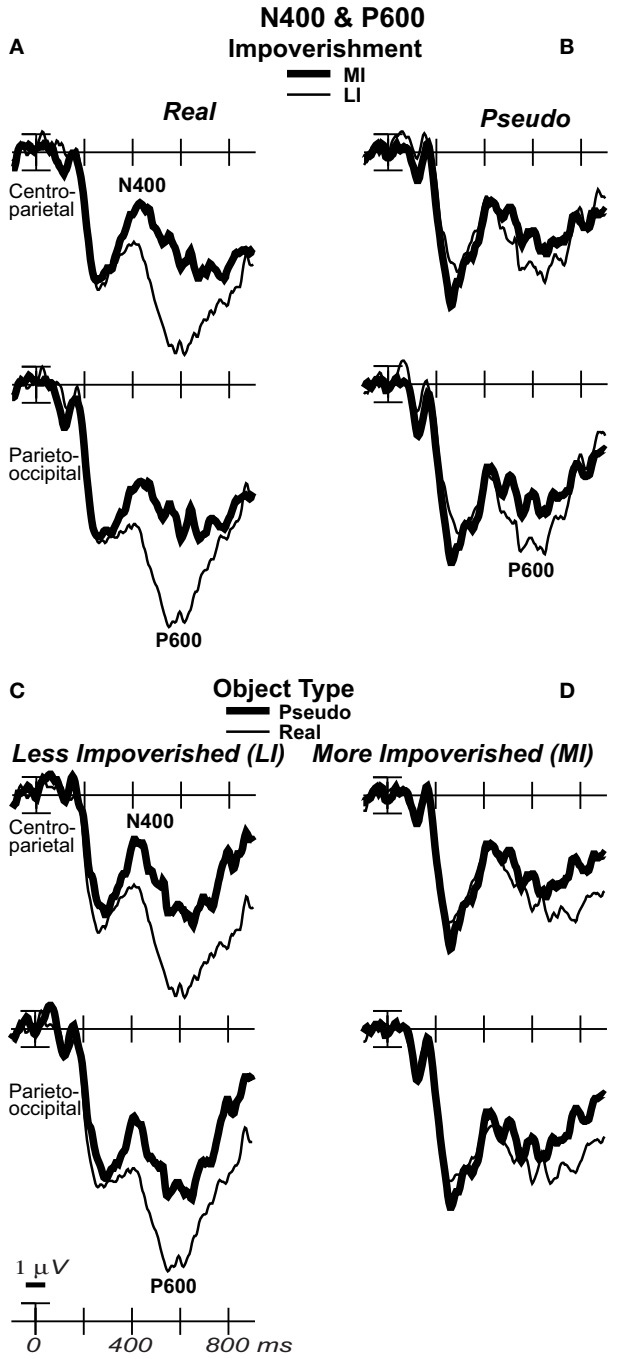


FIGURE 6 | Effects of impoverishment and object type on the N400 and P600. Grand average ERPs at focal sites of the centroparietal N400 and parietal P600 plotted negative up. N400 and P600 impoverishment effects shown for **(A)** real objects and **(B)** pseudo objects. **(C)** N400 and P600 object type effects shown on LI and **(D)** MI trials, which showed no type effect. From 400 to 700 ms, impoverished-real-object effects were found on the N400 and P600. Positivity was greater on LI than MI trials, and this impoverishment effect was larger for real than pseudo objects, which showed no such effect on the N400. The P600 was the first ERP to show impoverishment effects for both real and pseudo objects and in the same direction.

from this fragmentation based analysis but included in the RT based analysis and so stimulus differences were smaller with fragmentation instead of RT defining impoverishment. Further, as RT must completely capture all stimulus impoverishment that affects RTs, impoverishment effects should be larger for results based on RTs than any single factor such as fragmentation.

From 200 to 400 ms, the critical impoverishment by type interaction was found (200–300 ms: x lateral electrode, $F = 4.5$, $p = 0.002$; 300–400 ms, $F = 16.1$, $p = 0.001$; x lateral electrode, $F = 7.98$, $p < .001$; 300–400 ms: midline, $F = 7.15$, $p = 0.015$; x electrode, $Fs > 8.54$, $ps < 0.009$), as well as Type and Impoverishment main effects and/or their interactions with scalp site (200–300 ms: $Fs > 3.04$, $ps < 0.02$; midline, $Fs > 4.95$, $ps < 0.04$; 300–400 ms: $Fs > 2.13$, $ps < 0.04$; midline, $Fs > 8.41$, $ps < 0.01$). Focal results at frontopolar N3 pair 3–4 showed effects of type ($Fs > 35$, $ps < 0.001$) and impoverishment by type ($Fs > 5.07$, $ps < 0.04$), and simple effects tests showed impoverishment for real objects from 300 to 400 ms ($Fs > 8.59$, $ps < 0.009$), impoverishment for pseudo objects from 200 to 400 ms ($Fs > 4.98$, $ps < 0.04$), and type on LI and MI trials from 200 to 400 ms ($Fs > 6.15$, $ps < 0.03$). Occipitotemporal pair 21–22 showed type effects from 200 to 400 ms ($Fs > 5.93$, $ps < 0.03$), and, from 200 to 300 ms, impoverishment by type ($F = 5.84$, $p = 0.026$). From 300 to 400 ms, the frontocentral N3 (pair 29–30) showed effects of type ($Fs > 6.85$, $ps < 0.02$), and impoverishment by type ($Fs > 17.97$, $ps < 0.001$), and simple effects tests showed effects of impoverishment for both objects ($Fs > 6.85$, $ps < 0.02$), and type on LI trials ($Fs > 25.3$, $ps < 0.001$).

From 400 to 700 ms, the critical impoverishment by type interaction was found on the P600 ($Fs > 16.61$, $ps < 0.001$; x lateral electrode, $Fs > 2.82$, $ps < 0.02$; midline, $Fs > 6.86$, $ps < 0.02$; x electrode, $Fs > 5.17$, $ps < 0.03$), as well as Type and Impoverishment main effects and their interactions with electrode ($Fs > 4.81$, $ps < 0.05$; midline, $Fs > 7.84$, $ps < 0.02$). Focal results from 400 to 500 ms showed type and impoverishment main effects at frontopolar pair 3–4 ($Fs > 25$, $ps < 0.01$), and a marginal impoverishment by type interaction ($F = 3.82$, $p = 0.066$), and occipitotemporal pair 21–22 showed a type effect ($F = 15.71$, $p < 0.001$). The frontocentral N3 (pair 29–30) showed effects of type ($Fs > 12$, $ps < 0.003$), and impoverishment by type ($Fs > 10.79$, $ps < 0.005$), and simple effects tests showed impoverishment for real objects, and type for LI trials ($Fs > 4.56$, $ps < 0.048$). Focal results from 500 to 700 ms at P600 (pair 55–56) showed effects of type ($Fs > 56.95$, $ps < 0.001$), impoverishment by type ($Fs > 25.58$, $ps < 0.001$), and type by hemisphere ($F = 6.97$, $p = 0.017$), and simple effects tests showed impoverishment effects for both types, and type on only LI trials ($Fs > 13.22$, $ps < 0.002$).

From 700 to 900 ms, the critical impoverishment by type interaction was also found (x lateral electrode, $Fs = 2.62$, $p = 0.044$; midline, $Fs = 7.89$, $p = 0.012$; x lobe, $F = 9.9$, $p = 0.006$), as well as Type and/or Impoverishment main effects and their interactions with electrode ($Fs > 2.16$, $ps < 0.03$; midline, $Fs > 4.88$, $ps < 0.04$). Focal results at frontocentral pair 11–12 showed effects of type ($F = 137.35$), $p < 0.001$), and impoverishment by type ($F = 5.33$, $p = 0.04$).

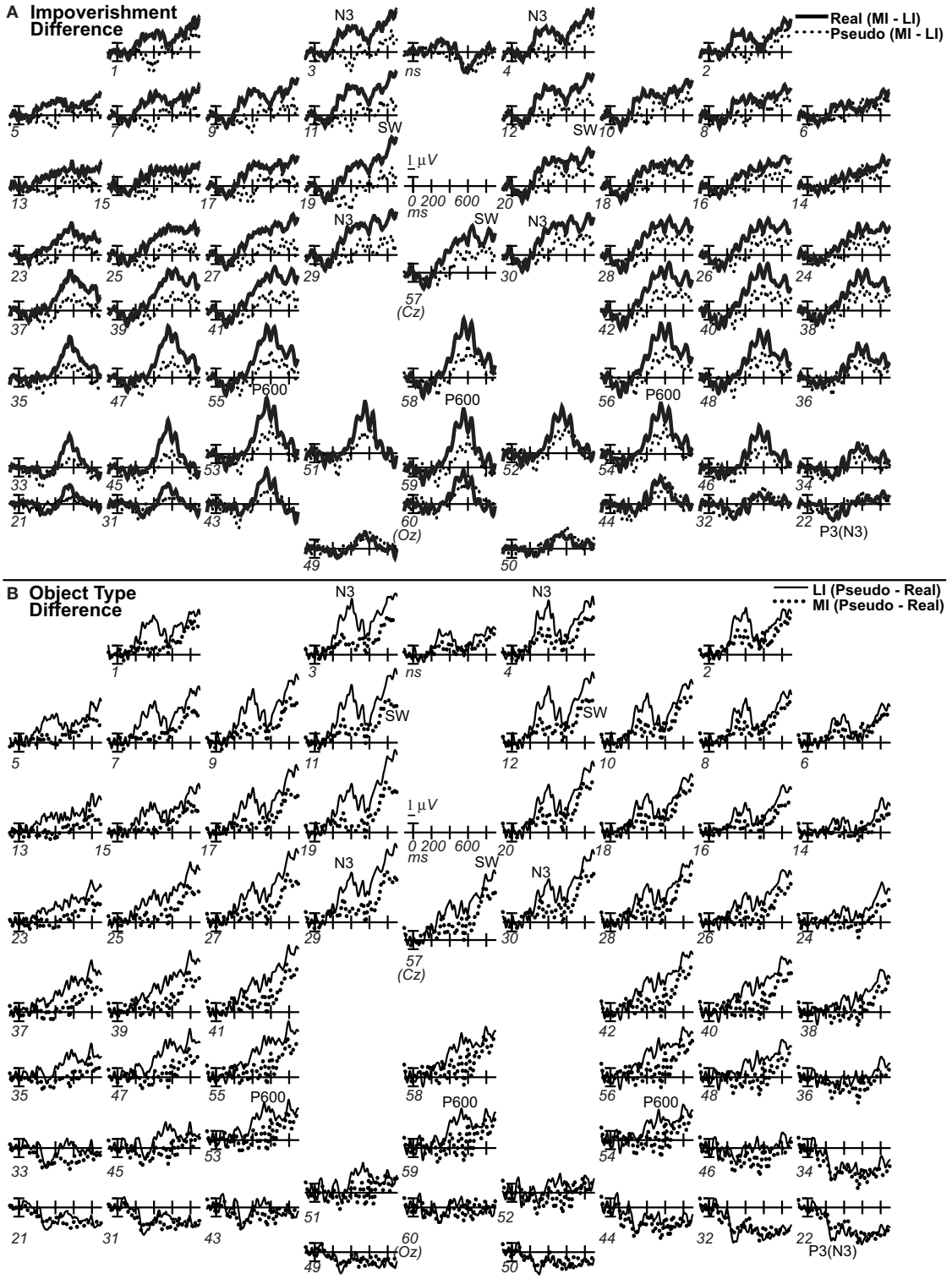


FIGURE 7 | Grand average difference ERPs computed by subtracting ERPs in two conditions. For display, waves were low pass filtered at 20 Hz. **(A)** *Difference waves of impoverishment effects*. Effects of impoverishment shown by subtracting the less impoverished (LI) condition from the more impoverished condition (MI). Up is negativity in MI greater than LI. Note, where the impoverishment difference wave was greater for real than pseudo objects reveals the impoverished-real-object effect. **(B)** *Difference waves of object type effects*. Effects of object knowledge shown by subtracting the real object condition from the pseudo object condition. Up is negativity for pseudo greater than real objects.

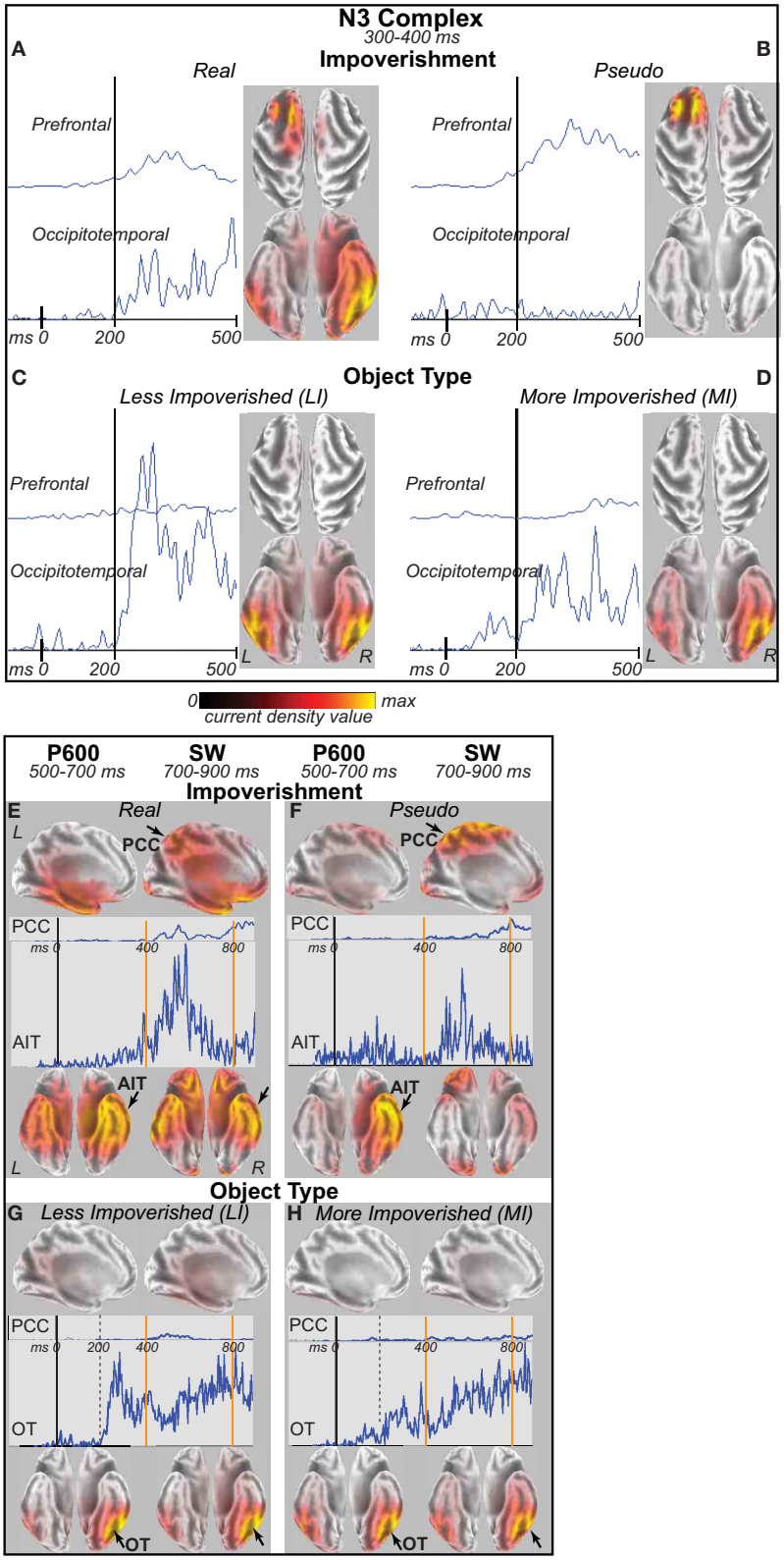


FIGURE 8 | The sLORETA maps show estimated sources of the difference waves (Figure 7) between two conditions (impoverishment = MI minus LI; type = pseudo minus real) in the grand average ERPs. Maps shown superimposed on an inflated, canonical MNI152 (Colin) brain. Dark areas are sulci; light

(Continued)

FIGURE 8 | Continued
areas are gyri. L, left hemisphere; R, right hemisphere. Each brain shows standardized cortical current density distributions, and source activity reflects the location of differential source activity between conditions but not the direction of effects. Scale uses hot colors (red, yellow) for maximal current density value differences. **(A–D)** *N3 Sources*. sLORETA maps shown for the N3 from 300 to 400 ms on dorsal (top) and ventral (bottom) cortical surfaces. Estimated intracranial ERPs plotted on the left for prefrontal (MNI *x y z* coordinates −15 20 65) and occipitotemporal sources (55 −45 −25) between −100 and 500 ms. **(A)** *N3 impoverishment sources for real objects*. Occipitotemporal sources: inferior (BA 20, 60 −40 −20; BA 37, 55 −45 −25) and middle temporal (BA 21, 65 −35 −15; BA 20, 55 −40 −15), fusiform (BA 37, 50 −50 −25; BA 20, 55 −35 −25; BA 19, 45 −70 −20; BA 36, 45 −40 −25), middle occipital (BA 19, 50 −70 −15), lingual (BA 18, 15 −85 −20), and parahippocampal (BA 36, 40 −30 −25) gyri. Prefrontal sources: superior (BA 6, −15 20 65; BA 8, −25 30 55), middle (BA 6, −25 20 60; BA 9, −35 40 40), and inferior frontal (BA 47, 20 25 −20) gyri. **(B)** *N3 impoverishment sources for pseudo objects.* Same prefrontal sources as for real objects. **(C)** *N3 object type sources for LI.* Occipitotemporal sources: fusiform (BA 37, 55 −60 −20, −50 −60 −25; BA 36, 45 −40 −30; BA 19, −50 −70 −20), inferior temporal (BA 20, 50 −55 −20; −60 −55 −20), middle temporal (BA 37, 55 −55 −15, −55 −65 −15; BA 21, 65 −50 −10), middle occipital (BA 37, 50 −65 −15, −50 −65 −15; BA 19, 50 −75 −15), parahippocampal (BA 19, 35 −45 −10) gyri. **(D)** *N3 object type sources for MI.* Same occipitotemporal sources as for LI. **(E–H)** *P600 and slow wave (SW) Sources*. sLORETA maps shown for left medial (top) and ventral (bottom) cortical surfaces. OT, occipitotemporal cortex; AIT, anterior inferior temporal cortex; PCC, posterior cingulate cortex, including precuneus and cuneus. Estimated intracranial ERPs plotted for the voxel showing maximum impoverished-real-object effects from 300 to 400 ms (same as later) in OT (55, −45, −25), 500 to 700 ms in AIT (25, 0 −45), and 700 to 900 ms in PCC (0 −55, 65). **(E)** *Late impoverishment sources for real objects.* AIT sources (maximum BA 20, 25–30 −5 −45) occurred from 450 to 700 ms when the P600 peaks: middle (BA 21, 65 −30 −20) and inferior temporal (BA 20, 60 −35 −20), fusiform (BA 20, 55 −35 −25), parahippocampal (BA 36, 35 −25 −30; BA 35, 30 −25 −25), and other limbic structures (BA 20, 25 0 −45; BA 38, 25 5 −45; BA 36, 25 −5 −40; BA 28, 25 −10 −35). From 500 to 700 ms, limbic lobe dominated (BA 20/38, 25 0 −45), including parahippocampal gyrus (BA 35, 25 −15, −30). From 700 to 900 ms, impoverishment effects in anterotemporal cortex continued and appeared in medial posterior cortex around cingulate (BA 25, 0 5 −10; BA 31, −10 −45 40), cuneus (BA 17, 5 −100 −5), and precuneus (BA 7, 5 −60 65; −5 −50 50), and occipital extrastriate regions (BA 18, 0 −95 −15). The SW effect in PCC is active after 700 ms. **(F)** *Late impoverishment sources for pseudo objects.* P600-like wave in AIT and SW in PCC shown. **(G)** *Late object type sources for LI*, and **(H)** *for* MI: Occipitotemporal cortex only.

Discussion

Altogether, a hybrid account that combines the MUSI framework, parietal-prefrontal PHT theories of vision, and decision theories best explains the findings (**Table 1** Results). Overall, the ERP time course indicates that knowledge and impoverishment modulate ERPs from 200 to 900 ms, all of which show the impoverished-real-object effect: the N3, centroparietal N400, parietal P600, and a late SW; note, as effects on the LPC were not distinguishable from the P600 and SW, henceforth we do not discuss the LPC. Earlier ERPs and later ERPs from 900 to 1400 ms provide no evidence of this effect, and later ERPs correlate with RTs and reflect supplementary motor cortex activity.

Early ERPs before 200 ms

Early ERPs are most consistent with the MUSI framework. Early ERPs before 200 ms show no impoverishment nor impoverished-real-object effects. Before 200 ms, there was no evidence that impoverishment affects activation of object knowledge, and the VPP/N170 showed only a small type effect (**Table 1**). However, a type effect would likely reflect low-level feature differences due to using a subset of pseudo versions of the real objects in order to keep decision rates around 50%; in contrast, prior work compared the full set of matched real and pseudo objects across two experiments and three tasks, finding no ERP differences until after 175 ms and none on the VPP/N170 (Schendan et al., 1998). More important, the VPP/N170 showed no impoverishment effect and no impoverished-real-object effect; note, sensory differences between MI and LI stimuli may have been too small and variable to be detected here. Thus, we found no evidence for early impoverishment effects, and only a small type effect likely reflecting spurious low-level sensory differences due to not using the full set of matched stimuli here. Other studies have also not found early impoverishment effects with these fragmented line drawings, even when level is held constant (e.g., Doniger et al., 2000; Schendan and Kutas, 2002; Schendan and Maher, 2009). With overall no evidence for early impoverishment and type effects independently, it is thus not surprising to find no evidence for an early impoverished-real-object effect. We are thus confident that early effects of impoverishment, type, and their interaction are minimal to none, in general.

Note, early top-down models that involve biasing attention (**Figure 1F**) assume a cue, context, or target determines task-relevant information, whereas the present task provided no such biasing signal, minimizing such top-down influences early on and consistent with no such evidence here. In real life, context may provide cues about object identity, but, when objects are categorized in scene contexts, similar to real life situations, the N3 complex shows the earliest context effect, not earlier ERPs (Ganis and Kutas, 2003). Possibly only strong, effortful, strategic biased attention would affect early visual processing, as when people visualize features mentally and effortfully before the picture appears, early VPP/N170 modulation is observed (Ganis and Schendan, 2008) and could be expected to be enhanced by impoverishment.

Late ERPs after 200 ms

Together, both early and later ERPs indicate that object cognition starts after initial bottom-up activation of the ventral stream. Only the MUSI framework, not other vision or decision theories, can explain this pattern. The rest of the discussion thus focuses on later effects and interpretation based on the full ERP time course. While early ERPs best fit the predictions of the MUSI framework, later ERPs best fit the predictions of parietal-prefrontal PHT theories, though MUSI, decision, and prefrontal theories can accommodate the results (**Table 1**). Thus, a hybrid MUSI account that combines MUSI with parietal-prefrontal PHT and decision theories best explains the findings.

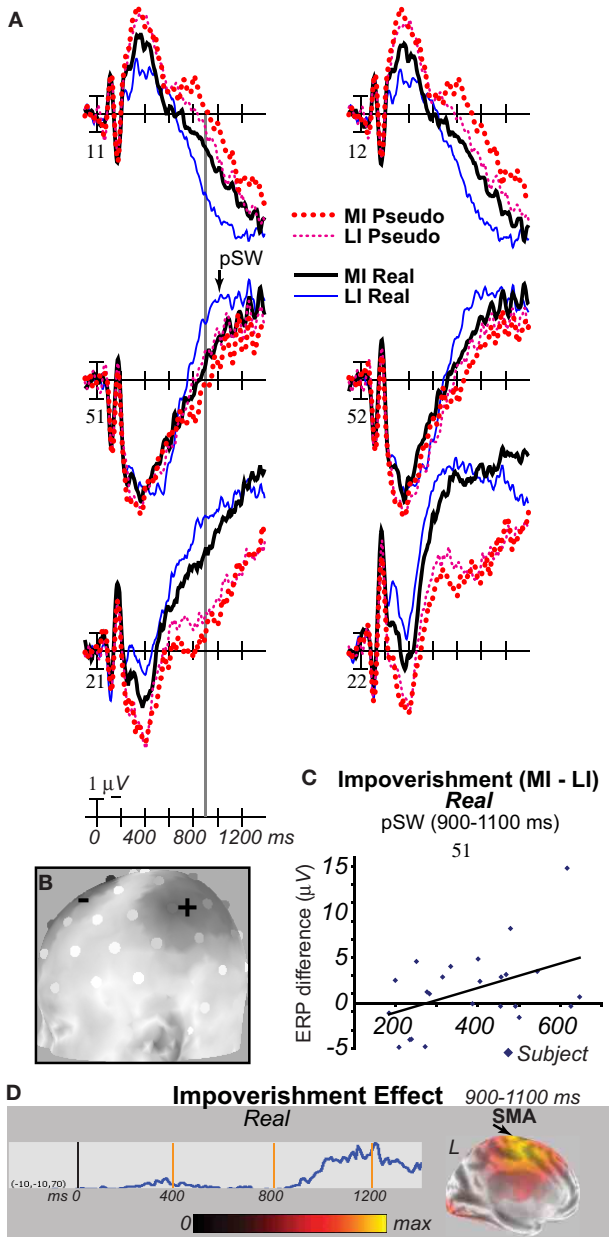


FIGURE 9 | Late ERP slow waves (SW) show main effects of impoverishment and object type after 900 ms until response, localize to supplementary motor area (SMA), and correlate with RT effects. (A) Grand average ERPs show effects of impoverishment (more [MI], less [LI] impoverished) and object type. ERPs at lateral sites of the SW (11–12), a posterior slow wave (pSW; 51–52), and type effects (21–22) are plotted negative up. Image type and impoverishment modulated distinct ERPs even after 900 and until the latest responses around 1400 ms for MI pseudo objects. An impoverished-real-object effect on a late pSW started after 900 ms (gray line). **(B)** Voltmap generated using sLORETA (default left mastoid reference) shows the distribution of voltage differences over the left hemisphere from 1100 to 1400 ms when only the SMA effect occurs; the distribution is similar from 900 to 1100 ms. Electrodes symbolized by half spheres. The + sites are where the pSW effect is strongest (MI - LI), whereas—sites are the location of the SW over frontal scalp (LI - MI). **(C)** Across subjects, the RT difference correlated significantly with the late pSW effect. Each diamond plots the RT and ERP values for each subject. RT difference on x-axis. ERP amplitude difference on y-axis. The computed linear regression line (solid) is shown. Impoverishment difference for real objects (MI minus LI) from 900 to 1100 ms at site 51 correlated such that, on MI relative to LI trials, as the pSW became more positive, RTs got slower. **(D)** Maps from sLORETA for 900 to 1100 ms on the left (L) medial surface show the late SMA (BA 6, −15 −10 55; 5 −5 65) impoverishment effect for real objects, extending into anterior cingulate gyrus (BA 24, −15 −10 50), and estimated intracranial ERPs show the SMA effect started after 900 ms. Specifically, from 900 to 1100 ms, sources of impoverishment effects for real objects continued in striate/extrastriate and anterior temporal cortex, and, for the first time, were located in left more than right SMA and anterior cingulate, and this effect appeared to correspond to the pSW. From 1100 to 1400 ms, the SMA effect continued, but extended dorsally into superior frontal gyrus (BA 6, −10 [−10 or −15] 70), and posterior effects were minimal or none. At these times, the impoverishment effect for pseudo objects localized to striate/extrastriate areas (BA 17/18) with weaker sources in temporal pole (BA 38, −40 20 −35) and inferior frontal gyrus (BA 47, −50 45 −10). Note, the sLORETA map shows estimated sources of the difference wave (MI - LI) in the grand average ERPs superimposed on an inflated, canonical MNI152 brain (Colin); dark areas represent sulci; light areas represent gyri. The depicted brain shows standardized cortical current density distributions, and source activity reflects the location of differential source activity between conditions but not the direction of effects. Scale shows yellow represents maximal current density value differences. Estimated intracranial ERPs from −100 to 1400 ms were extracted from the voxel showing maximum impoverished-real-object effects at MNI coordinates from 900 to 1400 ms in SMA (−10 −10 70). Solid tics mark the 0 ms stimulus onset and 400 ms intervals post-stimulus.

Knowledge

Real objects activate knowledge more than pseudo objects so type effects reveal the time course of knowledge activation. The frontal N3, N400, P600, and SW are more negative, and occipitotemporal counterparts of the N3 are more positive for real than pseudo objects, and these effects localize to occipitotemporal cortex. All these ERPs show type effects for LI stimuli, and N3 type effects start at 230 ms for LI stimuli. Altogether, these results replicate evidence for knowledge effects with fully intact (i.e., LI) pictures of real and pseudo objects on these ERPs, starting from 175 to 218 ms during the N3 and continuing onwards (Holcomb and McPherson, 1994; Schendan et al., 1998; McPherson and Holcomb, 1999; Gruber and Müller, 2005, 2006).

While the MUSI account might seem counter to ultra-rapid categorization and other early categorization findings before 150 ms, they are actually compatible. Consider the following. First, for example, eye movement findings (Kirchner and Thorpe, 2006) during ultra rapid categorization suggest an onset at the earliest possible time of around 124 ms when there are more correct than wrong responses. This time matches the 120 ms onset of categorical perception of objects and early object perception processes on the VPP/N170 during State 1 (Schendan et al., 1998; Schendan and Lucia, 2010). When behavior (e.g., a saccade) can be performed based on information from the initial bottom-up pass, then it will be carried out. However, this is a rare occurrence. The same eye movement findings during ultra rapid categorization suggest a mean minimum saccade RT of around 150 ms, and median saccade RT of around 228 ms, varying widely

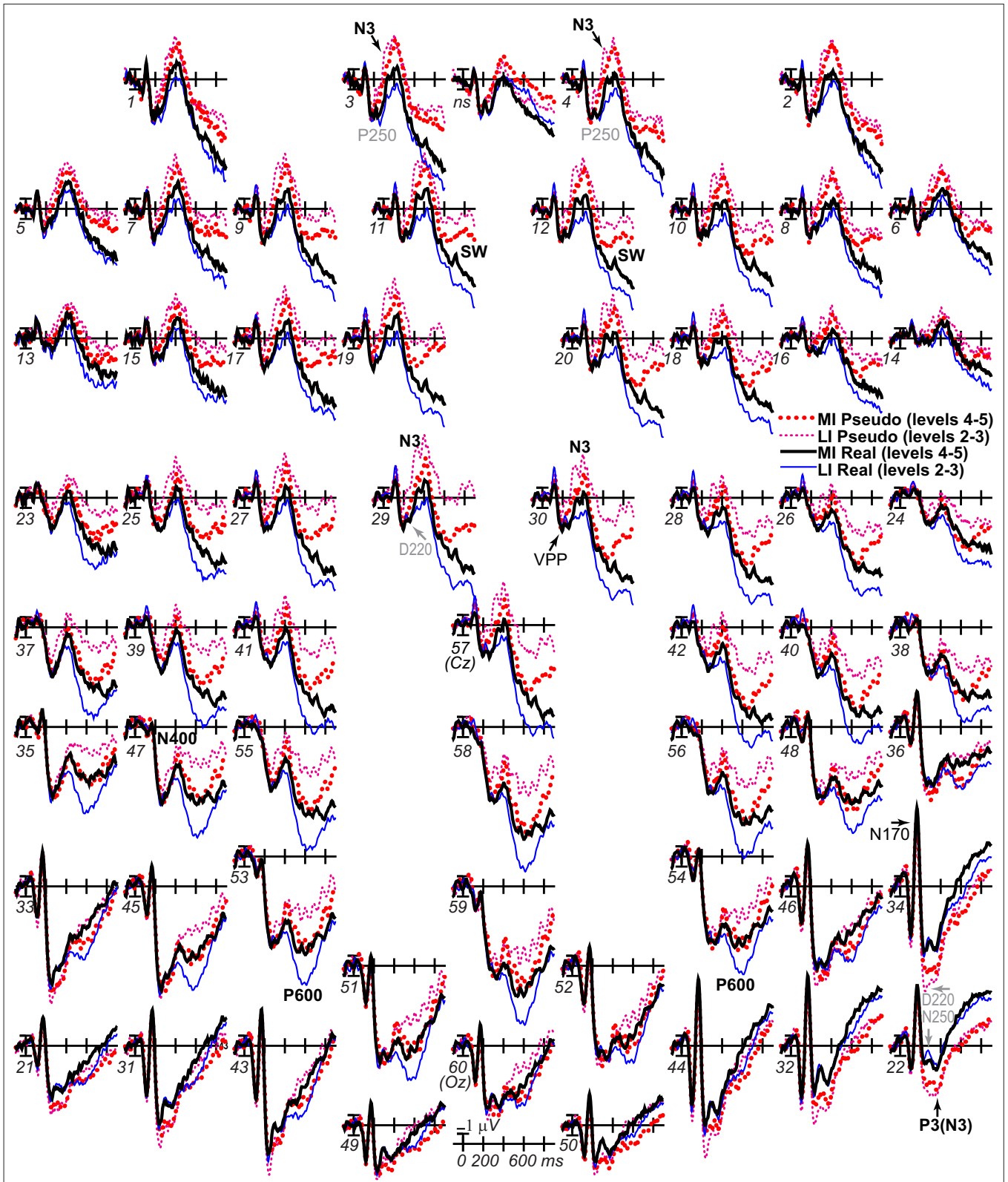


FIGURE 10 | ERP effects of impoverishment (by fragmentation level) and object type. Same as **Figure 3**, except impoverishment is defined using picture fragmentation level: LI is levels 2–3; MI is levels 4–5. Results from this analysis replicate those in **Figure 3** which used the median split of RTs to define impoverishment.

between people, from 159 to 301 ms. Thus while it is tempting to focus on the onset, it is more informative for most visual cognitive phenomena to realize that the fastest times represent a special case of the minimum speed of initial (low-level) visual feed forward processing that is sufficient to enable a decision and motor response (Kirchner and Thorpe, 2006). Most visual input and task situations require more time. Indeed, even ultra rapid categorization tasks pinpoint a typical onset of categorization of 150–228 ms or longer. In addition, longer time (e.g., 150–230 ms or longer) is associated with greater accuracy, even for eye movements, and this additional time is thought to reflect iterative (i.e., interactive, resonant) decision and motor processes (Kirchner and Thorpe, 2006).

Second, as reviewed by Fabre-Thorpe (2011), rapid visual categorization is associated with N3 effects by 150 ms and minimal reaction time (MinRT) of 250 to 300 ms for superordinate categorizations in go-no go and two-category decisions. The shortest time of 220 ms can only be achieved with extensive training, that is, on animal/non-animal decisions with a single overlearned animal scene among novel scenes. No set of easy, trivial, or optimal stimuli can explain this short RT, and MinRT does not shorten even for the simplest geometric images of square vs. circle. However, slower RTs are associated with difficult images, and experience can reduce these RTs (Fabre-Thorpe et al., 2001), consistent with greater repetition priming effects behaviorally and on the N3 and P600/LPC for more than less impoverished categorized objects (Schendan and Kutas, 2003). Altogether, the evidence has led to the conclusion that the role of rapid visual categorization on behavior is limited because it is based on "coarse and unconscious (achromatic) visual representations automatically activated by the first available magnocellular information" that is processed along the ventral visual pathway (Fabre-Thorpe, 2011). Notably, basic level categorization (e.g., dog) yields slightly higher accuracy (4%) than superordinate categorization (e.g., animal), and MinRT is about 50 ms slower (Fabre-Thorpe, 2011). This suggests that, even at the fastest possible time, categorization at the basic relative to superordinate level requires additional processing time, which also achieves a higher decision accuracy. This is consistent with the finding that entry level categorization of new objects is typically associated with an N3 onset time around 200 ms, and repetition priming can reduce this by about 50 ms down to around 150 ms with canonical views, which are not impoverished (Schendan and Kutas, 2003; Schendan and Maher, 2009).

Third, it remains open whether low level feature search can explain the fastest times achieved. Ultra rapid categorization involves giving the subject the category to search for beforehand, making it essentially a visual search task (Treisman, 2006). Hence, before the trial, the visual system has been placed in a "top down presetting" state through feedback processes that prepares it to detect the task relevant features of the category (Enns, 2004; Fabre-Thorpe, 2011). Thus, if a feature of the input matches the top-down search target by 120–150 ms of visual processing, then this can be used to execute a motor behavior (i.e., a saccade), but this does not mean that entry level categorization, meaning, phenomenological awareness, or object cognition has yet occurred. All we know is that a sensori-motor program has been executed within 120 to 150 ms. The MUSI argument is that state 1 may be sufficient for a simple sensori-motor program to be executed based on categorical perception or feature detection (as in visual search), but actual entry level categorization, decision, cognition, and phenomenological awareness do not happen until State 2. What is driving the fastest times in ultra rapid categorization tasks is categorical perception, not actual cognitive categorization. Indeed, it is thought that the 120 ms minimum time for the saccade behavior during ultra rapid categorization tasks may be due to low level visual area V4, bypassing higher order visual areas, such as inferotemporal cortex, sending input directly to lateral inferior parietal cortex and then to frontal eye fields. The earliest 120 to 150 ms times essentially reflect a low level sensorimotor decision that bypasses semantics and even categorical perception, and "is just the start of a series of complex events involving feedback loops…to (generate) conscious perception" (Kirchner and Thorpe, 2006): This is essentially the interactive resonant activity posited in State 2 of the MUSI account.

Fourth, the VPP/N170 and earlier P1 and C1 are thought to reflect predominantly the initial fast feedforward pass through the visual system (e.g., **Figure 1A**), as well as reflexive feedback (**Figures 1B,C**), whereas later ERPs are dominated by feedback inputs (David et al., 2005, 2006). Thus feedback has the greatest role in cognition after the initial bottom-up pass.

Fifth, the ERP that shows the earliest effect in ultra-rapid categorization studies is the N3 complex, not the VPP/N170, and across a variety of ultra-rapid categorization studies the N3 is modulated between about 150 and 500 ms (Johnson and Olshausen, 2003, 2005). Interestingly, the onset of the original effect was between 152 and 171 ms (Thorpe et al., 1996). This onset is consistent with ERP findings estimating when entry level categorization starts, that is, between about 150 and 250 ms, modulating the N3. For example, canonical and unusual (impoverished) views of objects differ between 140 and 250 ms (Schendan and Kutas, 2003), repetition effects with canonical (best) views of real objects start to modulate the N3 between 148 and 172 ms (Schendan and Kutas, 2003), and repetition effects for fragmented drawings of real objects that are named correctly start by 192 ms (Schendan and Maher, 2009) or 248 ms (Schendan and Kutas, 2007a). Consistent with the MUSI account, the early part of the N3 effect from 190 to 215 ms on ultra rapid categorization tasks localizes to occipitotemporal cortex (Delorme et al., 2004; Fize et al., 2005), and intracranial ERPs localize ERPs between 200 and 400 ms to VLPFC and occipitotemporal cortex (Allison et al., 1999; Puce et al., 1999).

Sixth, the N3 is the first ERP that modulates with categorization success, not the VPP/N170 (Schendan and Kutas, 2002, 2003; Schendan and Maher, 2009). This indicates that object cognition, entry level categorization, and phenomenological awareness of the meaning of the object do not start until feedback interactions dominate processing, especially from anterior temporal or prefrontal cortex down to occipitotemporal cortex, as indexed by the N3 (Lamme, 2003; Schendan and Kutas, 2007a; Folstein and van Petten, 2008; Schendan and Maher, 2009; Clarke et al., 2011). The N3 is

more negative with less successful category decisions, greater top-down processes of mental imagery (Schendan and Lucia, 2009; Schendan and Ganis, 2012), greater image atypicality and impoverishment (Doniger et al., 2000; Schendan and Kutas, 2002, 2003; Johnson and Olshausen, 2003), and for new relative to repeated meaningful objects (i.e., repetition priming) (Henson et al., 2004; Schendan and Maher, 2009; Voss et al., 2010). The N3 typically inverts polarity somewhat over occipitotemporal sites, where effects are most prominent with a common average reference (known as N250, Ncl, or L1) and associated with category learning and implicit memory (Gruber and Müller, 2006; Scott et al., 2006; Sehatpour et al., 2006; Soldan et al., 2006). Critically, ERPs and corresponding single-trial EEG and fMRI show that category decision processes that distinguish between faces and objects happen during the N3 complex in state 2 but not on the earlier VPP/N170 in state 1 (Philiastides and Sajda, 2006, 2007; Philiastides et al., 2006; Ratcliff et al., 2009; Rousselet et al., 2011). On functional and spatiotemporal grounds, such work suggests a D220 component of the N3 from 220 to 300 ms (**Figures 3**, **5A**, **9**, **11B**) varies with impoverishment and task difficulty and reflects anterior cingulate, eye field, insula, and dorsolateral prefrontal activity, and a so-called "late component" of the N3 from 300 to 450 ms reflects decision processes in which VLPFC accumulates evidence from lateral occipital cortex (Philiastides and Sajda, 2006, 2007). For example, the N3 complex and both decision components have similar scalp distribution patterns: Both invert polarity between similar frontal and posterior locations. The role of prediction in visual search (Enns and Lleras, 2008) is consistent with the present finding that interactions or resonance between bottom-up and feedback processes contributes to object constancy and the incorporation of parietal-prefrontal PHT ideas into the MUSI account at state 2.

Impoverishment and Knowledge

Knowledge activates around 230 ms, and impoverishment and impoverished-real-object effects start around the same time (∼250 ms). These onsets are consistent with parietal-prefrontal PHT theory ideas that, when initial bottom-up activation (by ∼175–230 ms) cannot categorize the object well enough to make a decision about MI real objects (Serre et al., 2007a), additional processes start to be recruited (∼250 ms) that use knowledge in posterior areas to achieve the visual constancy of the category decision. Critically, impoverishment affects real objects the most; note, the flip side of the interaction is that LI stimuli activate knowledge the most effectively. This timing is consistent with the finding from category decision studies of a ∼50 ms onset range of single trial EEG discrimination between faces and cars when their phase coherence varies between 30 and 45% (Philiastides and Sajda, 2006). The fMRI and these ERP results are compatible with both (a) top-down processes in the parietal-prefrontal PHT variants (e.g., Ganis et al., 2007) and (b) bottom-up accumulation in decision theories (e.g., Philiastides and Sajda, 2007).

However, only parietal-prefrontal PHT variants predict the interaction (**Table 1**), and findings from ERP studies of mental imagery indicate top-down processes operate after 200 ms. Mental imagery, which can be mediated only by top-down processes, modulates both the N3 and SW but not the P600 and minimally so the N400 (Schendan and Ganis, 2012). Moreover, ERP mental imagery effects resemble the spatiotemporal characteristics and direction of the impoverishment effects; for example, the N3 and SW are most negative when the need for top-down processes for mental imagery and when impoverishment are greatest. In contrast, adaptation effects, which primarily reflect bottom-up processes, can show ERP effects in the opposite direction to mental imagery and impoverishment effects (Ganis and Schendan, 2008; Schendan and Ganis, 2012). We thus conclude that the N3 impoverished-real-object effect reflects interactive top-down and bottom-up activity that facilitates the category decision because only the N3 reflects visual object knowledge (as argued above) and shows the expected pattern of knowledge, impoverishment, and decision effects across many studies that are predicted by PHT and decision theories.

Accordingly, the N3 impoverishment effects localize to lateral prefrontal cortex (LPFC), and, for real objects only, localize also to the same occipitotemporal region as knowledge activity. This is consistent with the MUSI proposal that the N3 complex reflects interactive activity between VLPFC and occipitotemporal cortex for model selection from object knowledge. After 450 ms, the N400/P600 impoverishment effects for real objects localize to anterior inferior temporal cortex and the mediotemporal lobe, consistent with intracranial studies showing memory effects in anterior mediotemporal lobe that resemble modulations of late posterior positivities on the scalp (Halgren et al., 1995; Guillem et al., 1999; Trautner et al., 2004). After 700 ms, SW impoverishment effects for both object types localize to PCC. As impoverished-real-object effects and their implications change over time, each ERP finding is next discussed in detail separately.

N3 complex

The N3 complex shows the earliest impoverishment and impoverished-real-object effects. These findings are consistent with prior evidence of impoverishment or category decision effects on only later ERPs, not at earlier times before ∼150–200 ms (Doniger et al., 2000; Schendan and Kutas, 2002, 2003; Johnson and Olshausen, 2003, 2005; Philiastides and Sajda, 2006, 2007; Philiastides et al., 2006; Sehatpour et al., 2006; Ratcliff et al., 2009; Schendan and Maher, 2009; Rousselet et al., 2011).

P250/N250 (D220)

The impoverished-real-object effect starts on a frontopolar P250 component of the N3 complex, and this effect inverts polarity occipitotemporally, where it is larger with a common average reference and modulates an N250 over the right hemisphere. At this time, only real objects are more negative frontally and more positive occipitotemporally for MI more than LI stimuli. The P250/N250 indexes processes of model selection from view-specific knowledge acquired based on prior experience categorizing objects at the subordinate level (Schendan and Kutas, 2003, 2007a; Scott et al., 2006). This knowledge also supports entry level categorization (Schendan and Maher, 2009) wherein the decision involves access to semantic memory about

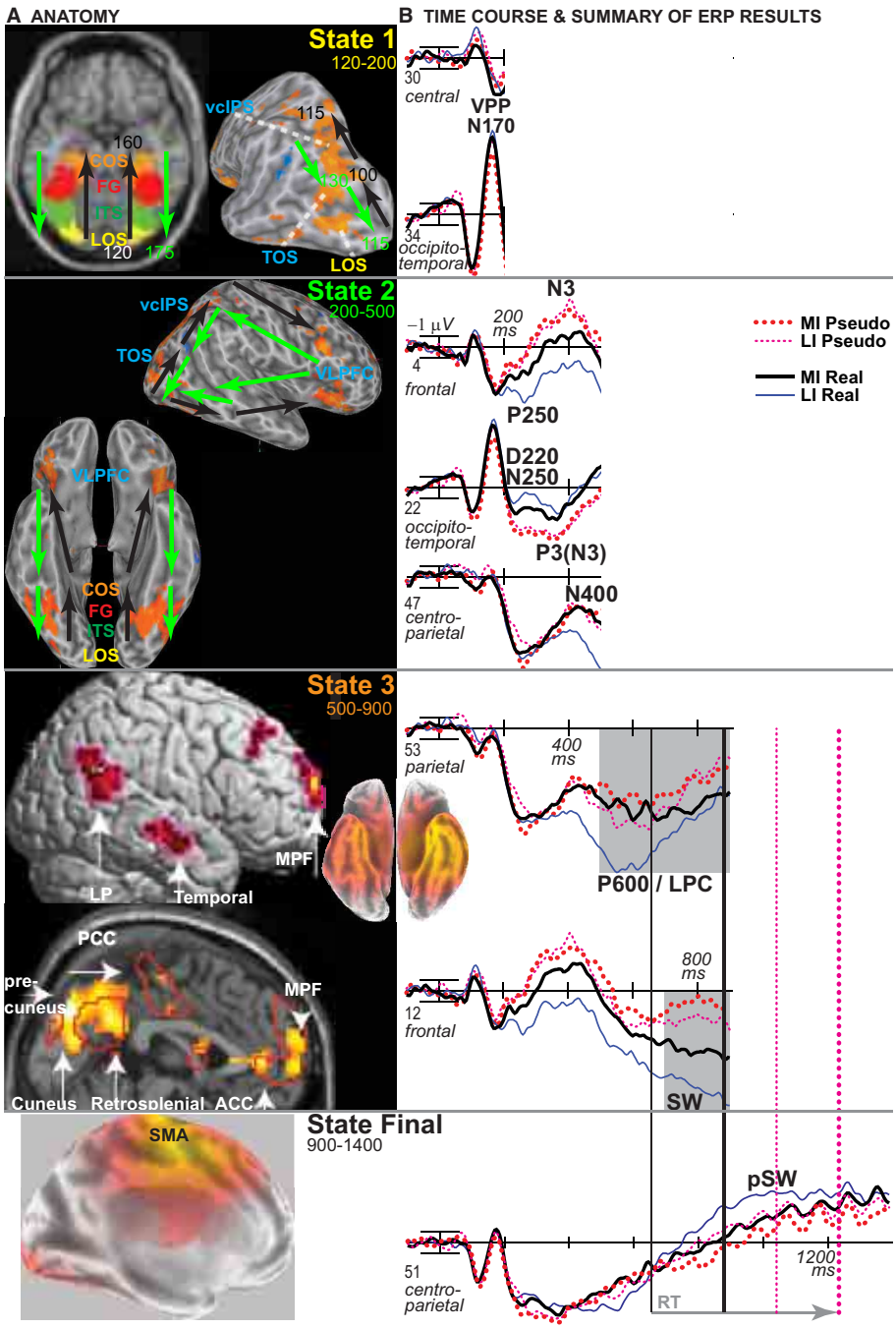


FIGURE 11 | Multiple-state interactive (MUSI) theory and summary of results. Numerals indicate time in milliseconds (ms). **(A)** Brain regions contributing to each state of cortical dynamics. Brain images in black box taken from impoverishment effects in fMRI studies of Ganis et al. (2007) for state 2 and dorsal stream of state 1 and Schendan and Stern (2008) for all others. Note: Green arrow is feedback. Black arrow is bottom-up. State 1 axial slice shows object-sensitive areas. VLPFC, ventral lateral prefrontal cortex; vcIPS, ventrocaudal intraparietal sulcus; TOS, transverse occipital sulcus; LOS, lateral occipital sulcus; ITS, inferotemporal sulcus; FG, fusiform gyrus; COS, collateral sulcus; LP, lateral inferior parietal cortex; MPF, medial prefrontal cortex; PCC, posterior cingulate cortex; ACC, anterior cingulate cortex; SMA, supplementary motor area. **(B)** Timing of cortical dynamics and summary of ERP and RT results. In state 1, the VPP/N170 in occipitotemporal cortex (see **A**) shows no impoverished-real-object effect. In state 2, the N3 complex (including N300, P3, P250, D220, N250 components), indexing an interactive network of occipitotemporal, occipitoparietal, and VLPFC regions (see **A**), shows the earliest impoverished-real-object effect. The later N400 also shows such an effect. In state 3, the P600/LPC in temporal lobe parts of a default mode network (see **A**) shows a later impoverished-real-object effect; note, the sLORETA brain source results from **Figure 8E** are copied here to show the location of P600/LPC effects. The latest such effect modulates an anterior slow wave (SW) in the PCC (see **A**). Gray shading indicates time course of the brain source of the P600 and SW impoverishment effects. A final posterior SW (pSW; state final) correlates with RTs and reflects SMA. Gray arrow points to mean RTs along ERP time course (same legend as for ERPs).

meaning (Jolicoeur et al., 1984). The underlying processes have roles in category learning, short-term repetition priming, and working memory, and these ERPs have been found to localize to areas (e.g., lateral occipital cortex) active also during the VPP/N170, consistent with the present source estimation and the MUSI account (Schweinberger et al., 2002; Foxe et al., 2005; Scott et al., 2006; Sehatpour et al., 2006; Ganis and Schendan, 2008; Schendan and Maher, 2009).

The P250/N250 is probably the same as a D220 observed in decision research, as these ERPs have similar time courses and scalp distributions. The D220 modulates with visual impoverishment defined by relative phase coherence of the image and corresponding category decision accuracy, which has been taken as a definition of task difficulty in the diffusion model of decision making (Philiastides et al., 2006). However, the present P250/N250 finding suggests that the D220 also shows the impoverished-real-object effect, arguing against a generic task difficulty interpretation. Instead, the P250/N250 (D220) reflects the interaction between decision processes, visual perception, and memory (i.e., category knowledge). If the D220 was related only to task difficulty, then it should also show an impoverishment effect for pseudo-objects, which it does not. In addition, the D220 is specific for decisions about the object's category as opposed to its color or episodic familiarity (Philiastides et al., 2006; Schendan and Lucia, submitted) so access to category knowledge is an integral part of the underlying neural processes. Relative to category decisions, color decisions were considered easier (Philiastides et al., 2006), but episodic recognition takes longer and so can be considered harder. Nonetheless, the N3 complex, including the D220, shows an impoverishment effect for category more than episodic memory decisions (Schendan and Lucia, submitted). Further, color decisions do not automatically activate category knowledge (Boucart and Humphreys, 1994; Pins et al., 2004). Thus knowledge activation, not task difficulty, explains why the D220 disappears when the task is color decision.

N3 significance

Altogether, the findings on the N3 complex indicate that PHT and decision processes start around 250 ms (i.e., the onset of when impoverishment affects knowledge activation) and lasts until around 500 ms post-stimulus onset. Impoverishment affects processing earlier for real than pseudo objects. For real objects, impoverishment makes the frontal N3 complex more negative from 250 to 450 ms, whereas, pseudo objects show no such effect. These N3 findings are consistent with previous work indicating that the frontopolar N3 varies with the success of categorization and degree of mental rotation (Schendan and Lucia, 2009; Schendan and Maher, 2009). They are also consistent with the idea that the underlying process primarily detects the relative match to stored information. Evidence indicates that the N3 complex indexes model selection from object information in occipitotemporal cortex based on the relative similarity of the shapes and parts in a specific view, regardless of the constituent small line segments, and working memory and long-term perceptual priming modulate these processes (Holcomb and McPherson, 1994; McPherson and Holcomb, 1999; Doniger et al., 2000, 2001; Daffner et al., 2000b; Schendan and Kutas, 2002, 2003, 2007a; Henson et al., 2004; Gruber and Müller, 2006; Sehatpour et al., 2006; Soldan et al., 2006; Ganis and Schendan, 2008). The neurophysiological processes underlying the N3, perhaps especially frontopolar components, likely contribute critically to processes of *similarity* evaluation for visual object cognition. Testing processes in PHT theories require evaluating the similarity of the spatial configuration (i.e., location) of features between object representations (e.g., between a predicted model and a perceived object). After all, shape similarity drives neural responses in monkey inferotemporal and human occipitotemporal cortex, and is important for category learning (Li et al., 1993; Rainer and Miller, 2000; Freedman et al., 2001, 2002, 2003; Sigala and Logothetis, 2002; Sigala et al., 2002; Sigala, 2004; Jiang et al., 2007; Kriegeskorte et al., 2008; Op de Beeck et al., 2008). Further, most categorization theories posit a central role for evaluation of similarity, especially perceptual similarity acquired through perceptual learning (Goldstone, 1994; Kruschke, 2008), and perceptual learning depends upon processing to a point at which perceptual constancy is achieved (Garrigan and Kellman, 2008).

N3 brain sources

Cortical source findings indicate that LPFC and occipitotemporal cortex activate together during the N3 and the posterior contribution includes knowledge-related processing, consistent with top-down parietal-prefrontal PHT, decision, and MUSI theories. While N3 impoverishment effects localize to the LPFC, regardless of knowledge, they also localize to occipitotemporal cortex only for real objects from 255 to ∼450 ms. By a PHT account, impoverishment of real objects recruits LPFC, which can succeed in modulating object knowledge stored in occipitotemporal cortex, resulting in an impoverishment effect there as well. Impoverishment of pseudo objects also recruits LPFC, but this has little or no modulatory influence on occipitotemporal activity because, by design, these unknown images activate knowledge minimally if at all. Intracranial ERPs extracted from LPFC and occipitotemporal sources show that these impoverished-real-object effects start only after the bottom-up pass (after ∼200 ms). While source estimates are inherently uncertain due to the inverse problem, our localizations fit the areas showing impoverished-real-object effects in fMRI (Ganis et al., 2007; Schendan and Stern, 2008) and are far distant from each other and so spatially resolvable (Pascual-Marqui, 2002; Wagner et al., 2004).

N400, P600, SW

Knowledge modulates the N3 and SW with both LI and MI stimuli but the N400 and P600 only with LI stimuli. Because subjects must activate knowledge in order to make a category decision with both LI and MI stimuli, this finding pinpoints the N3 and SW as candidates for reflecting the critical knowledge activity. However, the anterior SW does not differ between MI unusual and LI canonical views (Schendan and Kutas, 2003) and so is not a general impoverishment marker, and the SW does not show repetition effects with categorized real objects, as it should if it reflects memory (Schendan and Maher, 2009). Thus, the N3 is only viable candidate for a neurophysiological marker of PHT

and decision processes that mediate the impoverished-real-object effect.

Only ERPs from 400 to 700 ms show a knowledge (type) effect only for LI stimuli. Thus, during the N400 and P600, underlying semantic memory and decision evaluation processes, respectively, take place for LI but not MI stimuli. In contrast, the earlier N3 and later SW show knowledge effects at both impoverishment levels, though more for LI than MI, dissociating late ERPs from each other. This dissociation between the N3 and N400/P600 supports a dichotomy (Kousta et al., 2011) between experiential (sensorimotor, affect) knowledge, as indexed by the N3 for vision, and linguistic (verbal) knowledge, indexed by the N400, and later strategic evaluation of earlier category decision processes and secondary higher-order semantic memory analysis, indexed by the P600/LPC (Schendan and Kutas, 2002; Sitnikova et al., 2010).

While all later ERPs after 200 ms show the impoverished-real-object effect, the exact pattern of the interaction differs, dissociating the meaning of these effects. The N3 and N400 findings indicate that LI images of real objects activate knowledge, including meaning, more strongly than MI images of them. After all, the N3 and N400 show impoverishment effects for real objects only. In contrast, the P600 and SW show impoverishment effects also for pseudo objects. Indeed, impoverishment affects processing of pseudo objects for the first time only later, after 500 ms on the P600 and SW. As impoverishment effects apply also to pseudo objects, which cannot activate knowledge, this suggests that these latest effects to some extent reflect response related processes after the category decision. Consistent with this, the P600 seems to index evaluating how well or confidently a task goal or memory matching process has succeeded (Ruchkin and Sutton, 1978; Schendan and Maher, 2009). The P600 is larger on LI than MI trials because LI stimuli are more confidently categorized than MI stimuli, enhancing the P600 and related to faster RTs for LI than MI stimuli. Accordingly, source findings indicate that impoverished-real-object effects during the P600 reflect post-categorization processes in anterior inferior and mediotemporal cortex related to evaluating the decision and memory match, and, after 700 ms during the SW, response planning related processes in a PCC region. These regions show impoverished-real-object effects in fMRI, though PCC shows deactivation (i.e., more active for LI than MI) (Ganis et al., 2007; Schendan and Stern, 2008). Altogether, the ERP time course indicates that larger fMRI impoverishment activations for real (than pseudo) objects reflect both earlier processes during the N3 and N400 and later processes during the P600 and SW, whereas the smaller fMRI impoverishment activations for pseudo objects reflect only later processes after 500 ms.

N400 linguistic knowledge

From 400 to 500 ms, impoverishment modulates the centroparietal N400, which is smallest for LI real objects relative to all other conditions. The idea that name, semantic (i.e., indexed by N400), and object model (or "structural description," i.e., indexed by the N3) knowledge interact bidirectionally to achieve visual object categorization and naming is consistent with an interactive activation and competition model of object naming (Humphreys et al., 1999) and the MUSI account. By such accounts, the present finding of an impoverished-real-object effect on the N400 would indicate that interactive computations among knowledge systems, including linguistic semantic memory, also have a role in achieving visual constancy of the cognitive decision. However, source findings suggest only posterior contributions from occipitotemporal and anterotemporal cortex. As no evidence was found for prefrontal-posterior interactions during the N400, word-related semantic memory may not contribute to perceptual hypothesis testing but rather activates after the category decision.

P600

By the MUSI account, the P600 in state 3 reflects strategic evaluation. P600 (or LPC) knowledge effects may also in part reflect stimulus categorization (Dien et al., 2004). The ∼50% overall categorization rate confirms that the ERP effects do not reflect differences in subjective probability of categorization success associated with P3(00)-like ERPs (Polich and Bondurant, 1997). The P600 shows impoverishment effects for the first time for both real and pseudo objects. The P600 is more positive for LI than MI stimuli for real more than pseudo objects. The P600 effect for real objects replicates the finding that the P600 is larger to LI canonical than MI unusual views on categorization and recognition (Schendan and Kutas, 2003; Schendan and Lucia, submitted).

SW

After 700 ms, a broadly distributed SW with a midline central maximum differs among all conditions. The SW impoverished-real-object effect manifests as greater positivity for LI real objects relative to MI ones relative to LI pseudo objects relative to MI ones. The SW seems to index processes related to response execution and monitoring, being less positive when these processes are more challenging (Schendan and Maher, 2009). After 700 ms during the SW, impoverishment effects localize primarily to the PCC region that instead activates more for LI than MI real objects in fMRI (Ganis et al., 2007; Schendan and Stern, 2008). The PCC is part of a default mode network for internal evaluation, exogenous attention, episodic memory retrieval, and semantic memory computations with words that is anticorrelated in fMRI with the active task network that instead includes prefrontal and posterior processing areas that underlie the N3 and N400 (Fox et al., 2005; Buckner et al., 2008; Binder et al., 2009). The present time course would be consistent with the idea that the active task network operates from 200 to 500 ms during the N3 and N400, whereas the P600 and SW reflect activity in the mediotemporal and PCC parts of the default mode network, respectively. Intriguingly, after 700 ms, real objects activate anterior and medial temporal cortex and PCC, whereas pseudo objects activate only the PCC. This suggests that knowledge in temporal cortex contributes to PCC activity as part of default mode interactions with real objects but not pseudo objects, which cannot activate knowledge. Because the anterior and medial temporal cortex activity starts during the P600, the same activity during the later SW likely reflects a

continuation of the earlier posterior positivity and may best be considered an LPC contribution to posterior ERPs after 500 ms (P600, SW).

Alternative Explanations

Not Subjective Probability

Subjective probability of categorized vs. uncategorized responses cannot explain the results. Subjects were naïve that some objects were not real (pseudo-objects) and so uncategorizable, and categorized and uncategorized responses split about evenly: From the subjects' perspectives, any object, whether truly real or pseudo, that did not belong to a known category was merely an uncategorized object, and this happened about half the time, making the task essentially a reliable and simple two-choice decision between half categorized and half uncategorized images.

Not Early Motor Potentials

N3 effects do not reflect earlier time courses of motor potentials for LI than MI objects. (a) The N3 and RTs dissociate. The N3 complex shows impoverished-real-object effects well before the earliest RT to LI real objects (∼650 ms). Still, if the N3 is merely a motor potential, a larger N3 should always be associated with longer RTs. To the contrary, it has been found that, when people categorize fragmented pictures of objects that have been repeated (primed), the N3 is the same between all repetition conditions, whereas RTs and other ERPs, such as the P600, differ between the various repeated conditions (Schendan and Kutas, 2007a,b). Further, the N3 is larger when categorization RTs are faster (instead of slower) for scrambled than intact objects (Schendan and Lucia, 2010). (b) The N3 does not index a motor readiness potential. The readiness potential (RP) is a midline central negativity that is greater for contralateral than ipsilateral responses by ∼200 ms post-stimulus due to differential activity in primary motor cortex. The RP could make negativity greater for MI than LI stimuli but cannot explain these N3 effects. First, with a mastoid reference, as herein, the RP is maximal over central midline (C3, C4) and absent at frontal sites (F3, F4) (Kutas and Donchin, 1980) near the frontocentral N3 and far from the frontopolar ERPs. Second, N3 and RP waveforms differ. The N3 impoverishment effect and its LPFC sources return to baseline by 500 ms, which is ∼150 ms before the earliest RT. In contrast, the RP rises steadily over ∼500 ms preceding the RT (Coles, 1989). Third, no impoverishment effects were found in primary motor cortex in our N3 source estimates and neuroimaging studies of model verification (Kosslyn et al., 1994; Ganis et al., 2007; Schendan and Stern, 2008). (c) N3 impoverishment effects cannot merely be related to motor planning. An impoverishment effect in the supplementary motor area was found in the fMRI version for fragmented pictures (Ganis et al., 2007) but not for unusual vs. canonical views (Kosslyn et al., 1994; Schendan and Stern, 2008). Only ventral premotor cortex activity reflects a general process related to image impoverishment with objects (unusual views, fragmented pictures) (Ganis et al., 2007; Schendan and Stern, 2008) that has been implicated in evidence accumulation for a decision (Heekeren et al., 2008). Finally, note that N3 knowledge effects are unlikely to reflect differences in motor responses between real and pseudo objects because similarly large N3 differences have been found with the full set of these stimuli during passive viewing when both object types were non-targets (Schendan et al., 1998), but P600 (or LPC) knowledge effects may in part reflect stimulus categorization (Dien et al., 2004).

Late Motor Activity (pSW)

The most likely ERPs to include motor potentials are those around the time of the response. Indeed, after 900 ms, a posterior slow wave (pSW) (**Figure 9**) modulates independently with type and impoverishment but shows no evidence of impoverished-real-object effects. The pSW is more positive for real than pseudo objects and for MI than LI trials, matches and correlates with corresponding RT effects, and localizes to SMA and nearby anterior cingulate regions that show impoverishment effects in fMRI (Ganis et al., 2007). These sources are consistent with late slow intracranial ERPs in premotor and motor regions of epilepsy patients (Halgren et al., 1994). However, the SMA region and pSW findings do not reflect model verification *per se* but rather later processes related to generating a response under MI relative to LI conditions because it was not specific for real objects, and SMA shows no effects of impoverishment by viewpoint (Schendan and Stern, 2008).

Nonvisual Impoverishment Factors

The median split approach captures all possible factors that contribute to the visual constancy of a category decision, but using fragmentation level to define LI and MI conditions yields the same pattern (**Figure 10**), demonstrating that visuoperceptual factors were among those driving the effects. Further, impoverishment effects here resemble those found when fragmentation or viewpoint impoverishes the images (Doniger et al., 2000; Schendan and Kutas, 2003). Future work will need to tease apart each perceptual and cognitive factor using the times and regions of interest defined here.

Conclusion

Findings reveal the cortical dynamics to achieve visual constancy of a category decision. The time course of knowledge, impoverishment, and impoverished-real-object findings fit best a hybrid MUSI account that incorporates parietal-prefrontal PHT theories and decision theories to explain the visual constancy of object cognition. By such an account, for MI objects, the initial bottom-up pass may fail to yield a sufficiently accurate decision, thereby recruiting prefrontal cortex to send top-down modulatory inputs to occipitotemporal object processing areas to accumulate more perceptual and knowledge evidence for the decision. Critically, by examining both impoverishment and knowledge factors, the findings demonstrate that impoverishment adversely affects activation of knowledge (conveyed by real objects) more than merely perceptual processing of any object (including pseudo objects) by ∼250 ms after seeing an image. Convergent evidence, including from studies of the top-down processes for mental imagery, lead to the conclusion that, during the N3 complex, top-down processes posited in parietal-prefrontal PHT and decision

theories recruit LPFC to modulate not only perceptual evidence coming from posterior object processing areas but also the activation of knowledge in those areas. This happens after the initial bottom-up activation of object processing areas.

Altogether these findings suggest the following hybrid MUSI account, which incorporates parietal-prefrontal PHT and decision theories, to explain the cortical dynamics for the visual constancy of object cognition (**Figure 11**). State 1 during the VPP/N170 between 120 and 200 ms involves initial, bottom-up activation of ventral object processing cortex. Starting ∼230 ms (state 2), model selection based on both visual input and memory (e.g., knowledge) for a decision starts during a second state of interactive bottom-up, recurrent, and feedback (reflexive top-down) activity among object processing areas in occipitotemporal cortex and VLPFC, indexed by the N3 complex. When visual input is highly impoverished, top-down processes of PHT in parietal and LPFC areas, especially VLPFC regions, can modulate occipitotemporal activity to facilitate the visual object constancy of a decision, achieving accuracy at a cost of longer response times. Any impoverished image can recruit PHT processes, but these processes modulate knowledge-related computations in occipitotemporal cortex only when the image depicts a real object. Based on convergent evidence, we propose that top-down processes for PHT are recruited based on the shape similarity among perceived object(s) and stored models (i.e., match between percept and knowledge), which decreases as image impoverishment increases. Also in state 2, during a centroparietal N400 from 400 to 500 ms, interactive activation of linguistic (verbal) knowledge (e.g., the name) happens in temporal cortex. Later after ∼500 ms (state 3), anterotemporal cortex during the P600/LPC and posterior cingulate activity during a broad slow wave (SW), perhaps in the default mode network for internal evaluation of prior processing and memory activation, and secondary higher-order semantic memory. Finally, after 900 ms (in a final response state), SMA and anterior cingulate activity, indexed by a posterior slow wave correlated with RTs, plans the execution of the motor response.

Funding

Research supported by Research Executive Agency European Union, Seventh Framework Programme (FP7), Marie Curie Career Integration Grant: PCIG09-GA-2011-294144-COGNITSIMS, the Research Executive Agency European Union FP7 Marie Curie Initial Training Networks (ITN) FP7-PEOPLE-2013-ITN-604764 Innovative Doctoral Programme (IDP): COGNOVO, and Plymouth University Grants to HS from the Social Science Collaboration with the University of Exeter Scheme, the International Research, Networking and Collaboration Grant, and the Faculty Innovation Centre (FInC) Grant.

Acknowledgments

This project was completed in the School of Psychology at the University of Plymouth. Additional research support from Tufts University. HS and GG designed and set-up the experiment. HS assisted with testing and directed the plan for data acquisition and analyses, supervised data collection, analyzed some data, prepared tables and figures, and wrote the manuscript. GG contributed to manuscript preparation. Emily A. Slocombe (ES) collected the data. The authors are grateful to ES also for assisting with some data analyses, Emily Newman for assisting ES with analyses, and Lisa C. Lucia and Stephen M. Maher for assistance with data collection.

References

Allison, T., Puce, A., Spencer, D. D., and McCarthy, G. (1999). Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb. Cortex* 9, 415–430. doi: 10.1093/cercor/9.5.415

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454. doi: 10.1073/pnas.0507062103

Barrett, S. E., and Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain Cogn.* 14, 201–212. doi: 10.1016/0278-2626(90)90029-N

Bastin, J., Committeri, G., Kahane, P., Galati, G., Minotti, L., Lachaux, J. P., et al. (2013). Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages. *Hum. Brain Mapp.* 34, 1357–1370. doi: 10.1002/hbm.21515

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Boucart, M., and Humphreys, G. W. (1994). Attention to orientation, size, luminance, and color: attentional failure within the form domain. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 61–80. doi: 10.1037/0096-1523.20.1.61

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107. doi: 10.1016/S0165-0173(01)00085-6

Bülthoff, H. H., Edelman, S. Y., and Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cereb. Cortex* 5, 247–260. doi: 10.1093/cercor/5.3.247

Busch, N. A., Herrmann, C. S., Müller, M. M., Lenz, D., and Gruber, T. (2006). A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. *Neuroimage* 33, 1169–1177. doi: 10.1016/j.neuroimage.2006.07.034

Clarke, A., Taylor, K. I., and Tyler, L. K. (2011). The evolution of meaning: spatio-temporal dynamics of visual object recognition. *J. Cogn. Neurosci.* 23, 1887–1899. doi: 10.1162/jocn.2010.21544

Coles, M. G. (1989). Modern mind-brain reading: psychophysiology, physiology, and cognition. *Psychophysiology* 26, 251–269. doi: 10.1111/j.1469-8986.1989.tb01916.x

Curran, T., Tanaka, J. W., and Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18. doi: 10.3758/CABN.2.1.1

Daffner, K. R., Mesulam, M. M., Scinto, L. F., Acar, D., Calvo, V., Faust, R., et al. (2000a). The central role of the prefrontal cortex in directing

attention to novel events. *Brain* 123(Pt 5), 927–939. doi: 10.1093/brain/123.5.927

Daffner, K. R., Mesulam, M. M., Scinto, L. F., Calvo, V., Faust, R., and Holcomb, P. J. (2000b). An electrophysiological index of stimulus unfamiliarity. *Psychophysiology* 37, 737–747. doi: 10.1111/1469-8986.3760737

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. doi: 10.1038/380499a0

David, O., Harrison, L., and Friston, K. J. (2005). Modeling event-related responses in the brain. *Neuroimage* 25, 756–770. doi: 10.1016/j.neuroimage.2004.12.030

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., and Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *Neuroimage* 30, 1255–1272. doi: 10.1016/j.neuroimage.2005.10.045

Deco, G., and Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.* 44, 621–642. doi: 10.1016/j.visres.2003.09.037

Delorme, A., Richard, G., and Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res.* 40, 2187–2200. doi: 10.1016/S0042-6989(00)00083-3

Delorme, A., Rousselet, G. A., Macé, M. J., and Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cogn. Brain Res.* 19, 103–113. doi: 10.1016/j.cogbrainres.2003.11.010

Dien, J., Spencer, K. M., and Donchin, E. (2004). Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology* 41, 665–678. doi: 10.1111/j.1469-8986.2004.00193.x

Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481

Di Russo, F., Martínez, A., and Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cereb. Cortex* 13, 486–499. doi: 10.1093/cercor/13.5.486

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., et al. (2000). Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J. Cogn. Neurosci.* 12, 615–621. doi: 10.1162/089892900562372

Doniger, G. M., Foxe, J. J., Schroeder, C. E., Murray, M. M., Higgins, B. A., and Javitt, D. C. (2001). Visual perceptual learning in human object recognition areas: a repetition priming study using high-density electrical mapping. *Neuroimage* 13, 305–313. doi: 10.1006/nimg.2000.0684

Edelman, S. (1999). *Representation and Recognition in Vision.* Cambridge, MA: MIT Press.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Res.* 44, 1321–1331. doi: 10.1016/j.visres.2003.10.024

Enns, J. T., and Lleras, A. (2008). What's next? New evidence for prediction in human vision. *Trends Cogn. Sci.* 12, 327–333. doi: 10.1016/j.tics.2008.06.001

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Front. Psychol.* 2:243. doi: 10.3389/fpsyg.2011.00243

Fabre-Thorpe, M., Delorme, A., Marlot, C., and Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J. Cogn. Neurosci.* 13, 171–180. doi: 10.1162/089892901564234

Fazl, A., Grossberg, S., and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cogn. Psychol.* 58, 1–48. doi: 10.1016/j.cogpsych.2008.05.001

Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., and Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: associated ERPs and effects of lateralization. *Brain Cogn.* 59, 145–158. doi: 10.1016/j.bandc.2005.06.002

Folstein, J. R., and van Petten, C. (2004). Multidimensional rule, unidimensional rule, and similarity strategies in categorization: event-related brain potential correlates. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1026–1044. doi: 10.1037/0278-7393.30.5.1026

Folstein, J. R., and van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170. doi: 10.1111/j.1469-8986.2007.00602.x

Folstein, J. R., Van Petten, C., and Rose, S. A. (2008). Novelty and conflict in the categorization of complex stimuli. *Psychophysiology* 45, 467–479. doi: 10.1111/j.1469-8986.2007.00628.x

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Foxe, J. J., Murray, M. M., and Javitt, D. C. (2005). Filling-in in schizophrenia: a high-density electrical mapping and source-analysis investigation of illusory contour processing. *Cereb. Cortex* 15, 1914–1927. doi: 10.1093/cercor/bhi069

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316. doi: 10.1126/science.291.5502.312

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *J. Neurophysiol.* 88, 929–941.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* 23, 5235–5246.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Ganis, G., and Kosslyn, S. M. (2007). "Multiple mechanisms of top-down processing in vision," in *Representation and Brain,* ed S. Funahashi (Tokyo: Springer Verlag), 21–45.

Ganis, G., and Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Res. Cogn. Brain Res.* 16, 123–144. doi: 10.1016/S0926-6410(02)00244-6

Ganis, G., and Schendan, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *Neuroimage* 42, 1714–1727. doi: 10.1016/j.neuroimage.2008.07.004

Ganis, G., Schendan, H. E., and Kosslyn, S. M. (2007). Neuroimaging evidence for object model verification theory: role of prefrontal control in visual object categorization. *Neuroimage* 34, 384–398. doi: 10.1016/j.neuroimage.2006.09.008

Garrigan, P., and Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2248–2253. doi: 10.1073/pnas.0711878105

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron* 34, 161–171. doi: 10.1016/S0896-6273(02)00622-0

Gold, J. I., and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition* 52, 125–157. doi: 10.1016/0010-0277(94)90065-5

Gratton, C., Evans, K. M., and Federmeier, K. D. (2009). See what I mean? An ERP study of the effect of background knowledge on novel object processing. *Mem. Cogn.* 37, 277–291. doi: 10.3758/MC.37.3.277

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203. doi: 10.1016/S0896-6273(00)80832-6

Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677. doi: 10.1146/annurev.neuro.27.070203.144220

Gruber, T., and Müller, M. M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb. Cortex* 15, 109–116. doi: 10.1093/cercor/bhh113

Gruber, T., and Müller, M. M. (2006). Oscillatory brain activity in the human EEG during indirect and direct memory tasks. *Brain Res.* 1097, 194–204. doi: 10.1016/j.brainres.2006.04.069

Gruber, T., Trujillo-Barreto, N. J., Giabbiconi, C. M., Valdés-Sosa, P. A., and Muller, M. M. (2006). Brain electrical tomography (BET) analysis of induced gamma band responses during a simple object recognition task. *Neuroimage* 29, 888–900. doi: 10.1016/j.neuroimage.2005.09.004

Guillem, F., Rougier, A., and Claverie, B. (1999). Short- and long-delay intracranial ERP repetition effects dissociate memory systems in the human brain. *J. Cogn. Neurosci.* 11, 437–458. doi: 10.1162/089892999563526

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., et al. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. *Electroencephalogr. Clin. Neurophysiol.* 94, 229–250. doi: 10.1016/0013-4694(95)98475-N

Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K., Chauvel, P., et al. (1994). Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and Rolandic cortices. *J. Physiol. Paris* 88, 51–80. doi: 10.1016/0928-4257(94)90093-0

Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479. doi: 10.1038/nrn2374

Henson, R. N., Rylands, A., Ross, E., Vuilleumeir, P., and Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *Neuroimage* 21, 1674–1689. doi: 10.1016/j.neuroimage.2003.12.020

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804. doi: 10.1016/S0896-6273(02)01091-7

Holcomb, P. J. (2003). *"ERPSystem." 2.0a ed.* Medford: Tufts University.

Holcomb, P. J., and McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain Cogn.* 24, 259–276. doi: 10.1006/brcg.1994.1014

Humphreys, G. W., Lamote, C., and Lloyd-Jones, T. J. (1995). An interactive activation approach to object processing: effects of structural similarity, name frequency, and task in normality and pathology. *Memory* 3, 535–586. doi: 10.1080/09658219508253164

Humphreys, G. W., Price, C. J., and Riddoch, M. J. (1999). From objects to names: a cognitive neuroscience approach. *Psychol. Res.* 62, 118–130. doi: 10.1007/s004260050046

Humphreys, G. W., Riddoch, M. J., and Price, C. J. (1997). Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 352, 1275–1282. doi: 10.1098/rstb.1997.0110

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., and Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 53, 891–903. doi: 10.1016/j.neuron.2007.02.015

Johnson, J. S., and Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *J. Vis.* 3, 499–512. doi: 10.1167/3.7.4

Johnson, J. S., and Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J. Vis.* 5, 299–312. doi: 10.1167/5.4.2

Johnson, R. Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology* 23, 367–384. doi: 10.1111/j.1469-8986.1986.tb00649.x

Jolicoeur, P., Gluck, M. A., and Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cogn. Psychol.* 16, 243–275. doi: 10.1016/0010-0285(84)90009-4

Kirchner, H., and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res.* 46, 1762–1776. doi: 10.1016/j.visres.2005.10.002

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L., and Anderson, A. K. (1994). Identifying objects seen from different viewpoints. A PET investigation. *Brain* 117, 1055–1071. doi: 10.1093/brain/117.5.1055

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (2011). The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.* 140, 14–34. doi: 10.1037/a0021446

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141. doi: 10.1016/j.neuron.2008.10.043

Kroll, J. F., and Potter, M. C. (1984). Recognizing words, pictures, and concepts: a comparison of lexical, object, and reality decisions. *J. Verbal Learn. Verbal Behav.* 23, 39–66. doi: 10.1016/S0022-5371(84)90499-7

Kruschke, J. K. (2008). "Models of categorization," in *The Cambridge Handbook of Computational Psychology,* ed R. Sun (New York, NY: Cambridge University Press), 267–301. doi: 10.1017/CBO9780511816772.013

Kutas, M., and Donchin, E. (1980). Preparation to respond as manifested by movement-related brain potentials. *Brain Res.* 202, 95–115. doi: 10.1016/S0006-8993(80)80037-0

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. doi: 10.1146/annurev.psych.093008.131123

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18. doi: 10.1016/S1364-6613(02)00013-X

Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lee, T. S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448. doi: 10.1364/JOSAA.20.001434

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., and Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cereb. Cortex* 11, 287–297. doi: 10.1093/cercor/11.4.287

Li, L., Miller, E. K., and Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* 69, 1918–1929.

Lowe, D. G. (2000). "Towards a computational model for object recognition in IT cortex," in *First IEEE International Workshop on Biologically Motivated Computer Vision* (Seoul), 20–31.

Luck, S. J. (1999). Direct and indirect integration of event-related potentials, functional magnetic resonance images, and single-unit recordings. *Hum. Brain Mapp.* 8, 115–201.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Luck, S. J. (2006). "The operation of attention– millisecond by millisecond– over the first half second," in *The First Half Second: The Microgenesis and Temporal Dynamics of Unconscious and Conscious Visual Processes*, eds H. Ogmen and B. G. Breitmeyer (Cambridge: MIT), 187–206.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., and Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38, 487–497. doi: 10.1016/S0896-6273(03)00197-1

McPherson, W. B., and Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* 36, 53–65. doi: 10.1017/S0048577299971196

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to cousineau (2005). *Tutorial Quant. Methods Psychol.* 4, 61–64.

Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Op de Beeck, H. P., Wagemans, J., and Vogels, R. (2008). The representation of perceived shape similarity and its role for category learning in monkeys: a modeling study. *Vision Res.* 48, 598–610. doi: 10.1016/j.visres.2007.11.019

Palmer, S. E., Rosch, E., and Chase, P. (1981). "Canonical perspective and the perception of objects," in *Attention and Performance,* eds J. Long and A. Baddeley (Hillsdale, NJ: Erlbaum), 135–151.

Pascual-Marqui, R. D. (2002). Standardized low resolution electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24, 5–12.

Philiastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J. Neurosci.* 26, 8965–8975. doi: 10.1523/JNEUROSCI.1655-06.2006

Philiastides, M. G., and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb. Cortex* 16, 509–518. doi: 10.1093/cercor/bhi130

Philiastides, M. G., and Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *J. Neurosci.* 27, 13082–13091. doi: 10.1523/JNEUROSCI.3540-07.2007

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R. Jr., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi: 10.1111/1469-8986.3720127

Pins, D., Meyer, M. E., Foucher, J., Humphreys, G., and Boucart, M. (2004). Neural correlates of implicit object identification. *Neuropsychologia* 42, 1247–1259. doi: 10.1016/j.neuropsychologia.2004.01.005

Polich, J., and Bondurant, T. (1997). P300 sequence effects, probability, and interstimulus interval. *Physiol. Behav.* 61, 843–849. doi: 10.1016/S0031-9384(96)00564-1

Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S., and Friston, K. J. (1996). The neural regions sustaining object recognition and naming. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 263, 1501–1507. doi: 10.1098/rspb.1996.0219

Puce, A., Allison, T., and McCarthy, G. (1999). Electrophysiological studies of human face perception. III: effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458. doi: 10.1093/cercor/9.5.445

Rainer, G., and Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron* 27, 179–189. doi: 10.1016/S0896-6273(00)00019-2

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R., Philiastides, M. G., and Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6539–6544. doi: 10.1073/pnas.0812589106

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Rousselet, G. A., Gaspar, C. M., Wieczorek, K. P., and Pernet, C. R. (2011). Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Front. Psychol.* 2:137. doi: 10.3389/fpsyg.2011.00137

Ruchkin, D. S., and Sutton, S. (1978). Emitted P300 potentials and temporal uncertainty. *Electroencephalogr. Clin. Neurophysiol.* 45, 268–277. doi: 10.1016/0013-4694(78)90010-X

Schendan, H. E., and Ganis, G. (2012). Electrophysiological potentials reveal cortical mechanisms for mental imagery, mental simulation, and grounded (embodied) cognition. *Front. Psychol.* 3:329. doi: 10.3389/fpsyg.2012.00329

Schendan, H. E., and Ganis, G. (2013). Face-specificity is robust across diverse stimuli and individual people, even when interstimulus variance is zero. *Psychophysiology* 50, 287–291. doi: 10.1111/psyp.12013

Schendan, H. E., Ganis, G., and Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology* 35, 240–251. doi: 10.1111/1469-8986.3530240

Schendan, H. E., and Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia* 40, 931–945. doi: 10.1016/S0028-3932(01)00176-2

Schendan, H. E., and Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *J. Cogn. Neurosci.* 15, 111–135. doi: 10.1162/089892903321107864

Schendan, H. E., and Kutas, M. (2007a). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *J. Cogn. Neurosci.* 19, 734–749. doi: 10.1162/jocn.2007.19.5.734

Schendan, H. E., and Kutas, M. (2007b). Neurophysiological evidence for transfer appropriate processing of memory: processing versus feature similarity. *Psychon. Bull. Rev.* 14, 612–619. doi: 10.3758/BF03196810

Schendan, H. E., and Lucia, L. C. (2009). Visual object cognition precedes but also temporally overlaps mental rotation. *Brain Res.* 1294, 91–105. doi: 10.1016/j.brainres.2009.07.036

Schendan, H. E., and Lucia, L. C. (2010). Object-sensitive activity reflects earlier perceptual and later cognitive processing of visual objects between 95 and 500 ms. *Brain Res.* 1329, 124–141. doi: 10.1016/j.brainres.2010.01.062

Schendan, H. E., and Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage* 44, 1423–1438. doi: 10.1016/j.neuroimage.2008.09.061

Schendan, H. E., and Stern, C. E. (2008). Where vision meets memory: prefrontal-posterior networks for visual object constancy during categorization and recognition. *Cereb. Cortex* 18, 1695–1711. doi: 10.1093/cercor/bhm197

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., and Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Brain Res. Cogn. Brain Res.* 14, 398–409. doi: 10.1016/S0926-6410(02)00142-8

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., and Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *J. Cogn. Neurosci.* 18, 1453–1465. doi: 10.1162/jocn.2006.18.9.1453

Sehatpour, P., Molholm, S., Javitt, D. C., and Foxe, J. J. (2006). Spatiotemporal dynamics of human object recognition processing: an integrated high-density electrical mapping and functional imaging study of "closure" processes. *Neuroimage* 29, 605–618. doi: 10.1016/j.neuroimage.2005.07.049

Sehatpour, P., Molholm, S., Schwartz, T. H., Mahoney, J. R., Mehta, A. D., Javitt, D. C., et al. (2008). A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4399–4404. doi: 10.1073/pnas.0708418105

Serre, T., Oliva, A., and Poggio, T. (2007a). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429. doi: 10.1073/pnas.0700622104

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007b). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/TPAMI.2007.56

Sigala, N. (2004). Visual categorization and the inferior temporal cortex. *Behav. Brain Res.* 149, 1–7. doi: 10.1016/S0166-4328(03)00224-9

Sigala, N., Gabbiani, F., and Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *J. Cogn. Neurosci.* 14, 187–198. doi: 10.1162/089892902317236830

Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320. doi: 10.1038/415318a

Sitnikova, T., Perrone, C., Goff, D., and Kuperberg, G. R. (2010). Neurocognitive mechanisms of conceptual processing in healthy adults and patients with schizophrenia. *Int. J. Psychophysiol.* 75, 86–99. doi: 10.1016/j.ijpsycho.2009.12.002

Snodgrass, J. G., and Corwin, J. (1988a). Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. *Percept. Motor Skills* 67, 3–36. doi: 10.2466/pms.1988.67.1.3

Snodgrass, J. G., and Corwin, J. (1988b). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J. Exp. Psychol. Gen.* 117, 34–50. doi: 10.1037/0096-3445.117.1.34

Snodgrass, J. G., and Feenan, K. (1990). Priming effects in picture fragment completion: support for the perceptual closure hypothesis. *J. Exp. Psychol. Gen.* 119, 276–296. doi: 10.1037/0096-3445.119.3.276

Snodgrass, J. G., Smith, B., Feenan, K., and Corwin, J. (1987). Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications. *Behav. Res. Methods Instrum. Comput.* 19, 270–274. doi: 10.3758/BF03203798

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174

Snodgrass, J. G., and Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behav. Res. Methods Instrum. Comput.* 28, 516–536. doi: 10.3758/BF03200540

Soldan, A., Mangels, J. A., and Cooper, L. A. (2006). Evaluating models of object-decision priming: evidence from event-related potential repetition effects. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 230–248. doi: 10.1037/0278-7393.32.2.230

Spreng, R. N., Sepulcre, J., Turner, G. R., Stevens, W. D., and Schacter, D. L. (2013). Intrinsic architecture underlying the relations among the default, dorsal attention, and frontoparietal control networks of the human brain. *J. Cogn. Neurosci.* 25, 74–86. doi: 10.1162/jocn_a_00281

Stuss, D. T., Picton, T. W., and Cerri, A. M. (1986). Searching for the names of pictures: an event-related potential study. *Psychophysiology* 23, 215–223. doi: 10.1111/j.1469-8986.1986.tb00622.x

Stuss, D. T., Picton, T. W., Cerri, A. M., Leech, E. E., and Stethem, L. L. (1992). Perceptual closure and object identification: electrophysiological responses to incomplete pictures. *Brain Cogn.* 19, 253–266. doi: 10.1016/0278-2626(92)90047-P

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb. Cortex* 13, 90–99. doi: 10.1093/cercor/13.1.90

Tarr, M. J., and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cogn. Psychol.* 21, 233–282. doi: 10.1016/0010-0285(89)90009-1

Tarr, M. J., Williams, P., Hayward, W. G., and Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nat. Neurosci.* 1, 275–277. doi: 10.1038/1089

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522. doi: 10.1038/381520a0

Trautner, P., Dietl, T., Staedtgen, M., Mecklinger, A., Grunwald, T., Elger, C. E., et al. (2004). Recognition of famous faces in the medial temporal lobe: an invasive ERP study. *Neurology* 63, 1203–1208. doi: 10.1212/01.WNL.0000140487.55973.D7

Treisman, A. (2006). How the deployment of attention determines what we see. *Vis. cogn.* 14, 411–443. doi: 10.1080/13506280500195250

Turnbull, O. H., Carey, D. P., and McCarthy, R. A. (1997). The neuropsychology of object constancy. *J. Int. Neuropsychol. Soc.* 3, 288–298.

Ullman, S. (1996). *High-level Vision: Object Recognition and Visual Cognition*. Cambridge: MIT.

Viggiano, M. P., and Kutas, M. (2000). Overt and covert identification of fragmented objects inferred from performance and electrophysiological measures. *J. Exp. Psychol. Gen.* 129, 107–125. doi: 10.1037/0096-3445.129.1.107

Voss, J. L., Schendan, H. E., and Paller, K. A. (2010). Finding meaning in novel geometric shapes influences electrophysiological correlates of repetition and dissociates perceptual and conceptual priming. *Neuroimage* 49, 2879–2889. doi: 10.1016/j.neuroimage.2009.09.012

Wagner, M., Fuchs, M., and Kastner, J. (2004). Evaluation of sLORETA in the presence of noise and multiple sources. *Brain Topogr.* 16, 277–280. doi: 10.1023/B:BRAT.0000032865.58382.62

Wheeler, M. E., Petersen, S. E., Nelson, S. M., Ploran, E. J., and Velanova, K. (2008). Dissociating early and late error signals in perceptual recognition. *J. Cogn. Neurosci.* 20, 2211–2225. doi: 10.1162/jocn.2008.20155

Zhang, X. L., Begleiter, H., Porjesz, B., Wang, W. Y., and Litke, A. (1995). Event related potentials during object recognition tasks. *Brain Res. Bull.* 38, 531–538. doi: 10.1016/0361-9230(95)02023-5

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.

Copyright © 2015 Schendan and Ganis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 28 May 2014
doi: 10.3389/fpsyg.2014.00512

Affective and contextual values modulate spatial frequency use in object recognition

Laurent Caplette[1], Gregory West[1], Marie Gomot[2], Frédéric Gosselin[1]* and Bruno Wicker[3]

[1] Département de Psychologie, CERNEC, Université de Montréal, Montréal, QC, Canada
[2] INSERM U930 Imagerie et Cerveau, Université François-Rabelais de Tours, CHRU de Tours, Tours, France
[3] CNRS UMR 7289, Institut de Neurosciences de la Timone, Aix-Marseille Université, Marseille, France

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Elan Barenholtz, Florida Atlantic University, USA
Matt Craddock, University of Leeds, UK

***Correspondence:**
Frédéric Gosselin, Département de Psychologie, CERNEC, Université de Montréal, Pavillon Marie-Victorin, 90, Avenue Vincent d'Indy, Montréal, QC H2V 2S9, Canada
e-mail: frederic.gosselin@umontreal.ca

Visual object recognition is of fundamental importance in our everyday interaction with the environment. Recent models of visual perception emphasize the role of top-down predictions facilitating object recognition via initial guesses that limit the number of object representations that need to be considered. Several results suggest that this rapid and efficient object processing relies on the early extraction and processing of low spatial frequencies (LSF). The present study aimed to investigate the SF content of visual object representations and its modulation by contextual and affective values of the perceived object during a picture-name verification task. Stimuli consisted of pictures of objects equalized in SF content and categorized as having low or high affective and contextual values. To access the SF content of stored visual representations of objects, SFs of each image were then randomly sampled on a trial-by-trial basis. Results reveal that intermediate SFs between 14 and 24 cycles per object (2.3–4 cycles per degree) are correlated with fast and accurate identification for all categories of objects. Moreover, there was a significant interaction between affective and contextual values over the SFs correlating with fast recognition. These results suggest that affective and contextual values of a visual object modulate the SF content of its internal representation, thus highlighting the flexibility of the visual recognition system.

Keywords: object recognition, internal representations, affective value, context, spatial frequencies

INTRODUCTION

Rapid and accurate visual recognition of everyday objects encountered in different orientations, seen under various illumination conditions, and partially occluded by other objects in a visually cluttered environment is necessary for our survival. The first theoretical efforts to explain this feat relied on purely bottom-up mechanisms in the visual system: cells in early visual areas would be sensitive to low-level features and cells in higher areas would integrate this information in order to then match it to a representation in memory (e.g., Maunsell and Newsome, 1987). However, it is improbable that feedforward pathways alone can account for object recognition because of their severely limited information processing capabilities (Gilbert and Sigman, 2007). Moreover, since these early theoretical efforts, the essential role of such feedback mechanisms in vision has been amply demonstrated (e.g., Rao and Ballard, 1999; Tomita et al., 1999; Barceló et al., 2000; Pascual-Leone and Walsh, 2001). Nowadays, most top-down models of object recognition (e.g., Grossberg, 1980; Ullman, 1995; Friston, 2003) propose that the search for correspondence between the input pattern and the stored representations is a bidirectional process where the input activates bottom-up as well as top-down streams that simultaneously explore many alternatives; object recognition is achieved when the counter streams meet and a match is found. The content of these stored representations could depend on several factors such as task requirements (e.g., perception or action, basic-level vs. superordinate-level categorization) or categorical properties of the object (e.g., animate vs. inanimate, affective vs. non-affective, social vs. non-social; Logothetis and Sheinberg, 1996). Understanding the properties of the stored representations that lead to the generation of predictions thus is an important unexplored issue. In particular, it remains to be understood if different representational systems are used during recognition of different categories of visual objects.

Building on the predictive account of visual object recognition, Bar (2003) proposed a brain mechanism for the cortical activation of top-down processing during object recognition, where low spatial frequencies (LSFs) of the image input are projected rapidly and directly through quick feedforward connections, from early visual areas into the dorsal visual stream. Such LSF information activates a relatively small set of probable candidate interpretations of the visual input in higher prefrontal integrative centers. These initial guesses are then back-projected along the reverse hierarchy to guide further processing and gradually encompass high spatial frequencies (HSFs) available at lower cortical visual areas. This proposal is supported by neurophysiological, computational and psychophysical evidence that LSFs are processed earlier than HSFs (Watt, 1987; Schyns and Oliva, 1994; Bredfeldt and Ringach, 2002; Mermillod et al., 2005; Musel et al., 2012; for reviews, see Bullier, 2001; Bar, 2003; Hegdé, 2008) and that top-down processing in visual recognition relies on LSFs (Bar et al., 2006); moreover, magnocellular projections, which are

more sensitive to LSFs (Derrington and Lennie, 1984), seem to be implicated in initiation of top-down processing (Kveraga et al., 2007). Stored internal representations may thus be biased toward LSFs, since objects would be primarily matched in memory with an LSF draft.

Only a handful of studies have focused on the effect of specific SF band filtering during object recognition. In a name-picture verification task, low-pass filtering selectively impaired subordinate-level category verification (e.g., verify the "Siamese" category instead of the "animal" category at the superordinate level or the "cat" category at the basic level), while having little to no effect on basic-level category verification, suggesting that basic-level categorization does not particularly rely on LSFs (Collin and McMullen, 2005). On the other hand, Harel and Bentin (2009) reported that subordinate-level categorization was impaired by the removal of HSFs, but also that basic-level categorization was equally impaired by removal of either HSFs or LSFs, thus suggesting that neither of these bands is especially useful for recognition at the basic level. Finally, using a superordinate-level categorization task, Calderone et al. (2013) reported no difference in accuracy or response times between LSFs and HSFs. Overall, these studies suggest that, although this seems a bit different for subordinate-level categorization, neither LSFs or HSFs have a privileged role in object recognition. Even if LSFs do initiate a top-down processing, this suggests that their overall role in recognition is negligible; other SFs (neither low or high), however, may have a preponderant role.

Intrinsic properties of visual objects such as their affective value or contextual associativity may modulate the content of internal representations. Because of their great adaptive value, emotional objects might necessitate fast recognition, to facilitate an immediate behavioral response; this is likely to apply to both dangerous and pleasant stimuli, the former threatening survival and the latter promoting it (Bradley, 2009). In fact, the brain's prediction about the identity of a visual object may be partly based on its affective value, i.e., prior experiences of how perception of a given object has influenced internal body sensations. As such, affective value could be not just a label or judgment applied to the object post-recognition, but rather an integral component of mental object representations (Lebrecht et al., 2012) and could act as an additional clue to the object's identity to facilitate its recognition (Barrett and Bar, 2009). Since emotional objects need to be processed quickly, it is likely that LSFs, which are extracted rapidly, are particularly important for their recognition. In agreement with this idea, there is some evidence that LSFs are more present in representations of objects with strong affective value than in representations of neutral objects. Mermillod et al. (2010) reported that threatening stimuli were recognized faster and more accurately than neutral ones with LSFs but not with HSFs. Other behavioral and neuroimaging studies also suggested an interaction between emotional content and LSFs in various perceptual tasks. Bocanegra and Zeelenberg (2009), for instance, observed that in a Gabor orientation discrimination task, briefly presented fearful faces improved subjects' performance with LSF gratings while impairing it with HSF gratings. Moreover, early ERP amplitudes sensitive to affective content were found to be greater when unpleasant scenes were presented intact or in LSFs rather than in HSFs (Alorda et al., 2007). In the same vein, Vuilleumier et al. (2003) observed that the amygdala responded to fearful faces only if LSFs were present in the stimulus. In an intracranial ERP study where subjects were presented with both visible and invisible (masked) faces, Willenbockel et al. (2012) found that amygdala activation correlated mostly with SFs around 2 and 6 cycles/face, while insula activation correlated mostly with slightly higher SFs near 9 cycles/face. All these results suggest that the internal representations of objects with affective value would comprise more LSFs than representations of neutral objects.

Relatedly, the contextual associativity of a visual object—"what other objects or context might go with this object?" (Bar, 2004; Fenske et al., 2006)—could also impact on the SF content of its mental representation. It has been shown that recognition of an object that is highly associated with a certain context facilitates the recognition of other objects that share the same context (e.g., Bar and Ullman, 1996). A lifetime of visual experience would lead to contextual associations that guide expectations and aid subsequent recognition of associated visual objects through rapid sensitization of their internal representations (Biederman, 1972, 1981; Palmer, 1975; Biederman et al., 1982; Bar and Ullman, 1996). This associative processing is quickly triggered merely by looking at an object and would be critical for visual recognition and prediction (Bar and Aminoff, 2003; Aminoff et al., 2007). It has been suggested that the rapidly extracted LSFs of an object image are sufficient to activate these associated representations, and thus that the representations of contextual objects are likely to be biased toward LSFs (Bar, 2004; Fenske et al., 2006). However, this hypothesis has never been tested directly.

Affective and contextual values may also interact, so that representations of visual objects with affective value could be modulated by their contextual value or vice-versa (e.g., Storbeck and Clore, 2005; Brunyé et al., 2013; Shenhav et al., 2013). Indeed, the affective value of a given object is often defined by the context to which it has been associated to in memory. For example, a tomb elicits sadness, not because it is inherently sad, but because it evokes a context of cemetery/death. As such, affective objects might be differentially represented whether or not their affective value originates from their associated contexts. Interactions between both psychological properties have been reported. For instance, our affective state influences the breadth of the associations we make (Storbeck and Clore, 2005) and conversely, the generation of associations influences our affective state (Brunyé et al., 2013). Also, it seems that associative and affective processing both take place in the medial orbitofrontal cortex, and that both contextual and affective values might in fact relate to a more unified purpose (Shenhav et al., 2013).

The current study examined the SF content of stored internal representations of visual objects with different affective and contextual values, by evaluating what are the SFs in the stimuli that correlate with fast and accurate identification. Stimuli consisted of pictures of objects equalized in SF content and categorized as having low or high affective and contextual values. The SFs of these stimuli were randomly sampled on a trial-by-trial basis while subjects categorized the objects portrayed in the images. By varying affective value, contextual value and spatial frequencies available in the object image altogether, we aimed to clarify their roles in

visual recognition, and to study potential interactions between them.

METHODS

PARTICIPANTS

Forty-seven healthy participants (33 males) with normal or corrected-to-normal visual acuity were recruited on the campus of the Université de Montréal for an object recognition study. Participants were aged between 19 and 31 years ($M = 23.04$; $SD = 3.13$) and did not suffer from any reading disability. A written informed consent was obtained prior to the experiment, and a monetary compensation was provided upon its completion.

APPARATUS

The experimental program was run on a Mac Pro computer in the Matlab (Mathworks Inc.) environment, using functions from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A refresh rate of 120 Hz and a resolution of 1920 × 1080 pixels were set on the Asus VG278H monitor used for stimuli presentation. The relationship between RGB values and luminance levels was linearized. Luminance depth was 8 bits, and minimum and maximum luminance values were 1.1 cd/m^2 and 134.0 cd/m^2, respectively. A chin rest was used to maintain viewing distance at 76 cm.

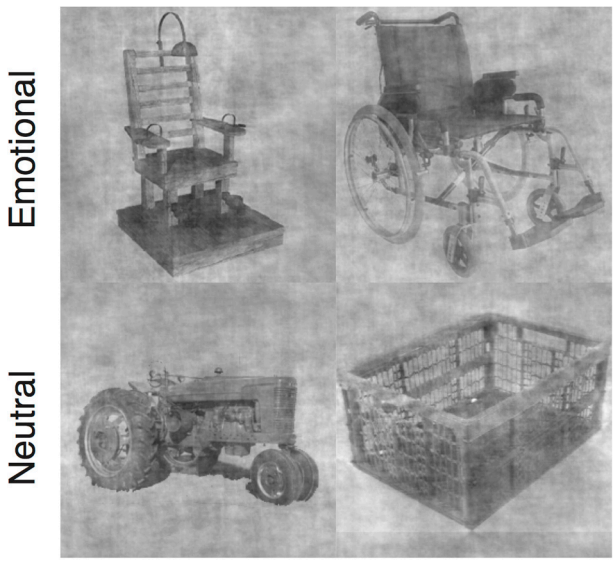


FIGURE 1 | Example images for each of the four categories of objects.

STIMULI

Selection and validation

One hundred fifty six object images were pre-selected mainly from the database used in Shenhav et al. (2013) but also from Internet searches. Each object image was presented to 30 raters who decided either (i) if they associated the object to a particular emotion, and if so, to which one or (ii) if they associated the object to a particular context, and if so, to which one. For the experiment, we selected 18 objects with clear consensus (or absence of) regarding their contextual and affective values in each of our four object categories: contextual emotional, non-contextual emotional, contextual neutral and non-contextual neutral (**Figure 1**, Table S1). Clear consensus about high affective or high contextual value meant that an object was associated to the same context or to the same emotion by more than 75% of raters; and clear consensus about low affective or contextual value meant that an object was associated to no particular context or emotion by more than 75% of the raters. Fifty-one of the selected images came from the Shenhav et al. (2013) database, and our affective and contextual ratings for these images closely matched theirs.

Control of low-level features

Stimuli thus consisted of 72 grayscaled object images of 256 × 256 pixels presented on a mid-gray background. The images subtended 6 × 6° of visual angle. Median object width was equal to 237 pixels. To target our investigation on stored internal representations and get rid of a potential interaction between the visual input and the representation, spatial frequency content and luminance were equalized across stimuli using the SHINE toolbox (Willenbockel et al., 2010a). Resulting images had a RMS contrast of 0.075. We reduced the undesired impact of psycho-linguistic factors, such as word length and lexical frequency, on response times by transforming these into z-scores for every object. For example, we computed the mean and standard deviation of the RTs of the correct positive trials in which the electric chair was presented, and we used these statistics to transform those RTs into z-scores. We did the same for all the other objects. As a result, the means and standard deviations of the RTs associated with every word were strictly identical, and all RT variations due to differences between the words were eliminated.

Sampling

SF content of the images properly padded was extracted via Fast Fourier Transform (FFT) and randomly filtered at each trial, according to the SF Bubbles method (Willenbockel et al., 2010b). In short, each spatial frequency filter was created by first generating a random vector of 10,240 elements consisting of 20 ones (the number of bubbles) among zeros. Second, the resulting vector was convolved with a Gaussian kernel that had a standard deviation of 1.8. Third, the vector was log transformed so that the SF sampling approximately fit the SF sensitivity of the human visual system (see De Valois and De Valois, 1990). The resulting sampling vector contained 256 elements representing each spatial frequency from 0.5 to 128 cycles per image. To create the two-dimensional spatial frequency filtered images, vectors were rotated about their origins and dot-multiplied with the FFT amplitudes (see Willenbockel et al., 2010b, for methodological details). Thus, several SF bandwidths were revealed in each stimulus; and objects were presented several times with different SF bandwidths revealed every time (**Figure 2**).

PROCEDURE

After they had completed a short questionnaire for general information (age, sex, education, language, etc.), participants sat

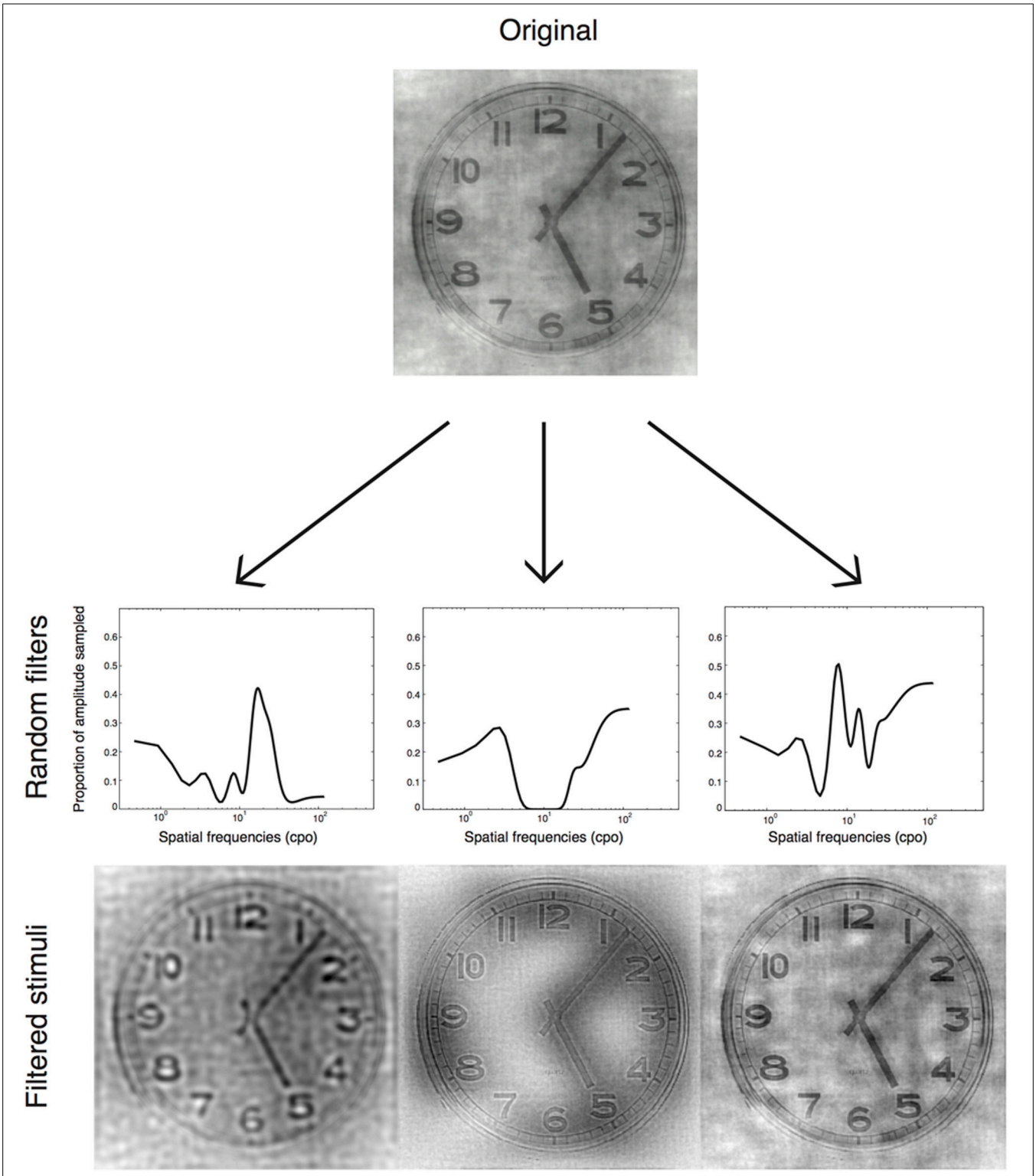


FIGURE 2 | Examples of stimuli presented in the experiment. These are generated by applying random filters to a base image.

comfortably in front of a computer monitor, in a dim-lighted room. Participants did two 500-trial blocks, with a short break in between. Each trial began with a central fixation cross lasting 300 ms, followed by a blank screen for 100 ms, the SF-filtered random object image for 300 ms, a central fixation cross for 300 ms, a blank screen for 100 ms, and finally a matching or mismatching object name that remained on the screen until the participant had answered or for a maximum of 1000 ms. Subjects

were asked to indicate with a keyboard key press as accurately and rapidly as possible whether or not the name matched the object depicted in the image. This picture-name verification task was chosen because it imposes a specific level of categorization to subjects (we chose the basic-level) without focusing attention explicitly on either affective or contextual value of the object. Name and object matched on half the trials.

SPATIAL FREQUENCY DATA ANALYSIS

To determine the spatial frequencies that contributed most to fast object recognition for each condition, we performed least-square multiple linear regressions between RTs and corresponding sampling vectors. Only correct positive trials (i.e., when the name matched the object, and the participant answered correctly) were included in the analysis. RTs were first z-scored for every object to minimize undesired sources of variability pertaining to psycholinguistic factors such as word length and lexical frequency (see Stimuli: Control of low-level features). They were further z-scored for each condition in each subject's session to diminish variability due to task learning. Trials associated with z-scores over 3 or below 3 were discarded (<1.8% of trials).

We call the resulting vectors of regression coefficients classification vectors. We first contrasted the classification vector for all objects against zero to examine what were the spatial frequencies used in general, regardless of affective or contextual values. We then contrasted the classification vectors for all emotional objects and all neutral objects, and the ones for all contextual objects and all non-contextual objects, to assess the main effects of contextual and affective values. Next, we examined if there was an interaction between these two dimensions. To do so, we contrasted classification vectors of all four subcategories of objects by applying the following formula:

$$(A_1B_1 - A_1B_2) - (A_2B_1 - A_2B_2)\,,$$

where A represents emotional value, B represents contextual value, and the number represents the level of the variable. We finally investigated the simple effects by comparing the conditions pairwise. The statistical significance of the resulting classification vectors was assessed by applying the Cluster test (Chauvin et al., 2005). Given an arbitrary z-score threshold, this test gives a cluster size above which the specified p-value is satisfied. We used this test rather than the Pixel test (Chauvin et al., 2005) because it is in general more sensitive, allowing us to detect weaker but more diffuse signals. Here, we used a threshold of ± 3 ($p < 0.05$, two-tailed). We report the size k of the significant cluster and its maximum Z-score $Z_{\max}$. We implemented the Cluster tests as bootstraps (Efron and Tibshirani, 1993); that is, we repeated all regressions 10,000 times pairing the sampling vectors with transformed RTs randomly selected in the observed transformed RT distribution. This resulted in 10,000 random classification vectors per condition. We used these random classification vectors to transform the elements of the observed classification vectors into z-scores and estimate their p-values. We corrected p-values for multiple comparisons in the pairwise comparisons by implementing Hochberg's step-up procedure (Hochberg, 1988).

RESULTS

EFFECTS OF CONDITION AND SPATIAL FREQUENCIES ON ACCURACY

The mean accuracy was 87.49% ($SD = 7.63$). To analyse possible effects of condition on accuracy, without taking SFs into account, we first conducted a 2 (Context: non-contextual or contextual) × 2 (Emotion: neutral or emotional) repeated-measures ANOVA on mean accuracies per participant. There was an effect of contextual value [$F_{(1,\ 46)} = 39.83$, $p < 0.001$, $\eta_p^2 = 0.46$]: non-contextual objects ($M = 81.92\%$; $SD = 9.21$) were recognized more easily than contextual ones ($M = 77.19\%$; $SD = 10.96$). There also was an effect of emotional value [$F_{(1,\ 46)} = 6.31$, $p < 0.05$, $\eta_p^2 = 0.12$]: neutral objects ($M = 80.30\%$; $SD = 9.48$) were recognized slightly more easily than emotional objects ($M = 78.81\%$; $SD = 10.49$).

There was an interaction between emotional and contextual values [$F_{(1,\ 46)} = 53.04$, $p < 0.001$, $\eta_p^2 = 0.53$]. This interaction was decomposed into simple effects. First, there was an effect of emotion on non-contextual objects [$F_{(1,\ 46)} = 49.63$, $p < 0.001$, $\eta_p^2 = 0.52$]. Non-contextual neutral objects ($M = 85.58\%$; $SD = 7.94$) were recognized more easily than non-contextual emotional objects ($M = 78.26\%$; $SD = 11.49$). Second, there was an effect of emotion on contextual objects as well [$F_{(1,\ 46)} = 20.87$, $p < 0.001$, $\eta_p^2 = 0.31$]. Contextual emotional objects ($M = 79.36\%$; $SD = 10.31$) were recognized more easily than neutral contextual objects ($M = 75.02\%$; $SD = 12.45$).

Accuracy did not correlate significantly with the presentation of any SF.

EFFECT OF CONDITION ON RESPONSE TIMES

The mean RT for correct positive trials was 623 ms ($SD = 83$). To analyse possible effects of condition on RTs, without taking SFs into account, we conducted a 2 (Context: non-contextual or contextual) × 2 (Emotion: neutral or emotional) repeated-measures ANOVA on $-\log(x+1)$-transformed RT means per participant (Ratcliff, 1993). Aberrant scores (over 2 s) were excluded from the analysis. There was an effect of contextual value on RTs [$F_{(1,\ 46)} = 161.29$, $p < 0.001$, $\eta_p^2 = 0.78$] whereby non-contextual objects (Md[1] = 596 ms; $SD = 60$) were recognized faster than contextual ones ($Md = 537$ ms; $SD = 67$). There was no effect of emotional value [$F_{(1,\ 46)} < 1$].

There also was an interaction between emotional value and contextual value [$F_{(1,\ 46)} = 18.46$, $p < 0.001$, $\eta_p^2 = 0.29$]. This interaction was decomposed into simple effects. First, there was an effect of emotion on non-contextual objects [$F_{(1,\ 46)} = 12.53$, $p < 0.001$, $\eta_p^2 = 0.21$]. Non-contextual neutral objects ($Md = 532$ ms; $SD = 57$) were identified faster than non-contextual emotional objects ($Md = 548$ ms; $SD = 68$). There also was an effect of emotion on contextual objects [$F_{(1,\ 46)} = 10.15$, $p < 0.01$, $\eta_p^2 = 0.18$]. Contextual emotional objects ($Md = 579$ ms; $SD = 64$) were identified faster than contextual neutral ones ($Md = 609$ ms; $SD = 80$).

[1]Median reaction times are given, since the ANOVA was performed on log transformed values. Given that the mean log values wouldn't be readily interpretable and that the median values don't change with a log transformation, we made this choice for purposes of clarity and transparency.

EFFECT OF SPATIAL FREQUENCIES ON RESPONSE TIME

To determine the spatial frequencies that contributed most to fast object recognition for each condition, we performed least-square multiple linear regressions between z-scored transformed RTs (see Methods: Spatial Frequency Data Analysis) and corresponding sampling vectors for correct positive trials. All object categories confounded, SFs between 13.71 and 24.31 cycles per object width (cpo) correlated negatively with RTs (peak at 19.45 cpo, $Z_{max} = 3.94$, $k = 23$, $p < 0.01$; **Figure 3A**). In other words, RTs were consistently reduced with the presentation of SFs within these boundaries. To examine a possible effect of emotional value, we contrasted classification vectors for all emotional objects and all neutral objects. There was no significant difference ($p > 0.05$). Similarly, there was no significant difference between non-contextual and contextual objects ($p > 0.05$).

We then examined the interaction between affective and contextual values (see Methods: Spatial frequency data analysis). We found a significant interaction for SFs between 5.52 and 6.69 cpo (peak at 6.02 cpo, $Z_{max} = 3.29$, $k = 3$, $p < 0.05$; **Figure 3B**).

We subsequently decomposed the interaction into simple effects. There was a significant effect of contextual value on neutral objects between 15.25 and 19.20 cpo; these SFs were correlated more negatively with RTs for contextual neutral objects than for non-contextual neutral objects (peak at 18.98 cpo, $Z_{max} = 3.36$, $k = 9$, $p < 0.05$, corrected for multiple comparisons; **Figure 3C**). However, the interaction was not significant for these SFs, making this effect difficult to interpret. There also was an effect of contextual value for emotional objects: SFs between 4.86 and 6.56 cpo correlated more positively with RTs for contextual emotional objects than for non-contextual emotional objects (peak at 5.56 cpo, $Z_{max} = 3.75$, $k = 4$, $p < 0.05$, corrected for multiple comparisons; **Figure 3D**). Moreover, there was an effect of emotional value on contextual objects: SFs between 4.86 and 6.09 cpo correlated more positively with RTs for contextual emotional objects than for contextual neutral objects (peak at 5.56 cpo, $Z_{max} = 3.21$, $k = 3$, $p < 0.05$, corrected for multiple comparisons; **Figure 3E**). Finally, we observed no significant difference between non-contextual neutral and non-contextual emotional objects ($p > 0.05$). The interaction thus seems to be caused by the significant effect of contextual value on emotional but not on neutral objects, combined with the significant effect of emotional value on contextual but not on non-contextual objects.

DISCUSSION

GENERAL SPATIAL FREQUENCY USE

A few studies have examined the effect of specific SF band filtering during name-picture verification tasks, similar to ours. Collin and McMullen (2005) reported that low-pass filtering objects had little impact on basic-level verification (e.g., verify the "cat" category instead of the "animal" category at the superordinate level or the "Siamese" category at the subordinate level), suggesting that basic-level categorization does not especially rely on LSFs. Furthermore, Harel and Bentin (2009) reported that basic-level categorization was equally impaired by removal of either HSFs or LSFs, thus suggesting that neither of these bands is especially

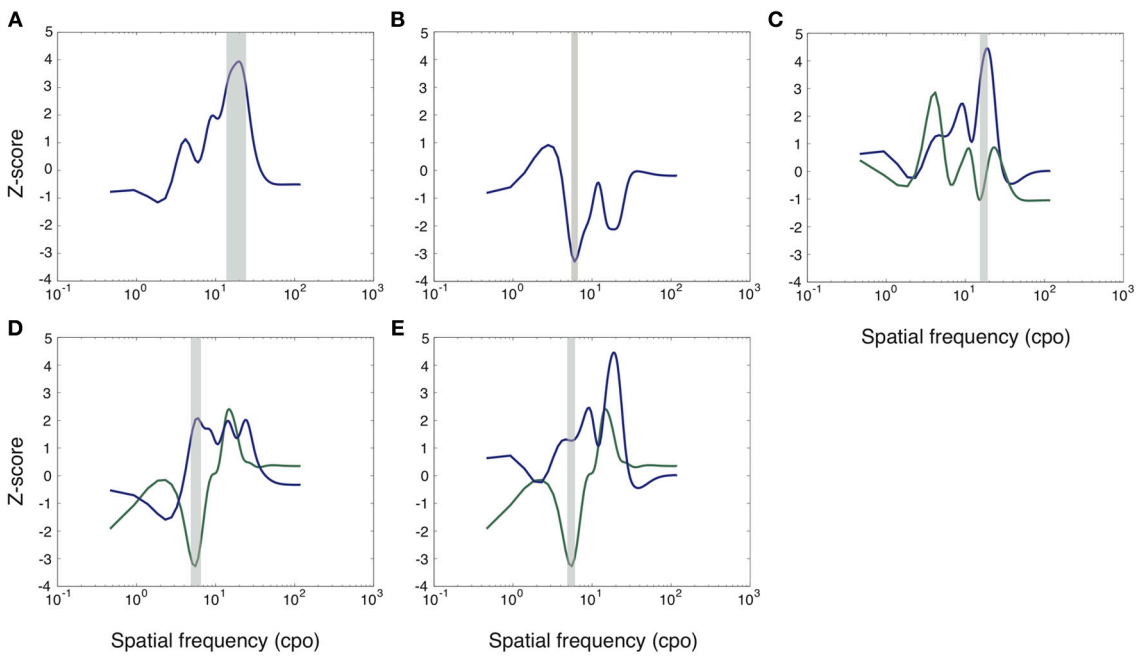


FIGURE 3 | Group classification vectors depicting the correlations between SFs and RTs for different conditions. Higher z-scores indicate a negative correlation (SFs leading to shorter RTs) while lower z-scores indicate a positive correlation (SFs leading to longer RTs). Highlighted gray areas are significant ($p < 0.05$). See text for details. **(A)** All objects together. **(B)** The vector depicting potential interactions between both variables, obtained by contrasting the contrasts of contextual value for both levels of emotional value. **(C)** Non-contextual neutral (green) objects and contextual neutral (blue) objects. **(D)** Contextual emotional (green) and non-contextual emotional (blue) objects. **(E)** Contextual emotional (green) and contextual neutral (blue) objects.

useful for recognition at the basic level. However, Harel and Bentin's cutoff for HSFs was especially high (65 cpo, or 6.5 cpd), thus preserving only very fine information typically not useful for object recognition. A large band of intermediate spatial frequencies was not explored in these studies.

An important aspect of our study is that instead of applying filters with fixed arbitrary cut-offs, we randomly sampled the entire SF spectrum. This allowed us to overcome the need of selecting arbitrary SF bands to evaluate. Indeed, there is no consensus in the literature about what consists of LSFs or HSFs: this seems to be more understood as a relative measure for SF bands inside a given study. Cut-offs for LSFs in the literature vary from 5 cpo (Boutet et al., 2003) to 15 cpo (Alorda et al., 2007). Similarly, cut-offs for HSFs vary from 20 cpo (Boutet et al., 2003) to 65 cpo (Harel and Bentin, 2009). When cut-offs are translated into cycles per degree (cpd), acknowledging that the diagnostic SFs may vary according to viewing distance, the discrepancy is even larger: cut-offs for LSFs vary from less than 0.4 cpd (Boutet et al., 2003) to more than 2.4 cpd (Alorda et al., 2007) and cut-offs for HSFs vary from 1.4 cpd (Boutet et al., 2003) to 6.5 cpd (Harel and Bentin, 2009). Quite interestingly, we note that some SFs (between 1.4 and 2.4 cpd) may be included either in LSFs or HSFs.

Our random sampling of the entire SF spectrum allowed us to evaluate the use of SFs considered as neither low nor high by most previous studies. Using this unbiased experimental approach, we found that intermediate SFs between about 14 and 24 cpo (2.3–4 cpd) are associated with fast RTs for basic-level verification. This suggests that objects are processed particularly rapidly through these SFs. Although this interpretation is the most straightforward, it is also possible that object processing was at least partly completed before the presentation of the words and, therefore, that the RTs reflect remnants of object processing rather than object processing *per se*.

Another unique aspect of our study is the fact that we equalized SF content of the object images prior to their sampling. This allows us to interpret results more confidently in terms of content of internal representations. Indeed, if SF content is not normalized among stimuli, results most likely reflect an interaction of the stored representation with the information available in the stimulus. Unfortunately, few studies have applied this procedure. As a notable exception, Willenbockel et al. (2010b) did equalize SF spectrum and randomly sample SFs in a face recognition task. Results revealed that SFs peaking at approximately 9 and 13 cycles/face (equivalent to 1.4 and 2 cpd, i.e., SFs that may be categorized as LSFs, HSFs, or most often neither of these) were most correlated with fast and accurate face identification. Although these SFs specific for images of faces are likely to differ from the SF content of object representations, they are an additional indicator that, as in the present study, intermediate SFs rather than LSFs occupy the greatest place in our representation of the world. It is plausible that stored representations consist of mostly these SFs because they are part of the intermediate band of SFs to which we are naturally most sensitive (e.g., Watson and Ahumada, 2005).

INTERACTION BETWEEN AFFECTIVE AND CONTEXTUAL VALUES

No main effect of contextual or affective value was observed in the SFs correlating with the objects' fast identification. However, we found a significant interaction between affective and contextual values for SFs centered on 6 cpo (or 1 cpd). This indicates that these LSFs, those usually associated with the magnocellular pathway (Derrington and Lennie, 1984), are sensitive in a non-linear manner to a combination of the visual object's intrinsic properties.

When testing the simple effects, we observed that affective value elicited a significant difference in the use of these SFs in contextual objects: they led to longer RTs for contextual emotional objects than for contextual neutral ones. This is not in accordance with the general effect of affective value usually reported in the literature (i.e., LSFs leading to faster RTs, e.g., Mermillod et al., 2010); however, our result is due to an interaction between affective and contextual values and is therefore difficult to compare to those of other studies. Moreover, our stimuli were equalized in their SF content and always comprised several randomly sampled SF bandwidths at the same time, whereas in studies using filters with fixed cut-offs, only some specific band of LSFs or HSFs is shown at a time.

SFs near 6 cpo (or 1 cpd) also led to longer RTs for contextual emotional objects than for non-contextual emotional objects. The effect of contextual value on SF content of object representations had not been tested before but it had been often proposed that rapidly extracted LSFs are sufficient to activate representations associated with an object (Bar, 2004; Fenske et al., 2006). Our data suggest that these presumed/hypothetical associative representations do not speed up the object's recognition. Why we observed this modulation only for emotional objects is not clear, but several interactions between affective and contextual processing have already been reported and could possibly explain the discrepancy (Storbeck and Clore, 2005; Brunyé et al., 2013; Shenhav et al., 2013). For example, affective value might influence the extent to which we associate a particular object to other objects (Bar, 2009; Shenhav et al., 2013).

CONCLUSION

The main findings of the present study are (i) that the SF content of object representations in general are in an intermediate band between 14 and 24 cpo (2.3–4 cpd), and (ii) that intrinsic high-level categorical properties of an object influence the SF content of its internally stored representation, more precisely that affective and contextual values interact in their modulation of the SF content of object representations.

According to predictive accounts of brain function (e.g., Rao and Ballard, 1999; Bar, 2003; Friston, 2003, 2010; Friston et al., 2006), our mind constantly generates predictions about our environment, and our understanding of a sensory input is based both on the available sensory information and on prior beliefs stored as internal representations (see Knill and Pouget, 2004). In this study, we investigated precisely the SF content of these stored representations, and its potential flexible modulation by affective and contextual properties of the stimulus. Our results reveal that stored representations of visual objects are composed of intermediate SFs that are often left over in studies using filters with fixed arbitrary cut-offs. Furthermore, we observed a modulation of this SF content by affective and contextual intrinsic values of the visual object, suggesting its flexibility and thus the multiplicity of visual recognition systems.

Our study cannot however address directly the issue of temporal dynamics of visual object recognition. While we observed that some SFs are more useful to identify some objects, we cannot conclude that these are extracted first. Further studies should address these issues and their links to potential initiation of top-down mechanisms.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00512/abstract

REFERENCES

Alorda, C., Serrano-Pedraza, I., Campos-Bueno, J. J., Sierra-Vázquez, V., and Montoya, P. (2007). Low spatial frequency filtering modulates early brain processing of affective complex pictures. *Neuropsychologia* 45, 3223–3233. doi: 10.1016/j.neuropsychologia.2007.06.017

Aminoff, E., Gronau, N., and Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb. Cortex* 17, 1493–1503. doi: 10.1093/cercor/bhl078

Barceló, F., Suwazono, S., and Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nat. Neurosci.* 3, 399–403. doi: 10.1038/73975

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629. doi: 10.1038/nrn1476

Bar, M. (2009). Predictions: a universal principle in the operation of the human brain. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1181–1182. doi: 10.1098/rstb.2008.0321

Bar, M., and Aminoff, E. (2003). Cortical analysis of visual context. *Neuron* 38, 347–358. doi: 10.1016/S0896-6273(03)00167-3

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454. doi: 10.1073/pnas.0507062103

Bar, M., and Ullman, S. (1996). Spatial context in recognition. *Perception* 25, 343–352. doi: 10.1068/p250343

Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1325–1334. doi: 10.1093/brain/106.2.473

Biederman, I. (1972). Perceiving real-world scenes. *Science* 177, 77–80. doi: 10.1126/science.177.4043.77

Biederman, I. (1981). Do background depth gradients facilitate object identification? *Perception* 10, 573–578. doi: 10.1068/p100573

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 14, 143–177. doi: 10.1016/0010-0285(82)90007-X

Bocanegra, B. R., and Zeelenberg, R. (2009). Emotion improves and impairs early vision. *Psychol. Sci.* 20, 707–713. doi: 10.1111/j.1467-9280.2009.02354.x

Boutet, I., Collin, C., and Faubert, J. (2003). Configural face encoding and spatial frequency information. *Percept. Psychophys.* 65, 1078–1093. doi: 10.3758/BF03194835

Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* 46, 1–11. doi: 10.1111/j.1469-8986.2008.00702.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Bredfeldt, C. E., and Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *J. Neurosci.* 22, 1976–1984.

Brunyé, T. T., Gagnon, S. A., Paczynski, M., Shenhav, A., Mahoney, C. R., and Taylor, H. A. (2013). Happiness by association: breadth of free association influences affective states. *Cognition* 127, 93–98. doi: 10.1016/j.cognition.2012.11.015

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Rev.* 36, 96–107. doi: 10.1016/S0165-0173(01)00085-6

Calderone, D. J., Hoptman, M. J., Martínez, A., Nair-Collins, S., Mauro, C. J., Bar, M., et al. (2013). Contributions of low and high spatial frequency processing to impaired object recognition circuitry in schizophrenia. *Cereb. Cortex* 23, 1849–1858. doi: 10.1093/cercor/bhs169

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., and Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J. Vis.* 5, 659–667. doi: 10.1167/5.9.1

Collin, C. A., and McMullen, P. A. (2005). Subordinate-level categorisation relies on high spatial frequencies to a greater degree than basic-level categorisation. *Percept. Psychophys.* 67, 354–364. doi: 10.3758/BF03206498

Derrington, A. M., and Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J. Physiol.* 357, 219–240.

De Valois, R. L., and De Valois, K. K. (1990). *Spatial Vision*. New York, NY: Oxford University Press.

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall/CRC.

Fenske, M. J., Aminoff, E., Gronau, N., and Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Prog. Brain Res.* 155, 3–21. doi: 10.1016/S0079-6123(06)55001-0

Friston, K. (2003). Learning and inference in the brain. *Neural Netw.* 16, 1325–1352. doi: 10.1016/j.neunet.2003.06.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 607–608. doi: 10.1038/nrn2787-c2

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol.* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001

Gilbert, C. D., and Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696. doi: 10.1016/j.neuron.2007.05.019

Grossberg, S. (1980). Biological competition: decision rules, pattern formation, and oscillations. *Proc. Natl. Acad. Sci. U.S.A.* 77, 2338–2342. doi: 10.1073/pnas.77.4.2338

Harel, A., and Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: implications for perceptual categorization hierarchies. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1264–1273. doi: 10.1037/a0013621

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439. doi: 10.1016/j.pneurobio.2007.09.001

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–802. doi: 10.1093/biomet/75.4.800

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Kveraga, K., Boshyan, J., and Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240. doi: 10.1523/JNEUROSCI.3481-07.2007

Lebrecht, S., Bar, M., Barrett, L. F., and Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Front. Psychol.* 3:107. doi: 10.3389/fpsyg.2012.00107

Maunsell, J., and Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annu. Rev. Neurosci.* 10, 363–401. doi: 10.1146/annurev.ne.10.030187.002051

Mermillod, M., Droit-Volet, S., Devaux, D., Schaefer, A., and Vermeulen, N. (2010). Are coarse scales sufficient for fast detection of visual threat? *Psychol. Sci.* 21, 1429–1437. doi: 10.1177/0956797610381503

Mermillod, M., Guyader, N., and Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: evidence from neuro-computational modeling. *Brain Cogn.* 57, 151–157. doi: 10.1016/j.bandc.2004.08.035

Musel, B., Chauvin, A., Guyader, N., Chokron, S., and Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLoS ONE* 7:e38493. doi: 10.1371/journal.pone.0038493

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Mem. Cognit.* 3, 519–526. doi: 10.3758/BF03197524

Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512. doi: 10.1126/science.1057099

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Ratcliff, R. (1993). Methods for dealing with reaction outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dpendent scene recognition. *Psychol. Sci.* 5, 195–200. doi: 10.1111/j.1467-9280.1994.tb00500.x

Shenhav, A., Barrett, L., and Bar, M. (2013). Affective value and associative processing share a cortical substrate. *Cogn. Affect. Behav. Neurosci.* 13, 46–59. doi: 10.3758/s13415-012-0128-4

Storbeck, J., and Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: mood and the false memory effect. *Psychol. Sci.* 16, 785–791. doi: 10.1111/j.1467-9280.2005.01615.x

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., and Miyashita, Y. (1999). Top-down from prefrontal cortex in executive control of memory retrieval. *Nature* 401, 699–703. doi: 10.1038/44372

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex* 5, 1–11. doi: 10.1093/cercor/5.1.1

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631. doi: 10.1038/nn1057

Watson, A. B., and Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *J. Vis.* 5, 6–6. doi: 10.1167/5.9.6

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *J. Opt. Soc. Am. A* 4, 2006–2021. doi: 10.1364/JOSAA.4.002006

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010b). Does face inversion change spatial frequency tuning? *J. Exp. Psychol. Hum. Percept. Perform.* 36, 122–135. doi: 10.1037/a0016465

Willenbockel, V., Lepore, F., Nguyen, D. K., Bouthillier, A., and Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions—an intracranial ERP study. *Front. Psychol.* 3:237. doi: 10.3389/fpsyg.2012.00237

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010a). Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* 42, 671–684. doi: 10.3758/BRM.42.3.671

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: 01 February 2014; accepted: 09 May 2014; published online: 28 May 2014.
Citation: Caplette L, West G, Gomot M, Gosselin F and Wicker B (2014) Affective and contextual values modulate spatial frequency use in object recognition. Front. Psychol. ***5****:512. doi: 10.3389/fpsyg.2014.00512*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Caplette, West, Gomot, Gosselin and Wicker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 26 June 2014
doi: 10.3389/fpsyg.2014.00672

The neural mechanisms for the recognition of face identity in humans

Stefano Anzellotti [1,2] **** and Alfonso Caramazza*** [1,2]

[1] *Department of Psychology, Harvard University, Cambridge, MA, USA*
[2] *Center for Mind/Brain Sciences, University of Trento, Trento, Italy*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Aude Oliva, Massachusetts Institute of Technology, USA
Chris Fields, New Mexico State University, USA (retired)

****Correspondence:***
Stefano Anzellotti, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA
e-mail: anzellot@fas.harvard.edu

Every day we encounter dozens of people, and in order to interact with them appropriately we need to recognize their identity. The face is a crucial source of information to recognize a person's identity. However, recognizing the identity of a face is challenging because it requires distinguishing between very similar images (e.g., the front views of two different faces) while categorizing very different images (e.g., a front view and a profile) as the same person. Neuroimaging has the whole-brain coverage needed to investigate where representations of face identity are encoded, but it is limited in terms of spatial and temporal resolution. In this article, we review recent neuroimaging research that attempted to investigate the representation of face identity, the challenges it faces, and the proposed solutions, to conclude that given the current state of the evidence the right anterior temporal lobe is the most promising candidate region for the representation of face identity.

Keywords: faces, identity, fMRI, object recognition, invariance

INTRODUCTION

In this paper, we focus on recent neuroimaging research that has investigated aspects of the neural mechanisms underlying the perceptual recognition of face identity. The ability to recognize individuals is crucial for guiding behavior – it allows us to retrieve information about people and interact with them in appropriate ways. Many different cues can be used to recognize an individual, including the appearance of the face, the sound of the voice, as well as the context in which we encounter a person and prior knowledge about his/her current general location (see Oliva and Torralba, 2007; Goesaert and Op de Beeck, 2013). A promising approach consists in studying how each of these cues is processed when other cues are controlled, to then proceed with an investigation of how the different cues are integrated. Among the different cues that can be used for person recognition, the face is a crucial source of information and is usually sufficient in isolation to recognize a person's identity. However, recognizing face identity is also computationally challenging: it requires discounting identity-irrelevant changes in sensory stimulation (such as changes in viewpoint and illumination) without losing the ability to perform fine-grained discriminations needed to distinguish the faces of similar individuals.

The earliest insights into the neural mechanisms underlying the ability to recognize face identity came from the study of patients with selective impairment for the recognition of faces (Charcot, 1883; Wilbrand, 1892; Heidenhain, 1927; Jossmann, 1929), which was subsequently named prosopagnosia (Bodamer, 1947). Hecaen and Angelergues (1962) investigated the location of lesions producing selective deficits for faces in a group of 22 patients, and observed that prosopagnosic patients tended to have lesions in the right hemisphere, often involving occipital regions. A review of the neuropsychological literature individuated the right occipitotemporal cortex as the most common location of the lesion in prosopagnosic patients (Meadows, 1974). Convergent evidence in support of the view that damage to the occipitotemporal cortex leads to prosopagnosia was reported in several studies (Whiteley and Warrington, 1977; Damasio et al., 1982; Malone et al., 1982).

Other neuropsychological studies reported deficits for the recognition of familiar and famous faces in patients with herpes simplex encephalitis (Warrington and Shallice, 1984; Warrington and McCarthy, 1988) and semantic dementia (Snowden et al., 2004), with more frequent face recognition deficits in the right than in the left temporal variant of semantic dementia (Thompson et al., 2003). These pathologies affect the anterior portions of the temporal lobe (Kapur et al., 1994; Mummery et al., 2000; Gitelman et al., 2001; Hodges and Patterson, 2007; Noppeney et al., 2007). Furthermore, the highest lesion overlap in patients with face recognition deficits was found the be in the right anterior temporal lobe (Tranel et al., 1997). Consistent with the neuropsychological literature, neuroimaging studies in healthy participants individuated regions showing stronger activity for faces than for other kinds of objects in occipitotemporal cortex [occipital face area (OFA) and fusiform face area (FFA); Sergent et al., 1992; Puce et al., 1996; Kanwisher et al., 1997; Gauthier et al., 2000; see Çukur et al., 2013 for an in-depth analysis of voxel response profiles] and the anterior temporal lobes (Rajimehr et al., 2009).

Both occipitotemporal regions and anterior temporal regions show stronger activity for faces than other objects, and lesions in these regions lead to face processing deficits. What are the respective contributions of the two brain regions in representing face identity? The finding that lesion to a brain region leads to a deficit for face recognition does not imply that that region encodes representations of face identity – it might just provide necessary input to another region that represents face identity. At the same time, neither occipitotemporal nor anterior temporal regions seem to be involved merely in the processing of "low level"

perceptual details. Patients with anterior temporal lesions have intact basic perceptual abilities (Warrington and Shallice, 1984), and while patients with occipitotemporal lesions often have visual field defects (Meadows, 1974), they are able to describe and draw individual face parts (Bodamer, 1947). A deeper understanding of the properties of representations in these regions is needed to clarify their respective roles for the recognition of face identity. This paper is concerned with the neuroimaging research pursuing this understanding. In particular, the focus is on perceptual representations of face identity, rather than on other aspects of person identity such as associated semantic knowledge (Tsukiura et al., 2002), or the sense of familiarity and emotional responses which can be impaired in disorders such as Capgras syndrome (Ellis and Lewis, 2001).

DISCRIMINATION OF FACE TOKENS

Before delving into the discussion of the literature, it is necessary to introduce some terms and clarify their use. We will use the term "face token" to refer to a specific image of a face, seen from a particular viewpoint and under a particular illumination. The recognition of face identity requires (1) to distinguish between face tokens that depict different people, and (2) to recognize when two different face tokens depict the same person. We will use the term "invariant face representations" to refer to representations that encode information about whether two face tokens depict the same person, for *some or all* pairs of face tokens that depict a same person. Note that invariance can be partial, for example, there might be representations that are invariant to changes in viewpoint of up to 35°. Therefore, not all invariant face representations are representations of face identity. We will reserve the term "representation of face identity" for representations that encode information that allows determining that two face tokens depict the same person for all pairs of face tokens that are recognized as a same person by a human observer. Whether or not there exists one brain region that encodes representations with invariance across all transformations that humans can generalize across is an empirical question. To search for representations of face identity, we can first search for representations that distinguish between face tokens that depict different people, and then test whether and to which extent they are invariant. Finding brain regions that distinguish between face tokens that depict different people provides us with a series of potential candidates for the representation of face identity.

The investigation of regions that distinguish between face tokens that depict different people with functional magnetic resonance imaging (fMRI) is challenging, because when properties like viewpoint and illumination are controlled, face tokens that depict different people do not produce significantly different blood-oxygen-level dependent (BOLD) responses when analyzed with standard univariate approaches. Nonetheless, fMRI remains one of the best methods available to localize regions that distinguish between face tokens that depict different people. This is because it allows coverage of a large extent of the human brain in a single study, and because among the methods with this property it is the one that offers the highest spatial resolution.

For this reason, in the course of the past two decades, researchers used fMRI to investigate the neural mechanisms underlying the recognition of face identity, developing and employing experimental designs and data analysis approaches to meet the challenge posed by the subtle differences in the BOLD responses produced by different face tokens.

One approach to individuating representations that distinguish between face tokens that depict different people involves using fMRI-adaptation (fMR-A). FMR-A is a phenomenon characterized by reduced BOLD responses to repeated stimuli (Grill-Spector et al., 1999). FMR-A has also been observed during the presentation of two stimuli that are not identical but are similar along some dimension (Grill-Spector et al., 1999; Vuilleumier et al., 2002). For example, fMR-A can occur for the presentation of different stimuli from the same category (Fairhall et al., 2011). FMR-A has been used to investigate representations of face tokens in a series of studies (Grill-Spector et al., 1999; Gauthier et al., 2000; Rotshtein et al., 2004; Furl et al., 2007). Greater adaptation for repetitions of a same face token than for the presentation of different face tokens has been observed in the FFA (Gauthier et al., 2000), as well as in occipitotemporal regions defined with a broader contrast between faces and textures (Grill-Spector et al., 1999).

As an alternative to fMR-A, some researchers have used multivariate pattern analysis (MVPA) to improve the sensitivity of fMRI (Haxby et al., 2001; Haynes and Rees, 2006). Multivariate approaches extract information from the pattern of activity in multiple voxels. They are more sensitive than univariate approaches, because they can distinguish between BOLD responses within a region that have the same mean but different spatial distributions.

A common method consists in using univariate analyses in order to individuate regions showing stronger responses to faces than other objects ("face-selective" regions) and subsequently investigate information content with MVPA within these regions. With this regions-of-interest (ROI) approach it has been shown that face-selective regions, including notably the FFA, encode information about face tokens (Nestor et al., 2011; Anzellotti et al., 2013; Goesaert and Op de Beeck, 2013; Verosky et al., 2013; but see Natu et al., 2010). However, this approach is based on the implicit assumption that localizing the brain regions showing the greatest mean difference between the activity in response to faces and the activity in response to other objects exhaustively captures the regions involved in the recognition of face identity. This assumption might not hold: there may be regions that do not show face-selectivity but still contribute to the recognition of face identity.

An alternative to the use of face selectivity is searchlight analysis (Kriegeskorte et al., 2006; Kriegeskorte and Bandettini, 2007) to individuate regions that distinguish between face tokens in the whole brain. In an early study (Kriegeskorte et al., 2007), searchlight was used to detect information that distinguishes between face tokens in the right anterior temporal lobe. The faces that were distinguished, though, were of different genders. A more recent study (Nestor et al., 2011) used searchlight and individuated information that distinguishes between face tokens of the same gender in the right anterior temporal lobe and posterior temporal cortex bilaterally.

Another method that can be used to individuate information that distinguishes between face tokens is recursive feature elimination (RFE), a type of MVPA (De Martino et al., 2008; Formisano et al., 2008). RFE has advantages (and some disadvantages) with respect to both ROI-based and searchlight methods. RFE can individuate information that is distributed beyond the extent of a searchlight sphere. It does not require that a set of contiguous voxels classify the different conditions significantly above chance; that is, informative voxels can be anywhere in the brain. This also means that feature selection approaches do not require making arbitrary choices about the size and shape of the regions within which to search for information. In addition, RFE requires that the individuated voxels contribute themselves to the discrimination, while in the case of searchlight an individuated voxel does not necessarily contribute to the discrimination: as long as other voxels within the sphere provide significant classification accuracy, the voxel will appear in the searchlight map, even if the voxel itself is not informative (this is especially true for SVM-based searchlight, see Etzel et al., 2013). The main disadvantage of RFE is that in its current form it allows localization of voxels that contribute to a given classification, but unlike searchlight and representational similarity analysis (RSA) it does not allow localization of regions based on a match between a neural dissimilarity matrix and a dissimilarity matrix hypothesized by the experimenter. However, for the purpose of localization of regions involved in the representation of face tokens this is not a major concern. To date, RFE has produced promising results for the localization of regions that distinguish between face tokens that depict different people (**Figure 1**), allowing localization of informative voxels for the discrimination between gender-matched faces in occipitotemporal and anterior temporal regions (Nestor et al., 2011; Anzellotti et al., 2013), and in the posterior cingulate and the posterior intraparietal sulcus (Anzellotti and Caramazza, 2014).

In sum, regions that distinguish between face tokens that depict different people have been found in occipitotemporal cortex bilaterally, in the anterior temporal lobes, in posterior cingulate and in bilateral IPS. Very recent studies (Cowen et al., 2014; Nestor et al., 2014) adopted principal component analysis (PCA) and independent component analysis (ICA) to investigate classification for larger numbers of face tokens, going beyond the small number of identities used in most studies to date.

INVARIANT FACE REPRESENTATIONS

Regions that distinguish between face tokens that depict different people are candidate regions for representing face identity, but not all of them necessarily encode representations of face identity. To individuate regions that represent face identity, it is important to investigate whether they encode invariant face representations. Studies investigating the invariance of face representations typically look for evidence of commonalities among representations of different face tokens that depict the same person. For this reason, it is particularly important to carefully control the stimuli used because the presence of commonalities in the low-level properties of different face tokens depicting a same person can lead to illusory invariance effects. Equating the average luminance, color and texture in the whole image is often insufficient as a control because visually responsive neurons at several stages of processing have local receptive fields that do not encompass the entire image. These challenges can be overcome by generating stimuli with computer graphics. Using computer graphics permits the careful control of the low-level

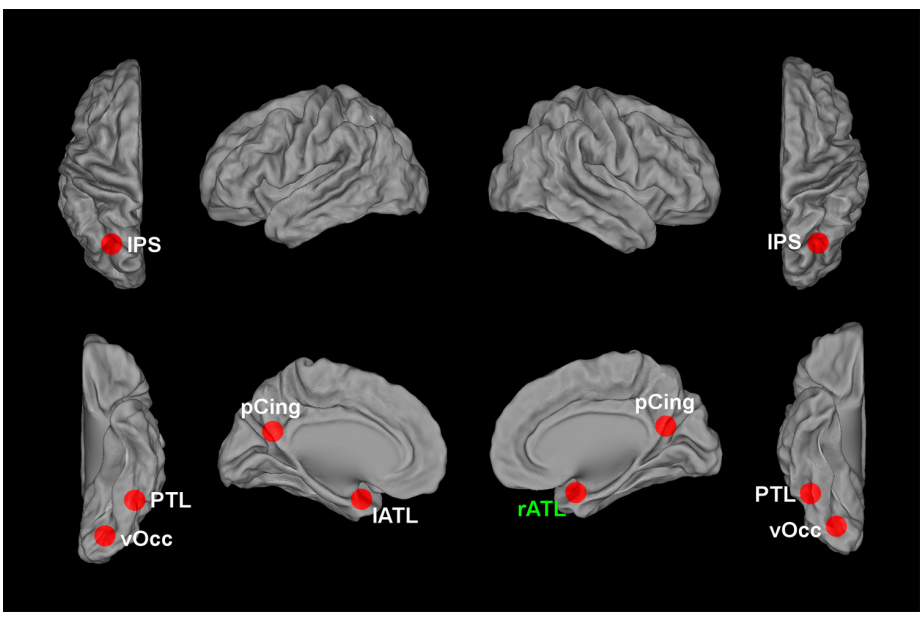


FIGURE 1 | Brain regions encoding information that contributes to the classification between different face tokens corresponding to different individuals. vOcc, ventral occipital cortex; PTL, posterior temporal lobe; ATL, anterior temporal lobe; pCing, posterior cingulate; IPS, intraparietal sulcus. The current evidence indicates the right ATL, marked in green, as the most likely candidate region for encoding invariant representations of face identity.

differences between face tokens at a local level (Anzellotti et al., 2013; Anzellotti and Caramazza, 2014). Since even cartoon faces elicit strong responses in face-selective neurons (Freiwald et al., 2009), it is unlikely that the use of realistic 3D renderings of faces would bias the results with respect to the use of photographs.

fMRI-adaptation can be used not only to individuate regions sensitive to differences in identity, but also to search for commonalities among representations of different face tokens that depict a same person. If a region encodes invariant face representations, the representations of different face tokens depicting the same person should overlap more than the representations of face tokens depicting different people, and therefore more fMR-A should be observed for the presentation of different face tokens that depict a same person than face tokens of different persons. One problem with the underlying assumptions motivating the use of fMR-A to study invariant face representations is that even if we accept that regions encoding invariant face representations should show fMR-A for the presentation of different face tokens depicting a same person, it does not follow that all regions that show fMR-A for the presentation of different face tokens depicting a same person encode invariant face representations. One way in which a region could show fMR-A for different face tokens depicting a same person despite encoding non-invariant face representations is through top-down influences. Via top-down influences, recognition of two different images as tokens depicting a same identity could lead to reduced activity not only in regions encoding invariant representations but also in early visual regions. Whether or not reduction in neural activity due to repetition can occur as a consequence of top-down influences is controversial (Xiang and Brown, 1998; Schendan and Kutas, 2003).

Several studies investigated invariant face representations using fMR-A, with mixed results: some studies found evidence for adaptation (Vuilleumier et al., 2002) while others did not (Pourtois et al., 2005). Ewbank and Andrews (2008) found fMR-A for repetition of face identity across different viewpoints in FFA when presenting familiar faces, but not when presenting novel faces. The likelihood of observing adaptation across different face tokens depicting a same person in fMR-A studies seems to be a function of the duration of the lag between two stimuli, with longer lags leading to more invariance in some studies (Andresen et al., 2009), but it remains unclear what are the mechanisms at the basis of this phenomenon. A recent study (Mur et al., 2010) found fMR-A for the repetition of face identity across different viewpoints in several regions, including early visual cortex. Given the current understanding of representations in early visual cortex, it is unlikely that this region carries invariant face representations. Findings such as this suggest that fMR-A can occur due to top-down influences.

To overcome the interpretative challenges that arise in fMR-A studies, invariant face representations have been investigated with MVPA. Experiments designed to investigate invariance with MVPA typically involve the presentation of multiple different tokens (e.g., different facial expressions, different viewpoints) of each face identity. The BOLD responses to those face tokens are then split into a subset used for the training of a classifier (for instance a support vector machine), and a subset used for the testing of the performance of the trained classifier. A possible approach is to split the data into subsets so that each part contains responses to all stimuli shown. In this case, the training and testing subsets contain the BOLD signal in response to *different* presentations of the *same identical* images. This analysis approach is *not* circular (data from different runs are used for the training and testing of classifiers), but since responses to the same images are used for training and testing, the classifier could potentially achieve significant classification accuracy relying on representations that are not invariant.

Despite these remarks, a recent study (Nestor et al., 2011) used this approach and found accuracies significantly above chance in FFA but at chance in early visual cortex for the classification of face identity in the presence of different facial expressions (Nestor et al., 2011). The robust classification accuracies obtained in this study (Nestor et al., 2011) are probably due to the contribution of invariant representations. However, other studies reported significant classification accuracy for faces seen from different viewpoints even in early visual cortex when using this method (Anzellotti et al., 2013). This is in contrast with the current understanding of representations in early visual cortex, and suggests that the conclusions obtained with this method should be interpreted with caution.

A more stringent method that overcomes the concerns discussed above consists in splitting the data into subsets so that the responses to different viewing conditions are included in the training and the testing set. In this case, the training and testing subsets contain the BOLD signal in response to *different* images. Using this method, classification across different viewpoints was at chance in early visual cortex, but was significant in other ventral stream regions (Anzellotti et al., 2013). In particular, even when using the responses to different stimuli for training and testing, and controlling carefully the "low-level" properties of images, significant classification generalizing across viewpoints was observed in both occipitotemporal and anterior temporal regions (Anzellotti et al., 2013). However, significant classification does not directly imply that a region carries representations of identity. The extent to which representations are invariant to transformations may vary, and a brain region could show invariance for some image transformations that humans can generalize across, but not for others. According to our definitions, such a representation would count as an invariant representation, but not as a representation of face identity.

Individuating significant classification accuracy across some specific transformations in multiple brain regions does not imply that the regions encode the same kind of representations. Therefore, occipitotemporal regions and anterior temporal regions might still encode different representations. To test this, a recent experiment investigated whether representations in different brain regions encoded information about face identity generalizing across different face halves (Anzellotti and Caramazza, 2014). For this manipulation, invariance was only found in the right anterior temporal lobe, and not in occipitotemporal cortex.

In the process of generating increasingly invariant representations, some information about identity-irrelevant differences between face tokens might be discarded or represented implicitly (DiCarlo and Cox, 2007). For this reason, the study of how and where identity-irrelevant information (e.g., information about viewpoint, illumination, and so on) is encoded can be seen as

a complementary investigation to the study of invariance. Several studies provide evidence that identity-irrelevant information declines moving from posterior to anterior regions in the ventral stream (Kietzmann et al., 2012; Anzellotti and Caramazza, 2014; see Freiwald and Tsao, 2010 for similar evidence in monkeys, and Yovel and Freiwald, 2013 for a discussion of issues of homology). However, some identity-irrelevant information might still be present in more anterior regions (DiCarlo and Maunsell, 2003; Kravitz et al., 2008).

CONCLUSION

Investigating the neural mechanisms underlying the recognition of face identity in humans is challenging, but the continuous development and improvement of design and analysis techniques has allowed the localization of representations that distinguish between face tokens depicting different people, and to begin to investigate their invariance. Given the current state of neuroimaging evidence, one region seems to encode face representations showing greatest invariance: the right anterior temporal lobe (Anzellotti et al., 2013; Anzellotti and Caramazza, 2014). This conclusion is consistent with neuropsychological evidence of deficits for face recognition after damage to the right anterior temporal lobe (Tranel et al., 1997), and with electrophysiology studies in monkeys (Freiwald and Tsao, 2010). However, it is important to note that current evidence does not establish that the right anterior temporal lobe is the only locus of face identity recognition: bilateral deficits are frequent in the anterior temporal lobes, and thus it remains possible that the left anterior temporal lobe also contributes, although to a lesser extent, to the recognition of face identity. In previous studies, the anterior temporal lobes have been implicated in semantic knowledge (Hodges et al., 1992; Tsukiura et al., 2002; Patterson et al., 2007). Invariant face representations could play an important role to link perceptual inputs to semantic knowledge about people.

Invariance does not appear only in the anterior temporal lobe, but builds up gradually, being present already to some extent in occipitotemporal regions (Kietzmann et al., 2012; Anzellotti et al., 2013; see Freiwald and Tsao, 2010 for consistent electrophysiology findings in monkeys), suggesting different roles for occipitotemporal and anterior temporal cortex for the recognition of face identity.

ACKNOWLEDGMENTS

Stefano Anzellotti was supported by a dissertation completion fellowship from Harvard University, Alfonso Caramazza's research was supported in part by the Fondazione Cassa di Risparmio di Trento e Rovereto.

REFERENCES

Andresen, D. R., Vinberg, J., and Grill-Spector, K. (2009). The representation of object viewpoint in human visual cortex. *Neuroimage* 45, 522–536. doi: 10.1016/j.neuroimage.2008.11.009

Anzellotti, S., and Caramazza, A. (2014). "Invariant representations of face identity in the ATL," *in Beyond the FFA: The role of the Ventral Anterior Temporal Lobes in Face Processing. Symposium Conducted at the Annual Meeting of the Vision Sciences Society*, eds J. Collins and I. Olson (Chairs), St. Pete Beach, FL.

Anzellotti, S., Fairhall, S. L., and Caramazza, A. (2013). Decoding representations of face identity that are tolerant to rotation. *Cereb. Cortex* doi: 10.1093/cercor/bht046 [Epub ahead of print].

Bodamer, J. (1947). Die Prosop-Agnosie. *Arch. Psychiatr. Nervenkr.* 179, 6–53. doi: 10.1007/BF00352849

Charcot, J. M. (1883). *Lectures on the Localisation of Cerebral and Spinal Diseases*, Vol. 102. London: New Sydenham Society, 178–179.

Cowen, A. S., Chun, M. M., and Kuhl, B. A. (2014). Neural portraits of perception: reconstructing face images from evoked brain activity. *Neuroimage* 94, 12–22. doi: 10.1016/j.neuroimage.2014.03.018

Çukur, T., Huth, A. G., Nishimoto, S., and Gallant, J. L. (2013). Functional subdomains within human FFA. *J. Neurosci.* 33, 16748–16766. doi: 10.1523/JNEUROSCI.1259-13.2013

Damasio, A. R., Damasio, H., and Van Hoesen, G. W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology* 32, 331–331. doi: 10.1212/WNL.32.4.331

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., and Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43, 44–58. doi: 10.1016/j.neuroimage.2008.06.037

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341. doi: 10.1016/j.tics.2007.06.010

DiCarlo, J. J., and Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J. Neurophysiol.* 89, 3264–3278. doi: 10.1152/jn.00358.2002

Ellis, H. D., and Lewis, M. B. (2001). Capgras delusion: a window on face recognition. *Trends Cogn. Sci.* 5, 149–156. doi: 10.1016/S1364-6613(00)01620-X

Etzel, J. A., Zacks, J. M., and Braver, T. S. (2013). Searchlight analysis: promise, pitfalls, and potential. *Neuroimage* 78, 261–269. doi: 10.1016/j.neuroimage.2013.03.041

Ewbank, M. P., and Andrews, T. J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *Neuroimage* 40, 1857–1870. doi: 10.1016/j.neuroimage.2008.01.049

Fairhall, S. L., Anzellotti, S., Pajtas, P. E., and Caramazza, A. (2011). Concordance between perceptual and categorical repetition effects in the ventral visual stream. *J. Neurophysiol.* 106, 398–408. doi: 10.1152/jn.01138.2010

Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322, 970–973. doi: 10.1126/science.1164318

Freiwald, W. A., and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851. doi: 10.1126/science.1194908

Freiwald, W. A., Tsao, D. Y., and Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nat. Neurosci.* 12, 1187–1196. doi: 10.1038/nn.2363

Furl, N., van Rijsbergen, N. J., Treves, A., and Dolan, R. J. (2007). Face adaptation aftereffects reveal anterior medial temporal cortex role in high level category representation. *Neuroimage* 37, 300–310. doi: 10.1016/j.neuroimage.2007.04.057

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.* 12, 495–504. doi: 10.1162/089892900562165

Gitelman, D. R., Ashburner, J., Friston, K. J., Tyler, L. K., and Price, C. J. (2001). Voxel-based morphometry of herpes simplex encephalitis. *Neuroimage* 13, 623–631. doi: 10.1006/nimg.2000.0734

Goesaert, E., and Op de Beeck, H. P. (2013). Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *J. Neurosci.* 33, 8549–8558. doi: 10.1523/JNEUROSCI.1829-12.2013

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203. doi: 10.1016/S0896-6273(00)80832-6

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. doi: 10.1126/science.1063736

Haynes, J. D., and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534. doi: 10.1038/nrn1931

Hecaen, H., and Angelergues, R. (1962). Agnosia for faces (prosopagnosia). *Arch. Neurol.* 7, 92–100. doi: 10.1001/archneur.1962.04210020014002

Heidenhain, D. A. (1927). Beitrag zur kenntnis der seelenblindheit. *Eur. Neurol.* 66, 61–74. doi: 10.1159/000165960

Hodges, J. R., and Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurol.* 6, 1004–1014. doi: 10.1016/S1474-4422(07)70266-1

Hodges, J. R., Patterson, K., Oxbury, S., and Funnell, E. (1992). Semantic dementia progressive fluent aphasia with temporal lobe atrophy. *Brain* 115, 1783–1806. doi: 10.1093/brain/115.6.1783

Jossmann, P. (1929). Zur Psychopathologie der optisch-agnostischen Storungen, *Mschr. Psychiatr. Neurol.* 77, 81–149. doi: 10.1159/000164409

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.

Kapur, N., Barker, S., Burrows, E. H., Ellison, D., Brice, J., Illis, L. S., et al. (1994). Herpes simplex encephalitis: long term magnetic resonance imaging and neuropsychological profile. *J. Neurol. Neurosurg. Psychiatry* 57, 1334–1342. doi: 10.1136/jnnp.57.11.1334

Kietzmann, T. C., Swisher, J. D., König, P., and Tong, F. (2012). Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. *J. Neurosci.* 32, 11763–11772. doi: 10.1523/JNEUROSCI.0126-12.2012

Kravitz, D. J., Vinson, L. D., and Baker, C. I. (2008). How position dependent is visual object recognition? *Trends Cogn. Sci.* 12, 114–122. doi: 10.1016/j.tics.2007.12.006

Kriegeskorte, N., and Bandettini, P. (2007). Combining the tools: activation- and information-based fMRI analysis. *Neuroimage* 38, 666–668. doi: 10.1016/j.neuroimage.2007.06.030

Kriegeskorte, N., Formisano, E., Sorger, B., and Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20600–20605. doi: 10.1073/pnas.0705654104

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3863–3868. doi: 10.1073/pnas.0600244103

Malone, D. R., Morris, H. H., Kay, M. C., and Levin, H. S. (1982). Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *J. Neurol. Neurosurg. Psychiatry* 45, 820–822. doi: 10.1136/jnnp.45.9.820

Meadows, J. C. (1974). Disturbed perception of colours associated with localized cerebral lesions. *Brain* 97, 615–632. doi: 10.1093/brain/97.1.615

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., and Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann. Neurol.* 47, 36–45. doi: 10.1002/1531-8249(200001)47:1<36::AID-ANA8>3.0.CO;2-L

Mur, M., Ruff, D. A., Bodurka, J., Bandettini, P. A., and Kriegeskorte, N. (2010). Face-identity change activation outside the face system: "release from adaptation" may not always indicate neuronal selectivity. *Cereb. Cortex* 20, 2027–2042. doi: 10.1093/cercor/bhp272

Natu, V. S., Jiang, F., Narvekar, A., Keshvari, S., Blanz, V., and O'Toole, A. J. (2010). Dissociable neural patterns of facial identity across changes in viewpoint. *J. Cogn. Neurosci.* 22, 1570–1582. doi: 10.1162/jocn.2009.21312

Nestor, A., Plaut, D. C., and Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9998–10003. doi: 10.1073/pnas.1102433108

Nestor, A., Plaut, D. C., and Behrmann, M. (2014). "Facial identity: an investigation of neural encoding and image reconstruction," in *Face Perception: Neural Mechanisms. Talk Session Conducted at the Annual Meeting of the Vision Sciences Society*, St. Pete Beach, FL.

Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., et al. (2007). Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain* 130, 1138–1147. doi: 10.1093/brain/awl344

Oliva, A., and Torralba, A. (2007). The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527. doi: 10.1016/j.tics.2007.09.009

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.* 8, 976–987. doi: 10.1038/nrn2277

Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., and Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: time course and topographic evoked – potentials mapping. *Hum. Brain Mapp.* 26, 65–79. doi: 10.1002/hbm.20130

Puce, A., Allison, T., Asgari, M., Gore, J. C., and McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16, 5205–5215.

Rajimehr, R., Young, J. C., and Tootell, R. B. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1995–2000. doi: 10.1073/pnas.0807304106

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., and Dolan, R. J. (2004). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat. Neurosci.* 8, 107–113. doi: 10.1038/nn1370

Schendan, H. E., and Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *J. Cogn. Neurosci.* 15, 111–135. doi: 10.1162/089892903321107864

Sergent, J., Ohta, S., and Macdonald, B. (1992). Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* 115, 15–36. doi: 10.1093/brain/115.1.15

Snowden, J. S., Thompson, J. C., and Neary, D. (2004). Knowledge of famous faces and names in semantic dementia. *Brain* 127, 860–872. doi: 10.1093/brain/awh099

Thompson, S. A., Patterson, K., and Hodges, J. R. (2003). Left/right asymmetry of atrophy in semantic dementia: behavioral–cognitive implications. *Neurology* 61, 1196–1203. doi: 10.1212/01.WNL.0000091868.28557.B8

Tranel, D., Damasio, H., and Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* 35, 1319–1327. doi: 10.1016/S0028-3932(97)00085-7

Tsukiura, T., Fujii, T., Fukatsu, R., Otsuki, T., Okuda, J., Umetsu, A., et al. (2002). Neural basis of the retrieval of people's names: evidence from brain-damaged patients and fMRI. *J. Cogn. Neurosci.* 14, 922–937. doi: 10.1162/089892902760191144

Verosky, S. C., Todorov, A., and Turk-Browne, N. B. (2013). Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia* 51, 2100–2108. doi: 10.1016/j.neuropsychologia.2013.07.006

Vuilleumier, P., Henson, R. N., Driver, J., and Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat. Neurosci.* 5, 491–499. doi: 10.1038/nn839

Warrington, E. K., and McCarthy, R. A. (1988). The fractionation of retrograde amnesia. *Brain Cogn.* 7, 184–200. doi: 10.1016/0278-2626(88)90029-2

Warrington, E. K., and Shallice, T. (1984). Category specific semantic impairments. *Brain* 107, 829–853. doi: 10.1093/brain/107.3.829

Whiteley, A. M., and Warrington, E. K. (1977). Prosopagnosia: a clinical, psychological, and anatomical study of three patients. *J. Neurol. Neurosurg. Psychiatry* 40, 395–403. doi: 10.1136/jnnp.40.4.395

Wilbrand, H. (1892). Ein Fall von Seelenblindheit und Hemianopsie mit Sectionsbefund. *Dtsch. Z. Nervenheilkd.* 2, 361–387. doi: 10.1007/BF01667704

Xiang, J. Z., and Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* 37, 657–676. doi: 10.1016/S0028-3908(98)00030-6

Yovel, G., and Freiwald, W. A. (2013). Face recognition systems in monkey and human: are they the same thing? *F1000prime Rep.* 5, 10. doi: 10.12703/P5-10

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 February 2014; accepted: 10 June 2014; published online: 26 June 2014.
Citation: Anzellotti S and Caramazza A (2014) The neural mechanisms for the recognition of face identity in humans. Front. Psychol. ***5****:672. doi: 10.3389/fpsyg.2014.00672*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Anzellotti and Caramazza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 02 July 2014
doi: 10.3389/fpsyg.2014.00695

The transition in the ventral stream from feature to real-world entity representations

Guy A. Orban[1]*, Qi Zhu[2] and Wim Vanduffel[2]

[1] *Department of Neuroscience, University of Parma, Parma, Italy*
[2] *Laboratorium voor Neuro-en Psychofysiologie, Department of Neuroscience, KU Leuven, Leuven, Belgium*

Edited by:
Chris Fields, New Mexico State University, USA (Retired)

Reviewed by:
Natasha Sigala, University of Sussex, UK
Shin'Ya Nishida, NTT Communication Science Laboratories, Japan

****Correspondence:***
Guy A. Orban, Department of Neuroscience, University of Parma, Via Volturno 39, 43100 Parma, Italy
e-mail: guy.orban@med.kuleuven.be

We propose that the ventral visual pathway of human and non-human primates is organized into three levels: (1) ventral retinotopic cortex including what is known as TEO in the monkey but corresponds to V4A and PITd/v, and the phPIT cluster in humans, (2) area TE in the monkey and its homolog LOC and neighboring fusiform regions, and more speculatively, (3) TGv in the monkey and its possible human equivalent, the temporal pole. We attribute to these levels the visual representations of features, partial real-world entities (RWEs), and known, complete RWEs, respectively. Furthermore, we propose that the middle level, TE and its homolog, is organized into three parallel substreams, lower bank STS, dorsal convexity of TE, and ventral convexity of TE, as are their corresponding human regions. These presumably process shape in depth, 2D shape and material properties, respectively, to construct RWE representations.

Keywords: 3D shape, retinotopy, actions, 2D shape, material properties

INTRODUCTION

This brief thought-provoking perspective paper complements the review devoted to the extrastriate neuronal properties published in Physiological reviews (Orban, 2008). At that time (Orban, 2008; Nassi and Callaway, 2009) the properties of infero-temporal neurons were not well understood, preventing a coherent picture of the function of monkey TE and its equivalent regions in man to be drawn. The present perspective paper attempts to correct this shortcoming. Since fMRI became available (Dubowitz et al., 1998; Logothetis et al., 1998; Stefanacci et al., 1998; Vanduffel et al., 1998) for systematic investigation in the alert monkey (Vanduffel et al., 2001), considerable progress has been made, through fMRI-guided monkey single-cell studies, and by parallel comparative imaging in humans and monkeys. In addition, the connections of TE cortex have recently been reassessed (Saleem et al., 2007, 2008; Ungerleider et al., 2008; Gerbella et al., 2010; Kravitz et al., 2013), allowing a tight comparison between anatomical connectivity and functionality.

RETINOTOPIC ORGANIZATION OF THE VISUAL SYSTEM

Our understanding of the retinotopic organization of the human visual system is largely due to fMRI. It is now established that human occipital cortex and neighboring parts of temporal and parietal cortex includes 15–17 distinct representations of the visual field. In addition to the three early visual areas V1-3, there is agreement (Wandell et al., 2007; Arcaro et al., 2009; Kolster et al., 2010) concerning hV4, LO1-2, the four areas of the MT cluster (MT, pMSTv, pFST, and pV4t), phPITd and phPITv (**Figure 1A**), and V6 (Pitzalis et al., 2006). There still is debate concerning the V3A complex which is subdivided into either two (V3A/B; Larsson and Heeger, 2006) or four areas (V3A/B/C/D; Georgieva et al., 2009). Dorsally, the V3A complex is bordered by V7 (Tootell et al., 1997), which is in fact the first parietal area, also designated IPS0 (Silver et al., 2005). Recently, V7 was reported to be part of a cluster of two areas, V7 (IPS0) and V7A

Abbreviations: *Cortical areas and regions*: AIP, anterior intraparietal area; CIP, caudal intraparietal area; DP, dorsal parietal area, located dorsal from V4; FST, fundus of superior temporal area, third element of the MT cluster; pFST, human homolog of FST; IPS0-5 is a set of successive retinotopic areas near the intraparietal sulcus (IPS) defined solely by reversal of polar angle (visual field is defined by both polar angle and eccentricity); IPS0-1 corresponds to V7/V7A; IT, infero-temporal cortex, includes three cytoarchitectonic fields, TEO, TE, and TGv; the first two have also been parceled into three antero-posterior subdivisions, posterior IT (PIT), central IT (CIT) and anterior IT (AIT), with PIT largely corresponding to TEO and CIT and AIT to TE; It includes the lower bank of the superior temporal sulcus (STS); LO1, LO2 lateral occipital area 1 and 2; LOC, lateral occipital cortex defined by the contrast *intact* vs *scrambled images* of objects. Includes LO1-2 but extends rostrally into occipito-temporal sulcus and fusiform cortex; LST, lateral superior temporal area, a motion area located in the monkey STS in front of FST; MSTv, medial superior temporal area ventral part, second component of the MT cluster; pMSTv human homologue of MSTv; MT, middle temporal area; first element of the MT cluster; OTd, occipito-temporal dorsal area; PFG, cytoarchitectonic field in IPL (others are PF, PG, and opt); PITd, posterior infero-temporal dorsal area; phPITd, putative human homologue of PITd; PITv, posterior infero-temporal ventral area; phPITv, putative human homologue of PITv; PPC, posterior parietal cortex (part of parietal cortex behind primary somato-sensory cortex); STPm, superior temporal posterior middle area, a motion area located in the upper bank of monkey STS (middle level); TF, TH cytoarchtectonic regions of parahippocampal cortex; TFO, cytoarchitectonic area posterior to TF/TH and medial to TEO; has been labeled previously VTF (visual part of TF) by Boussaoud et al. (1991), but is now recognized as a separate cytoarchitectonic entity (Kravitz et al., 2013); V1, V2-V7, visual area 1, 2, to 7. The designation "V7" has been used only in humans; V5 corresponds to MT; While homology for V1-3 and V5/MT and V6 is relatively well established, hV4 refers a human area in positioned similarly to monkey V4 but having a different retinotopic organization; V3A, V4A, ad V7A, areas in neighborhood of V3, V4, and V7; V4t, fourth area of the MT cluster, initially considered incomplete now, accepted as corresponding to a complete hemifield; pV4t, human homologue of V4t; VO1, VO2, ventral occipital area 1 and 2; *Anatomical structures*: IPS intraparietal sulcus separating the superior parietal lobule (SPL) from the inferior parietal lobule (IPL); MTG, middle temporal gyrus; OTS occipito-temporal sulcus; STS, superior temporal sulcus; STG, superior temporal gyrus; TPJ, temporo-parietal junction; *Other abbreviations:* AL, anterior lateral (face patch); BM, biological motion; ML, middle lateral (face patch); RWE, real world entity.

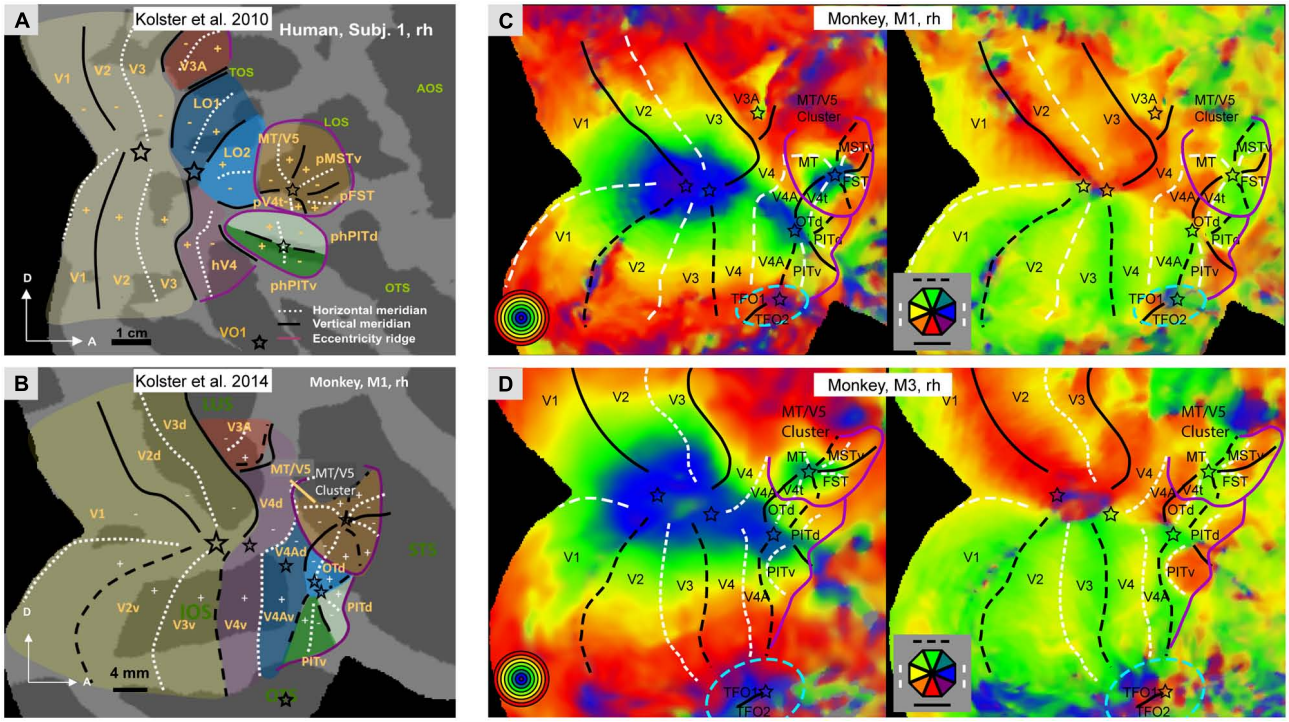


FIGURE 1 | (A,B) Schematic representation of the retinotopic organization of occipital cortex: in humans (**A**, subject 1, rh) and in monkeys (**B**, monkey M1, rh); Modified from Kolster et al. (2014).**C,D**: Polar angle and eccentricity maps for monkeys M1 **(C)** and M3 **(D)**, same data as Janssens et al. (2014) but lower threshold. Black lines: vertical meridians (full: upper, dashed: lower), white dashed lines: horizontal meridians, stars: central visual field representation; purple lines: eccentricity ridges; In **A,B**: LuS: lunate sulcus, STS: superior temporal sulcus; OTS occipito-temporal sulcus; TOS: transverse occipital sulcus, LOS: lateral occipital sulcus, AOS: anterior occipital sulcus, OTS occipito-temporal sulcus; Other nomenclature: see Abbreviations. In **C,D** blue stippled elliptic outlines mark additional retinotopic regions (TFO1/2) ventral to V4A/PITv.

(IPS1), sharing a central representation (Georgieva et al., 2009), a finding confirmed by using stereoscopically- instead of luminance-defined phase-encoded retinotopy stimuli (Kolster et al., 2011). This test also suggested that at more rostral levels the posterior parietal cortex (PPC) is retinotopically organized into 3–6 additional areas. Their complete characterization requires further work, since investigations thus far have relied mainly on polar angle analyses to define IPS2-5 (Silver and Kastner, 2009). On the other, ventral side of the occipital cortex Kolster et al. (2010) have described a single VO1 area (**Figure 1A**), although these data are also compatible with the presence of a second VO2 area, as described by Brewer et al. (2005). Finally, Arcaro et al. (2009) have shown that VO1-2 borders two additional retinotopic areas, PH1 and PH2, extending into the parahippocampal cortex. Thus in humans, a major difference exists between the dorsal and ventral visual pathways with respect to their retinotopic representation. The dorsal pathway retains a retinotopic organization, while the ventral pathway discards this organization beyond the phPIT cluster. It needs to be noted, however, that the most ventrally located occipito-temporal cortex processing scene information remains retinotopically organized. It has been suggested that at higher levels of the ventral pathway, eccentricity remains an important principle of organization (Levy et al., 2001), but this largely reflects the representation of large eccentricities in scene-processing regions.

The situation is very similar in the macaque. Its occipital cortex and neighboring parts of temporal and parietal cortex includes 14 retinotopic maps (**Figure 1B**): the three early areas V1-3, V4, and its two satellites (V4A and OTd), the two PITs (Janssens et al., 2014; Kolster et al., 2014), V3A, the four areas of the MT cluster (Kolster et al., 2009), and V6 in the parieto-occipital sulcus (Galletti et al., 1999). Cytoarchitectonic area TEO, which initially was proposed to contain a single retinotopic map (Boussaoud et al., 1991), in fact includes four different retinotopic maps: V4A, OTd, PITd, and PITv (Janssens et al., 2014; Kolster et al., 2014). It may be that neighboring cytoarchitectonic area TFO will undergo the same fate. Indeed, ventrally in occipital cortex, in front of the most peripheral part of V4 and below V4A, there is preliminary evidence (Janssens et al., 2014; Kolster et al., 2014) for another central representation, defining a cluster including two areas joined by that central representation. These areas have been tentatively labeled TFO1 and TFO2 (**Figures 1C,D**). The location in the dorsal bank of OTS and internal organization of this cluster suggest they may correspond to VO1-2 of humans. In humans VO1/2 are sensitive to color (Brewer et al., 2005) and color responses have been reported in a monkey PET study in a region that likely corresponds to TFO (Takechi et al., 1997). We propose that TFO1/2 are the starting point of the scene-processing pathway, consistent with recent fMRI activation and single cell recordings (Kornblith et al., 2013, but see Nasr et al., 2011). As

in humans this pathway emphasizes the peripheral visual field (Kravitz et al., 2013). A number of parietal regions are retinotopically organized. Arcaro et al. (2011) described, in addition to DP, a pair of areas, CIP1 and CIP2, in the caudal part of the lateral bank of the IPS. In keeping with their location caudal to an extensive representation of peripheral visual field, CIP1/2 might be the monkey counterparts of the V7/V7A pair (Durand et al., 2009). This implies that human areas V3B-D have no counterpart in the monkey and are evolutionary novel areas. This is consistent with the caudal elongation of the IPS which in humans includes an occipital portion needed to bridge the enlargement of IPL (Grefkes and Fink, 2005). Further forward in monkey IPS, Arcaro et al. (2011) described a single hemifield representation, LIP, of which the central representation had been described by Fize et al. (2003).

In summary, the retinotopic organization of occipital cortex is remarkably similar in human and non-human primates, more than initially appreciated (Wandell et al., 2007). In addition, the organization beyond occipital cortex is also rather similar. The dorsal visual pathway of both humans and monkeys maintains a retinotopic organization, while the ventral pathway abandons this organization beyond TEO/the PIT monkey areas and their human homologs (phPITs). In both species the rostral limit of retinotopic cortex represents the peripheral visual field (purple lines in **Figure 1**). The most ventral, scene-processing pathway transiting through the parahippocampal cortex retains this organization at least in humans and possibly in monkeys (this ventral cortex is difficult to image in the monkey given the susceptibility artifacts, see Ku et al., 2011). Insofar as scene processing might be considered the qualitative counterpart of the metric processing of space in the dorsal pathway, the underlying principle may be that areas processing space, either quantitatively or qualitatively retain a crude retinotopic organization. In the monkey, the temporal cortex beyond TEO/the PITs includes mainly areas TE and TGv near the temporal pole (**Figure 2A**). In humans, LOC, which primarily corresponds to TE (Denys et al., 2004; Sawamura et al., 2005) is located several cm away from the temporal pole, suggesting that the TGv region has greatly expanded in humans. This raises the question by which functional organization principle, if any, the retinotopic organization has been replaced in these regions of temporal cortex.

PITd PROCESSES 3D SHAPE FROM SHADING, ONE OF THE BUILDING BLOCKS OF SHAPE REPRESENTATION FOR REAL-WORLD ENTITIES

In monkeys, the fMRI study of Nelissen et al. (2009) indicates that the dorsal PIT is involved in processing 3D shape from shading. The fMRI activation of PITd corresponds to stronger neuronal responses for shading patterns reflecting 3D structure (Köteles et al., 2008). In humans, 3D shape from shading is similarly processed in a restricted occipito-temporal region (Georgieva et al., 2008). Matching the local maximum of this activation to a maximum-probability map of occipital retinotopic areas (Abdollahi et al., 2013) suggests that it is located near or in phPITd. In an effort to dissociate 3D shape from shading from simple flat luminance patterns, both Nelissen et al. (2009) and Georgieva et al. (2008) required joined activation in several specific contrasts for a region to be considered processing 3D shape from shading. Sereno et al. (2002) also reported 3D shape from shading responses in a somewhat broader region near PITd, including MT and FST in which several 3D shape cues, motion, shading, and texture converged. The importance of these observations derives from the fact that the image of any real-world object is necessarily (because of optics) characterized by two complementary components: a boundary that defines its 2D shape and a luminance pattern inside this boundary that defines its relief (shape in depth or 3D shape). These two complementary components depend in complex ways on the material properties and shape of the objects, as well as the direct and indirect light sources present in the scene. Nevertheless, 2D shape and 3D shape from shading combine to unambiguously define a visual representation of a real-world entity (RWE), whether an object, a plant, an animal, or a conspecific. RWE is preferred to the term *object* which is ambiguous, as the above listing shows. It is well established that boundary information is processed in V4 (Pasupathy and Connor, 2001) and is further elaborated in what is commonly called TEO (Brincat and Connor, 2004). Thus the most rostral retinotopic regions of the ventral pathway (**Figure 1B**), parts of cytoarchitectonic TEO, contain the elements required to generate visual representations of RWE. We propose that the primary function of TE, located beyond the retinotopic cortex, is to house the visual representations of RWEs, built by combining lower-level inputs from retinotopic cortex. The visual representation of RWEs can also be triggered by their images (Tanaka et al., 1991), and by even more simplified stimuli such as drawings (Denys et al., 2004).

The visual representations of RWEs are supposedly assembled in TE by combining inputs representing a boundary (or external contour) as well as elements of the luminance distribution inside that boundary. These internal elements can be either contours corresponding to extremes in the luminance distribution, or regions of constant or smoothly varying luminance. Indeed, this combinatorial view is supported by recent recordings in the ML face patch of the monkey, located just at the edge of retinotopic cortex. Almost all neurons in this patch are face selective (Tsao et al., 2006) and this selectivity arises from combining the geometry of the boundary with that of key internal features such as the eyes, nose, or mouth (Freiwald et al., 2009), but also includes the contrast levels in certain positions with respect to these features (Ohayon et al., 2012). However, this combination of 2D shape and 3D from shading does not exhaust the possible visual representations of RWEs, since the nature of RWEs is specified by not only their shape but also their material properties. Hence the representation of RWEs is build up from three main sources: features related to the 2D shape of the boundary in the image, and to the 3D shape, and material properties of the region enclosed by the boundary.

REPRESENTATIONS OF REAL-WORLD ENTITIES IN TE

Recent anatomical data suggest that three parallel substreams operate within TE (**Figure 2A**), located in the lower bank of STS and in the dorsal and ventral parts of TE. We suggest that these three streams preferentially use features of 3D shape, 2D shape, and material properties, respectively, to build up RWE

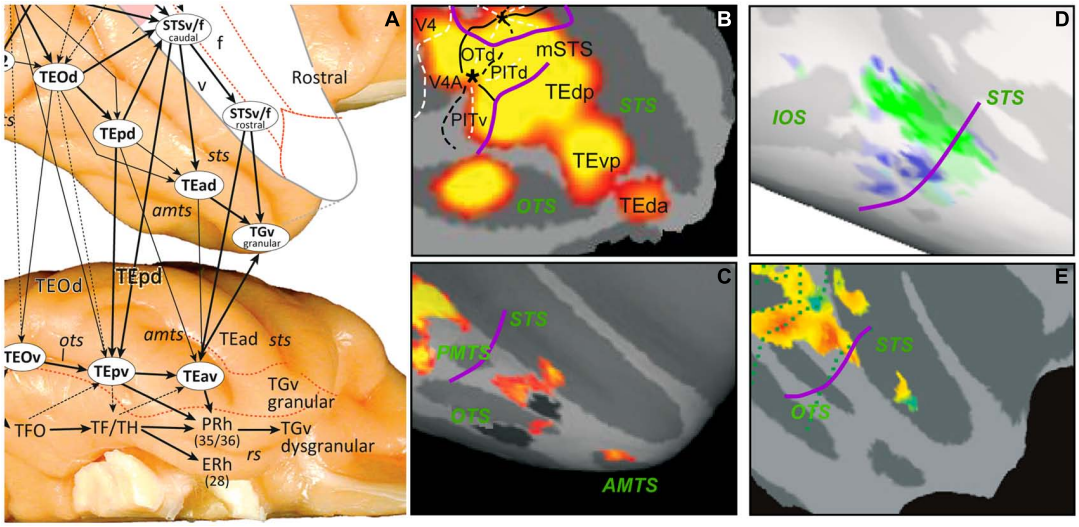


FIGURE 2 | (A) The anatomical organization of monkey TE into three parallel substreams (from Kravitz et al., 2013); **(B–E)** SPMs showing activation sites in right IT for 2D shape, color, shape vs. no shape, and gloss. These were defined by the following subtractions: *intact* vs. *scrambled images* of objects **(B)**, *color* vs. *no color* mondrians **(C)**, *inact* vs. *scrambled images* of objects **(D)** main effect of *gloss*, independent of contrast **(E)**. In D the non-shape, selective voxels were strongly selective for material property, whereas shape-selective ones were not. Purple curved lines in **B–E**: approximate caudal boundary of TE. From Denys et al. (2004; **B**), Harada et al. (2009; **C**), Goda et al. (2014; **D**), and Okazawa et al. (2012; **E**).

representations (**Figure 3**). This implies that functional segregation between these substreams is maximal at the transition between the retinotopic, feature level and the middle level (i.e., the TEO/TEp border in **Figures 2A,D**) and gradually blurs toward the rostral end of TE. Indeed, the three aspects defining RWEs (3D shape, 2D shape, and material properties) contribute in different proportions to the definition of given RWEs, and some cues belonging to one of the aspects may remain represented at more rostral levels, as for example color, one of the material cues (see below). According to this scheme the middle substream carries mainly 2D shape information, as evidenced by the subtraction

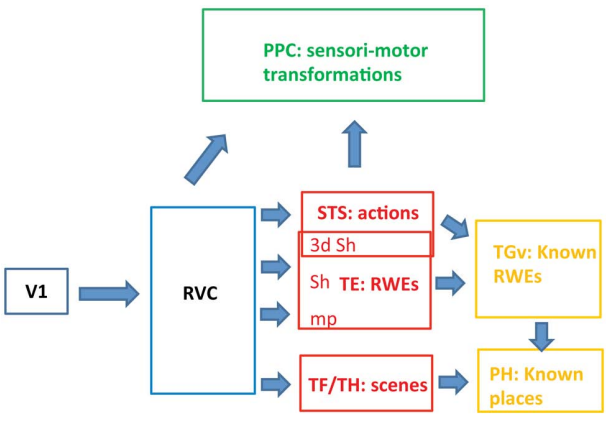


FIGURE 3 | Schematic view of the functional organization of the ventral pathway in the three levels (blue, red, and yellow). RVC: retinotopic visual cortex includes the PITs, i.e., the posterior part of the IT complex; RWE: real world entity; sh: shape, mp: material properties, PH: parahippocampal cortex.

intact minus *scrambled images* of objects, which mainly activates dorsal TE (**Figure 2B**; Denys et al., 2004; Sawamura et al., 2005; Lafer-Sousa and Conway, 2013). A long list of single-cell studies have been devoted to 2D shape selectivity in IT cortex (Logothetis and Sheinberg, 1996; Tanaka, 1996; Orban, 2008 for review), with some stressing the affine nature of the representation (Kayaert et al., 2005). This 2D shape substream also contains several face patches, such as the ML, and AL patches (Moeller et al., 2008).

The ventral TE substream may process material properties (for review see Fleming, 2014) which also contribute to the definition of RWEs (e.g., a tomato is red and smooth). This is supported by the color activation sites in ventral TE (**Figure 2C**; Harada et al., 2009; Lafer-Sousa and Conway, 2013). The other principal material property cue is texture (texture is also a cue for 3D shape; see Sereno et al., 2002; Orban, 2011). Little is known about texture processing in monkeys (see Köteles et al., 2008), but in humans ventral occipito-temporal cortex is heavily involved in texture processing (Peuskens et al., 2004; Cant and Goodale, 2007). Ku et al. (2011) have reported face patches in and around the ventral temporal cortex of the monkey: in ventral TE, area TF, entorhinal cortex, hippocampus, and region labeled ventral V4, which might have included TFO. Since the hairy monkey face and control stimuli (fruits, houses, and fractals) differed in texture, some of these activation sites (in particular the posterior ones) might actually reflect the texture differences rather than the presence of the face. Regions in PIT processing material properties have been investigated recently by Goda et al. (2014), who showed a clear segregation between shape and material properties at the level of PIT (**Figure 2D**), in agreement with our proposal. We propose that the third substream in the lower bank of STS processes 3D shape (Sereno et al., 2002; Yamane et al., 2008). This

proposal is consistent with the presence in the lower bank of a small patch concerned with gloss (**Figure 2E**; Okazawa et al., 2012), a marker of 3D convexity for certain materials, and TEs, a region extracting curvature from disparity (Janssen et al., 2000). This substream overlaps with action-processing regions located in both banks of the STS, especially their deeper regions (Nelissen et al., 2011). One of the main cues for extracting actions is the deformation of body shape (Vangeneugden et al., 2009; Singer and Sheinberg, 2010), explaining the proximity of shape, and action processing areas. Similarly, material properties contribute heavily to scene processing, which may explain their location in ventral TE, as it neighbors the scene-processing stream in parahippocampal TF/TH.

Both the general anatomy, that indicates serial processing (**Figure 2A**), and studies specific to the face-processing system suggest that the representation of RWEs might be further elaborated rostrally within TE. A detailed study of the face patches (Freiwald and Tsao, 2010) suggests that the first step is the extraction of the face category in ML; that additional properties, such as the viewpoint from which the face is seen, are represented in subsequent patches; and that finally at the highest level, exemplars, individual faces, are represented, implying that sufficient invariance has been achieved. Similarly Lafer-Sousa and Conway (2013) have suggested that the representation of color is more elaborated in anterior than in posterior TE. Koida and Komatsu (2007) demonstrated the task dependent activity of TE color selective neurons. Task dependent processing and other aspects of TE processing such as extending the neural representation beyond the stimulus presentation (Kovacs et al., 1995) or buffering the last representation (Orban and Vogels, 1998) are beyond the scope of the present perspective paper.

Despite this elaboration of RWE representations, including becoming gradually more invariant (DiCarlo et al., 2012), the representation in TE remains incomplete in the sense that the entire RWE is generally not represented (a few neurons may do so, as suggested for target-paired association neurons; Hirabayashi et al., 2013). Even in the anterior face patches, only the face is represented, not the whole person; also, patches related to color represent only one material aspect of the RWE. The partial representation of the RWE at the middle level can be considered a generalization of the selectivity of TE neurons for 2D shape components (Tanaka et al., 1991). The RFs of TE neurons are relatively large (about 15° diameter), located primarily in the contralateral visual field, and generally included the fovea (Op De Beeck and Vogels, 2000). Hence a certain spatial coding remains possible, in particular that of the relative positions of shape or RWE parts. Several rationales can be advanced for the incomplete representation of RWEs in TE having to do with more flexible representations. In particular, some material properties define the exemplar but not the category (e.g., John may have black hair but not all men have black hair), accommodation of slow changes in properties, e.g., due to aging, or seasons (color changes of the leaves), and finally detection of uncommon associations of shape and color (see Zeki and Marini, 1998; e.g., John generally looks healthy, but can be very pale because of illness).

Thus far, views about the organization of TE have been dominated by the presence of patches in TE, among which face and body patches (Tsao et al., 2003; Pinsk et al., 2009; Bell et al., 2011; Popivanov et al., 2012) are the best known. Initially it was assumed that the non-face and non-body objects were processed outside these patches (Ishai et al., 1999; Tsao et al., 2003), implying that RWE of different types were processed in different compartments of TE. This view, however, is inconsistent with recent evidence for patches for color, 3D shape from disparity, or gloss (Harada et al., 2009; Joly et al., 2009; Okazawa et al., 2012). A recent study by Srihasam et al. (2012) sheds new light on the exact organization of TE. These authors showed that when monkeys are trained to use numerical or letter symbols from a young age, these stimuli are represented in patches within TE, but are not present in untrained monkeys or those trained to use these symbols as adults (and not learning the task as well). While others (Vogels and Orban, 1994; Kobatake et al., 1998; Sigala and Logothetis, 2002) have reported plasticity at the single-cell level after training, the Srihasam study was the first to report functional architectural changes in TE, rather than just changes in neuronal properties. Srihasam et al. (2012) suggest that patches arise because neurons with similar selectivity tend to group together to increase computational efficiency (shorter connections). In retinotopic cortex, these groupings are constrained by the retinotopic organization, but in TE this is not the case, thus giving rise to varying degrees of aggregation, probably depending on the behavioral relevance of the selectivity. Those aspects or components of RWEs with strong behavioral relevance are grouped into complex systems of multiple connected patches, of which the face patches are probably the most elaborated. Those with limited relevance, such as properties/parts of objects encountered only infrequently, have small representations in columnar-like structures (Tanaka et al., 1991). Those with intermediate relevance have a somewhat broader representation, in one or two patches, such as color or 3D shape. Thus the processing of RWEs of different type or nature is interwoven, their properties being represented more or less extensively depending on behavioral relevance. Such size differences of functional TE modules are consistent with the findings of Sato et al. (2013), with our largest and smallest modules corresponding to their domains and columns, respectively. In humans these domains may include the word form areas (Cohen et al., 2000) analyzing strings of symbols during reading, even if words are not actually RWEs.

REPRESENTATIONS OF ACTIONS IN STS

Several lines of investigation suggest that actions (purposeful movements of an agent: animal, human, or even robot) are processed in the middle and rostral STS largely in parallel with RWEs in TE (**Figure 3**). Recent evidence suggests that actions are extracted in LST and STPm, two motion-sensitive regions just anterior to the MT cluster. In these regions the configuration and kinematic cues of BM interact (Jastorff et al., 2012), which is the definition of action. Indeed, action-selective neurons have been recorded at this level, and both cues appear operative: deforming shape in the lower bank, and motion patterns in the upper bank (Vangeneugden et al., 2009). We have begun to understand the homology of monkey STS (Orban and Jastorff, 2014): The lower bank corresponds to posterior OTS and fusiform cortex in humans, overlapping with LOC (in which actions and shape

overlap, as in the lower bank of STS; Jastorff and Orban, 2009), while the upper bank of monkey STS corresponds to posterior MTG and posterior STS in humans (Jastorff and Orban, 2009; Jastorff et al., 2012).

We have recently shown that the action-sensitive regions of STS devoted to grasping project to the ventral premotor cortex (F5), where mirror neurons occur, via two way stations in the PPC: AIP and PFG (Nelissen et al., 2011). We believe that this is a general strategy within the primate visual system, not merely for grasping and manipulative actions, but for all types of action. The STS action-processing regions project to the PPC in order to extract action category which requires that a large number of invariances to be solved: not only for size, position, and in plane orientation, as for RWEs, but also for viewpoint and posture. The available evidence (Freiwald and Tsao, 2010) suggests that TE and neighboring regions achieve invariance only at the expense of large neuronal pools and that therefore the many invariances required for understanding body actions involve too much neuronal hardware to be realistically achieved in the STS. Hence, we propose that the STS regions send the visual information about *which action* is observed to the PPC housing the schema of specific actions, i.e., the sensori-motor transformation underlying various actions. By projecting these visual signals onto the corresponding motor plan, invariance is automatically achieved and categorization becomes feasible. This invariance problem is less stringent for facial expressions, as the viewpoints, and postural invariance requirements are much more limited. Hence what applies to body action may not necessarily apply to facial expressions, explaining the presence of face patches in the upper bank of STS, where dynamic face expressions are processed (Polosecki et al., 2013).

These action signals sent to the PPC concern the nature/goal of the action defining *which action* is observed. However, actions are also further processed in the STS itself, analysis probably related to *how the action* is performed, e.g., slowly or quickly, with difficulty or easily, physiologically or pathologically. The latter sort of processing provides information about the state of the actor, even if the actor itself, an RWE, is processed in TE. The state of the agent reflects his/her emotions, but also the physiological state, and perhaps also vitality (Di Cesare et al., 2013). The latter aspect is related to the rank of the actor in the group or the social organization in general and may be dealt with in human TPJ, a region which may have arisen from some middle part of the STS (Sallet et al., 2011; Mars et al., 2013). TPJ is often considered the starting point (Saxe et al., 2004) for processing other agents (theory of mind), but recent studies (Jastorff and Orban, 2009) alternatively suggest that there might be a representation of an agent in the scene in posterior STG. Activity in posterior STS and TPJ would then specify properties of the agent, such as rational or efficient behavior (Jastorff et al., 2011).

THREE LEVELS OF PROCESSING IN THE VENTRAL STREAM (FIGURE 3)

TE corresponds to the middle level of the ventral stream in the monkey. It builds a partial representation of RWEs and operates in parallel with STS, processing actions and TF/TH processing scenes (**Figure 3**). TE receives input from retinotopic cortex (first level) where image features are processed to generate higher-order features related to 3D shape, 2D shape, or material properties in specific parts of the visual field. The retinotopic visual cortex not only processes a range of elementary image features (Zeki, 1978) but also resolves image segmentation by establishing topological relationships between the features: inside vs. outside and in front vs. behind (Zhang and von der Heydt, 2010). The anatomy indicates, however, that the ventral pathway in monkeys may include, in addition to the retinotopic cortex and TE, a third level beyond TE. A small temporal region, TGv, receiving input from the three substreams in TE, is situated in front of TE near the temporal pole (Kravitz et al., 2013). The TGv region projects to rhinal cortex in which memory of the association between two images is constructed by the convergence of their representations in TE (Naya et al., 2003a; Hirabayashi et al., 2013). We propose that the TGv region, which is greatly expanded in humans and is referred to as the temporal pole, builds on the partial representations of individual RWEs achieved at the rostral TE (Freiwald and Tsao, 2010) to generate representations of *known* RWEs (Damasio et al., 2004; Quiroga et al., 2005). The association of the elements present in TE detected in rhinal cortex (Hirabayashi et al., 2013), may be back-projected (Naya et al., 2003b; Takeuchi et al., 2011) onto the most rostral visual part of temporal cortex, giving rise to representations of known RWEs (Takeda et al., 2005). For example, exemplars of a shape category, e.g., face plus body, and particular material properties define a given individual and this association gives rise to the representation of that known individual in TGv, perhaps supplemented by information about *how* he acts and the scenes in which he appears. In contrast to the TE level, the representation here is that of the complete RWE, e.g., a conspecific, and no longer simply a face. A similar operation may be applied to scene information in parahippocampal areas, giving rise to known places, although no direct link between TF/TH and TGv has been described. Interestingly, recent fMRI data (Miyamoto et al., 2014) indicate that monkey rhinal cortex encodes familiar items, operationalized as middle items in a serial probe task. This type of encoding is appropriate for known RWEs, and by extension, semantic knowledge. In humans, this third level of the ventral stream, the temporal pole, may correspond to the anterior part of the semantic system (Vandenberghe et al., 1996). This association between the temporal pole and semantic memory has its basis in the connections of the pole to memory structures such as rhinal cortex. The third level may also be linked with the amygdala, the structure underlying association between known person and emotions, which has been referred to as personal semantic memory (Olson et al., 2007).

The visual representation of known RWEs at the third level also seems consistent with single cell recorded in the human hippocampal complex showing neuronal selectivity for familiar persons or places, sometimes referred to as visual concept neurons (Quiroga, 2012). This might suggest that visual episodes (events) are also represented at this third level and probably beyond, e.g., in entorhinal cortex and hippocampus. The latter view is supported by the recent study of Miyamoto et al. (2014), who showed that the memory trace of recalled items, operationalized as the first item in a serial probe task, is located in caudal entorhinal cortex and hippocampus of the monkey. A relatively small region may suffice for representing episodes, as this representation may be short-lived. Indeed, if the event is repeated or memorable it may become knowledge

(the fact that somebody looks ill may become part of medicine or history); if it is important for the subject it may become part of autobiographic memory. The dissociation of episodic and semantic memory within the third, known-RWE level is also supported by patients studies (Hirni et al., 2013).

For simplicity we have described the three levels, those processing features, partial RWEs, and known RWEs, as separate components, using anatomy (Kravitz et al., 2013) as a guide. It is possible, however, that the transitions between these levels are gradual. Indeed, as mentioned, the ML face patch is located at the edge of retinotopic cortex and the overlap between retinotopic cortex and some of the more caudal face or body patches may be larger in humans than in monkeys. In the monkey, the body patch is anterior to the MT cluster (Jastorff et al., 2012), but in humans EBA overlaps the retinotopic MT cluster to a large extent (Ferri et al., 2012). Moreover, segregation between the third level, TGv, and the levels below, TE, and beyond, rhinal cortex, might be incomplete, insofar as the anterior parts of TE and the lower bank of STS also exchange bidirectional projections with rhinal cortex. At this level, differences between humans and monkeys may have arisen due to the enlargement of the temporal pole in humans.

The three levels of the ventral stream also appear to differ in the way they develop. The experiment of Srihasam et al. (2012) suggests that the middle level (TE, and by extension perhaps also STS and TF/TH) reflects the individual development, while the earlier retinotopic level is probably species-specific. This explains that although the different retinotopic regions are present in all individual subjects, albeit with some variation in size and location, the number of patches in TE seems more variable among individuals (Bell et al., 2011; Lafer-Sousa and Conway, 2013). The third and final level would remain the most plastic and dependent on lifelong mental activity. Its internal organization is presently unknown.

In *conclusion* we propose that the ventral stream is organized into three levels comprising the ventral retinotopic cortex known as TEO, TE, and TGv in the monkey, and their homologs in human cortex. We attribute to these levels the visual representation of features, partial RWEs, and more speculatively, known, complete RWEs, respectively. Furthermore, the middle level TE and its human equivalent is organized into three parallel substreams related to processing shape in depth, 2D shape, and material properties in order to build up RWE representations.

ACKNOWLEDGMENTS

This work was supported by ERC grant Parietalaction and IUAP grant 7/11.

REFERENCES

Abdollahi, R. O., Glasser, M. F., Dierker, D., VanEssen, D. C., and Orban, G. A. (2013). A probabilistic atlas of 18 human retinotopic areas. *Soc. Neurosci. Abstr.* 824:15.

Arcaro, M. J., McMains, S. A., Singer, B. D., and Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *J. Neurosci.* 29, 10638–10652. doi: 10.1523/JNEUROSCI.2807-09.2009

Arcaro, M. J., Pinsk, M. A., Li, X., and Kastner, S. (2011). Visuotopic organization of macaque posterior parietal cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 31, 2064–2078. doi: 10.1523/JNEUROSCI.3334-10.2011

Bell, A. H., Malecek, N. J., Morin, E. L., Hadj-Bouziane, F., Tootell, R. B., and Ungerleider, L. G. (2011). Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J. Neurosci.* 31, 12229–12240. doi: 10.1523/JNEUROSCI.5865-10.2011

Boussaoud, D., Desimone, R., and Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *J. Comp. Neurol.* 306, 554–575. doi: 10.1002/cne.903060403

Brewer, A. A., Liu, J., Wade, A. R., and Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat. Neurosci.* 8, 1102–1109. doi: 10.1038/nn1507

Brincat, S. L., and Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886. doi: 10.1038/nn1278

Cant, J. S., and Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex* 17, 713–731. doi: 10.1093/cercor/bhk022

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., et al. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307. doi: 10.1093/brain/123.2.291

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., and Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition* 92, 179–229. doi: 10.1016/j.cognition.2002.07.001

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., et al. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study. *J. Neurosci.* 24, 2551–2565. doi: 10.1523/JNEUROSCI.3569-03.2004

Di Cesare, G., Di Dio, C., Rochat, M. J., Sinigaglia, C., Bruschweiler-Stern, N., Stern, D. N., et al. (2013). The neural correlates of 'vitality form' recognition: an fMRI study: this work is dedicated to Daniel Stern, whose immeasurable contribution to science has inspired our research. *Soc. Cogn. Affect. Neurosci.* doi: 10.1093/scan/nst068 [Epub ahead of print].

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010

Dubowitz, D. J., Chen, D. Y., Atkinson, D. J., Grieve, K. L., Gillikin, B., Bradley, W. G., et al. (1998). Functional magnetic resonance imaging in macaque cortex. *Neuroreport* 9, 2213–2218. doi: 10.1097/00001756-199807130-00012

Durand, J. B., Peeters, R., Norman, J. F., Todd, J. T., and Orban, G. A. (2009). Parietal regions processing visual 3D shape extracted from disparity. *Neuroimage* 46, 1114–1126. doi: 10.1016/j.neuroimage.2009.03.023

Ferri, S., Kolster, H., Jastorff, J., and Orban, G. A. (2012). The overlap of the EBA and the MT/V5 cluster. *Neuroimage* 66C, 412–425. doi: 10.1016/j.neuroimage.2012.10.060

Fize, D., Vanduffel, W., Nelissen, K., Denys, K., Chef d'Hotel, C., Faugeras, O., et al. (2003). The retinotopic organization of primate dorsal V4 and surrounding areas: a functional magnetic resonance imaging study in awake monkeys. *J. Neurosci.* 23, 7395–7406.

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Res.* 94, 62–75. doi: 10.1016/j.visres.2013.11.004

Freiwald, W. A., and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851. doi: 10.1126/science.1194908

Freiwald, W. A., Tsao, D. Y., and Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nat. Neurosci.* 12, 1187–1196. doi: 10.1038/nn.2363

Galletti, C., Fattori, P., Gamberini, M., and Kutz, D. F. (1999). The cortical visual area V6: brain location and visual topography. *Eur. J. Neurosci.* 11, 3922–3936. doi: 10.1046/j.1460-9568.1999.00817.x

Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., and Orban, G. A. (2009). The processing of three-dimensional shape from disparity in the human brain. *J. Neurosci.* 29, 727–742. doi: 10.1523/JNEUROSCI.4753-08.2009

Georgieva, S. S., Todd, J. T., Peeters, R., and Orban, G. A. (2008). The extraction of 3D shape from texture and shading in the human brain. *Cereb. Cortex* 18, 2416–2438. doi: 10.1093/cercor/bhn002

Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2010). Cortical connections of the macaque caudal ventrolateral prefrontal areas 45A and 45B. *Cereb. Cortex* 20, 141–168. doi: 10.1093/cercor/bhp087

Goda, N., Tachibana, A., Okazawa, G., and Komatsu, H. (2014). Representation of material properties of objects in the visual cortex of nonhuman primates. *J. Neurosci.* 34, 2660–2673. doi: 10.1523/JNEUROSCI.2593-13.2014

Grefkes, C., and Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *J. Anat.* 207, 3–17. doi: 10.1111/j.1469-7580.2005.00426.x

Harada, T., Goda, N., Ogawa, T., Ito, M., Toyoda, H., Sadato, N., et al. (2009). Distribution of colour-selective activity in the monkey inferior temporal cortex revealed by functional magnetic resonance imaging. *Eur. J. Neurosci.* 30, 1960–1970. doi: 10.1111/j.1460-9568.2009.06995.x

Hirabayashi, T., Takeuchi, D., Tamura, K., and Miyashita, Y. (2013). Microcircuits for hierarchical elaboration of object coding across primate temporal areas. *Science* 341, 191–195. doi: 10.1126/science.1236927

Hirni, D. I., Kivisaari, S. L., Monsch, A. U., and Taylor, K. I. (2013). Distinct neuroanatomical bases of episodic and semantic memory performance in Alzheimer disease. *Neuropsychologia* 51, 930–937. doi: 10.1016/j.neuropsychologia.2013.01.013

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., and Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9379–9384. doi: 10.1073/pnas.96.16.9379

Janssen, P., Vogels, R., and Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288, 2054–2056. doi: 10.1126/science.288.5473.2054

Janssens, T., Zhu, Q., Popivanov, I. D., and Vanduffel, W. (2014). Probalistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. *J. Neurosci.* (in press).

Jastorff, J., Clavagnier, S., Gergely, G., and Orban, G. A. (2011). Neural mechanisms of understanding rational actions: middle temporal gyrus activation by contextual violation. *Cereb. Cortex* 21, 318–329. doi: 10.1093/cercor/bhq098

Jastorff, J., and Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *J. Neurosci.* 29, 7315–7329. doi: 10.1523/JNEUROSCI.4870-08.2009

Jastorff, J., Popivanov, I. D., Vogels, R., Vanduffel, W., and Orban, G. A. (2012). Integration of shape and motion cues in biological motion processing in the monkey STS. *Neuroimage* 60, 911–921. doi: 10.1016/j.neuroimage.2011.12.087

Joly, O., Vanduffel, W., and Orban, G. A. (2009). The monkey ventral premotor cortex processes 3D shape from disparity. *Neuroimage* 47, 262–272. doi: 10.1016/j.neuroimage.2009.04.043

Kayaert, G., Biederman, I., Op de Beeck, H. P., and Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *Eur. J. Neurosci.* 22, 212–224. doi: 10.1111/j.1460-9568.2005.04202.x

Kobatake, E., Wang, G., and Tanaka, K. (1998). Effects of shape discrimination training on the selectivity of inferotemporla cells in adult monkeys. *J. Neurophysiol.* 80, 324–330.

Koida, K., and Komatsu, H. (2007). The effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nat. Neurosci.* 10, 108–116. doi: 10.1038/nn1823

Kolster, H., Janssen, T., Orban, G. A., and Vanduffel, W. (2014). The retinotopic organization of macaque ocipitotemporal cortex anterior to V4 and caudo-ventral to the MT cluster. *J. Neurosci.* (in press).

Kolster, H., Mandeville, J. B., Arsenault, J. T., Ekstrom, L. B., Wald, L. L., and Vanduffel. W. (2009). Visual field map clusters in macaque extrastriate visual cortex. *J. Neurosci.* 29, 7031–7039. doi: 10.1523/JNEUROSCI.0518-09.2009

Kolster, H., Peeters, R., and Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *J. Neurosci.* 30, 9801–9820. doi: 10.1523/JNEUROSCI.2069-10.2010

Kolster, H., Peeters, R., and Orban, G. A. (2011). Ten retinotopically organized areas in the human parietal cortex. *Soc. Neurosci. Abstr.* 851:10.

Kornblith, S., Cheng, X., Ohayon, S., and Tsao, D. Y. (2013). A network for scene processing in the macaque temporal lobe. *Neuron* 79, 766–781. doi: 10.1016/j.neuron.2013.06.015

Köteles, K., De Mazière, P. A., Van Hulle, M., Orban, G. A., and Vogels, R. (2008). Coding of images of materials by macaque inferior temporal cortical neurons. *Eur. J. Neurosci.* 27, 466–482. doi: 10.1111/j.1460-9568.2007.06008.x

Kovacs, G. Y., Vogels, R., and Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proc. Nat. Acad. Sci. U.S.A.* 92, 5587–5592. doi: 10.1073/pnas.92.12.5587

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49. doi: 10.1016/j.tics.2012.10.011

Ku, S. P., Tolias, A. S., Logothetis, N. K., and Goense, J. (2011). fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70, 352–362. doi: 10.1016/j.neuron.2011.02.048

Lafer-Sousa, R., and Conway, B. R. (2013). Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat. Neurosci.* 16, 1870–1878. doi: 10.1038/nn.3555

Larsson, J., and Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *J. Neurosci.* 26, 13128–13142. doi: 10.1523/JNEUROSCI.1657-06.2006

Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539.

Logothetis, N., Peled, S., and Pauls, J. (1998). Development and application of fMRI for visual studies in monkeys. *Soc. Neurosci. Abstr. U.S.A.*, 28.

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Ann. Rev. Neurosci.* 19, 577–621. doi: 10.1146/annurev.ne.19.030196.003045

Mars, R. B., Sallet, J., Neubert, F. X., and Rushworth, M. F. (2013). Connectivity profiles reveal the relationship between brain areas for social cognition in human and monkey temporoparietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10806–10811. doi: 10.1073/pnas.1302956110

Miyamoto, K., Adachi, Y., Osada, T., Watanabe, T., Kimura, H. M., Setsuie, R., et al. (2014). Dissociable memory traces within the macaque medial temporal lobe predict subsequent recognition performance. *J. Neurosci.* 34, 1988–1197. doi: 10.1523/JNEUROSCI.4048-13.2014

Moeller, S., Freiwald, W. A., and Tsao, D. Y. (2008). Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320, 1355–1359. doi: 10.1126/science.1157436

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., et al. (2011). Scene-selective cortical regions in human and nonhuman primates. *J. Neurosci.* 31, 13771–13785. doi: 10.1523/JNEUROSCI.2792-11.2011

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372. doi: 10.1038/nrn2619

Naya, Y., Yoshida, M., and Miyashita, Y. (2003a). Forward processing of long-term associative memory in monkey inferotemporal cortex. *J. Neurosci.* 23, 2861–2871.

Naya, Y., Yoshida, M., Takeda, M., Fujimichi, R., and Miyashita, Y. (2003b). Delay-period activities in two subdivisions of monkey inferotemporal cortex during pair association memory task. *Eur. J. Neurosci.* 18, 2915–2918. doi: 10.1111/j.1460-9568.2003.03020.x

Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., et al. (2011). Action observation circuits in the macaque monkey cortex. *J. Neurosci.* 31, 3743–3756. doi: 10.1523/JNEUROSCI.4803-10.2011

Nelissen, K., Joly, O., Durand, J. B., Todd, J. T., Vanduffel, W., and Orban, G. A. (2009). The extraction of depth structure from shading and texture in the macaque brain. *PLoS ONE* 4:e8306. doi: 10.1371/journal.pone.0008306

Ohayon, S., Freiwald, W. A., and Tsao, D. Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron* 74, 567–581. doi: 10.1016/j.neuron.2012.03.024

Okazawa, G., Goda, N., and Komatsu, H. (2012). Selective responses to specular surfaces in the macaque visual cortex revealed by fMRI. *Neuroimage* 63, 1321–1333. doi: 10.1016/j.neuroimage.2012.07.052

Olson, I. R., Plotzker, A., and Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 130, 1718–1731. doi: 10.1093/brain/awm052

Op De Beeck, H., and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *J. Comp. Neurol.* 426, 505–518. doi: 10.1002/1096-9861(20001030)426:4<505::AID-CNE1>3.0.CO;2-M

Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiol. Rev.* 88, 59–89. doi: 10.1152/physrev.00008.2007

Orban, G. A. (2011). The extraction of 3D shape in the visual system of human and nonhuman primates. *Annu. Rev. Neurosci.* 34, 361–388. doi: 10.1146/annurev-neuro-061010-113819

Orban, G. A., and Jastorff, J. (2014). "Functional mapping of motion regions in human and non-human primates," in *The New Visual Neurosciences*, eds J. S. Werner and L. M. Chalupa (Cambridge: MIT press), 777–791.

Orban, G. A., and Vogels, R. (1998). The neuronal machinery involved in successive discrimination. *Prog. Neurobiol.* 55, 117–147. doi: 10.1016/S0301-0082(98)00010-0

Pasupathy, A., and Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86, 2505–2519.

Peuskens, H., Claeys, K. G., Todd, J. T., Norman, J. F., Van Hecke, P., and Orban G. A. (2004). Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J. Cogn. Neurosci.* 16, 665–682. doi: 10.1162/089892904323057371

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., et al. (2009). Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. *J. Neurophysiol.* 101, 2581–2600. doi: 10.1152/jn.91198.2008

Pitzalis, S., Galletti, C., Huang, R. S., Patria, F., Committeri, G., Galati, G., et al. (2006). Wide-field retinotopy defines human cortical visual area v6. *J. Neurosci.* 26, 7962–7973. doi: 10.1523/JNEUROSCI.0178-06.2006

Polosecki, P., Moeller, S., Schweers, N., Romanski, L. M., Tsao, D. Y., and Freiwald, W. A. (2013). Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *J. Neurosci.* 33, 11768–11773. doi: 10.1523/JNEUROSCI.5402-11.2013

Popivanov, I. D., Jastorff, J., Vanduffel, W., and Vogels, R. (2012). Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *Neuroimage* 63, 723–741. doi: 10.1016/j.neuroimage.2012.07.013

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* 13, 587–597. doi: 10.1038/nrn3251

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107. doi: 10.1038/nature03687

Saleem, K. S., Kondo, H., and Price, J. L. (2008). Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *J. Comp. Neurol.* 506, 659–693. doi: 10.1002/cne.21577

Saleem, K. S., Price, J. L., and Hashikawa, T. (2007). Cytoarchitectonic and chemoarchitectonic subdivisions of the perirhinal and parahippocampal cortices in macaque monkeys. *J. Comp. Neurol.* 500, 973–1006. doi: 10.1002/cne.21141

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S., et al. (2011). Social network size affects neural circuits in macaques. *Science* 334, 697–700. doi: 10.1126/science.1210027

Sato, T., Uchida, G., Lescroart, M. D., Kitazono, J., Okada, M., and Tanifuji, M. (2013). Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *J. Neurosci.* 33, 16642–16656. doi: 10.1523/JNEUROSCI.5557-12.2013

Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., and Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J. Neurosci.* 25, 4294–4306. doi: 10.1523/JNEUROSCI.0377-05.2005

Saxe, R., Carey, S., and Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.* 55, 87–124. doi: 10.1146/annurev.psych.55.090902.142044

Sereno, M. E., Trinath, T., Augath, M., and Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. *Neuron* 33, 635–652. doi: 10.1016/S0896-6273(02)00598-6

Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320. doi: 10.1038/415318a

Silver, M. A., and Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends Cogn. Sci.* 13, 488–495. doi: 10.1016/j.tics.2009.08.005

Silver, M., A., Ress, D., and Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *J. Neurophysiol.* 94, 1358–1371. doi: 10.1152/jn.01316.2004

Singer, J. M., and Sheinberg, D. L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *J. Neurosci.* 30, 3133–3145. doi: 10.1523/JNEUROSCI.3211-09.2010

Srihasam, K., Mandeville, J. B., Morocz, I. A., Sullivan, K. J., and Livingstone, M. S. (2012). Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron* 73, 608–619. doi: 10.1016/j.neuron.2011.12.022

Stefanacci, L., Reber, P., Costanza, J., Wong, E., Buxton, R., Zola, S., et al. (1998). fMRI of monkey visual cortex. *Neuron* 20, 1051–1057. doi: 10.1016/S0896-6273(00)80485-7

Takechi, H., Onoe, H., Shizuno, H., Yoshikawa, E., Sadato, N., Tsukuda, H., et al. (1997). Mapping of cortica area sinvolved in color vision in non-human primates. *Neurosci. Lett.* 230, 17–20. doi: 10.1016/S0304-3940(97)00461-8

Takeda, M., Naya, Y., Fujimichi, R., Takeuchi, D., and Miyashita, Y. (2005). Active maintenance of associative mnemonic signal in monkey inferior temporal cortex. *Neuron* 48, 839–84 doi: 10.1016/j.neuron.2005.09.028

Takeuchi, D., Hirabayashi, T., Tamura, K., and Miyashita, Y. (2011). Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* 331, 1443–1447. doi: 10.1126/science.1199967

Tanaka, K. (1996). Inferotemporal cortex, object vision. *Annu. Rev. Neurosci.* 19, 109–139. doi: 10.1146/annurev.ne.19.030196.000545

Tanaka, K., Saito, H., Fukuda, Y., and Moriya, M. (1991). Coding visual images of objects in infero-temporal cortex of the macaque monkey. *J. Neurophysiol.* 66, 170–189.

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., et al. (1997). Functional analysis of V3A and related areas in human visual cortex. *J. Neurosci.* 17, 7060–7078.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995. doi: 10.1038/nn1111

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674. doi: 10.1126/science.1119983

Ungerleider, L. G., Galkin, T. W., Desimone, R., and Gattass, R. (2008). Cortical connections of area V4 in the macaque. C*ereb. Cortex* 18, 477–499. doi: 10.1093/cercor/bhm061

Vandenberghe, R., Price, C., Wise, R., Josephs, O., and Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature* 383, 254–256. doi: 10.1038/383254a0

Vanduffel, W., Beatse, E., Sunaert, S., Van Hecke, P., Tootell, R. B. H., and Orban, G. A. (1998). Functional magnetic resonance imaging in an awake rhesus monkey. *Soc. Neurosci. Abstr.* 24, 11.

Vanduffel, W., Fize, D., Mandeville, J. B., Nelissen, K., Van Hecke, P., Rosen, B. R., et al. (2001). Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. *Neuron* 32, 565–577. doi: 10.1016/S0896-6273(01)00502-5

Vangeneugden, J., Pollick, F., and Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. *Cereb. Cortex* 19, 593–611. doi: 10.1093/cercor/bhn109

Vogels, R., and Orban, G. A. (1994). Does practice in orientation discrimination lead to changes in the response properties of macaque inferior temporal neurons? *Eur. J. Neurosci.* 6, 1680–1694. doi: 10.1111/j.1460-9568.1994.tb00560.x

Wandell, B. A., Dumoulin, S. O., and Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron* 56, 366–383. doi: 10.1016/j.neuron.2007.10.012

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360. doi: 10.1038/nn.2202

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature* 274, 423–428. doi: 10.1038/274423a0

Zeki, S., and Marini, L. (1998). Three cortical stages of colour processing in the human brain. *Brain* 121, 1669–1685. doi: 10.1093/brain/121.9.1669

Zhang, N. R., and von der Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *J. Neurosci.* 30, 6482–6496. doi: 10.1523/JNEUROSCI.5168-09.2010

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 February 2014; accepted: 16 June 2014; published online: 02 July 2014.
*Citation: Orban GA, Zhu Q and Vanduffel W (2014) The transition in the ventral stream from feature to real-world entity representations. Front. Psychol. **5**:695. doi: 10.3389/fpsyg.2014.00695*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Orban, Zhu and Vanduffel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 17 July 2014
doi: 10.3389/fpsyg.2014.00730

Visuo-haptic multisensory object recognition, categorization, and representation

***Simon Lacey*[1] * *and K. Sathian*[1,2,3,4]**

[1] *Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA*
[2] *Department of Rehabilitation Medicine, Emory University School of Medicine, Atlanta, GA, USA*
[3] *Department of Psychology, Emory University School of Medicine, Atlanta, GA, USA*
[4] *Rehabilitation Research and Development Center of Excellence, Atlanta Veterans Affairs Medical Center, Decatur, GA, USA*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Carl M. Gaspar, Hangzhou Normal University, China
Mounia Ziat, Northern Michigan University, USA

****Correspondence:***
Simon Lacey, Department of Neurology, Emory University School of Medicine, WMB-6000, 101 Woodruff Circle, Atlanta, GA 30322, USA
e-mail: slacey@emory.edu

Visual and haptic unisensory object processing show many similarities in terms of categorization, recognition, and representation. In this review, we discuss how these similarities contribute to multisensory object processing. In particular, we show that similar unisensory visual and haptic representations lead to a shared multisensory representation underlying both cross-modal object recognition and view-independence. This shared representation suggests a common neural substrate and we review several candidate brain regions, previously thought to be specialized for aspects of visual processing, that are now known also to be involved in analogous haptic tasks. Finally, we lay out the evidence for a model of multisensory object recognition in which top-down and bottom-up pathways to the object-selective lateral occipital complex are modulated by object familiarity and individual differences in object and spatial imagery.

Keywords: cross-modal, effective connectivity, fMRI, viewpoint dependence, face processing, visual imagery

INTRODUCTION

Despite the fact that object perception and recognition are invariably multisensory processes in real life, the haptic modality was for a long time the poor relation in a field dominated by vision science, with the other senses lagging even further behind (Gallace and Spence, 2009; Gallace, 2013). Two things have happened to change this: firstly, from the 1980s, haptics has developed as a field in its own right; secondly, from the 1990s, there has been an accelerated interest in multisensory interactions. Here, we review the interactions and commonalities in visuo-haptic multisensory object processing, beginning with the capabilities and limits of haptic and visuo-haptic recognition. One way to facilitate recognition is to group like objects together: hence, we review recent work on the similarities between visual and haptic categorization and cross-modal transfer of category knowledge. Changes in orientation and size present a major challenge to within-modal object recognition. However, these obstacles seem to be absent in cross-modal recognition and we show that a shared representation underlies both cross-modal recognition and view-independence. We next compare visual and haptic representations from the point of view of individual differences in preferences for object or spatial imagery. A shared representation for vision and touch suggests shared neural processing and therefore we review a number of candidate brain regions, previously thought to be selective for visual aspects of object processing, which have subsequently been shown to be engaged by analogous haptic tasks. This reflects the growing consensus around the concept of a "metamodal" brain with a task-based organization and multisensory inputs, rather than organization around discrete unisensory inputs (Pascual-Leone and Hamilton, 2001; Lacey et al., 2009a; James et al., 2011). Finally, we draw these threads together and discuss the evidence for a model of multisensory visuo-haptic object recognition in which representations are flexibly accessible by either top-down or bottom-up pathways depending on object familiarity and individual differences in imagery preference (Lacey et al., 2009a).

HAPTIC AND VISUO-HAPTIC OBJECT RECOGNITION

The speed and accuracy of visual object recognition is well-established. Haptic recognition, albeit less well studied, is somewhat slower than visual recognition, but, at least for everyday objects, is still fairly fast and highly accurate with 96% correctly named: 68% in less than 3 s and 94% within 5 s (Klatzky et al., 1985); indeed, a "haptic glance" of less than 1 s suffices in some circumstances (Klatzky and Lederman, 1995). Longer response times in the study of Klatzky et al. (1985) likely reflect the time taken to explore some of the larger items such as a tennis racket or hairdryer. A remarkable fact about haptic processing is that it can be achieved with the feet as well as the hands, albeit more slowly and less accurately, with hand and foot performance being highly correlated across individuals (Lawson, 2014). Haptic identification proceeds, with increasing accuracy, from a "grasp and lift" stage that extracts basic low-level information about a variety of object properties to a series of hand movements that extract more precise information (Klatzky and Lederman, 1992). These hand movements, known as "exploratory procedures," are property-specific, for example, lateral motion is used to assess texture and contour-following to precisely assess shape (Lederman and Klatzky, 1987). These properties differ in salience to haptic processing depending on the context: under neutral instructions, salience progressively decreases in this order: hardness > texture > shape; under instructions that emphasized haptic processing, the order changes to texture > shape > hardness

(Klatzky et al., 1987). Note that the saliency order under neutral instructions is reversed to shape > texture > hardness/size in simultaneous visual and haptic perception, and in haptic perception under instructions to use concurrent visual imagery (Klatzky et al., 1987).

Overall, cross-modal visuo-haptic object recognition, while fairly accurate, comes at a cost compared to within-modal recognition (e.g., Bushnell and Baxt, 1999; Casey and Newell, 2007; and see Lacey et al., 2007). Cross-modal performance is generally better when visual encoding is followed by haptic retrieval than the reverse (e.g., Jones, 1981; Streri and Molina, 1994; Lacey and Campbell, 2006). This asymmetry appears to be a consistent feature of visuo-haptic cross-modal memory but has generally received little attention (e.g., Easton et al., 1997a,b; Reales and Ballesteros, 1999; Nabeta and Kawahara, 2006). One explanation for cross-modal asymmetry might be that shape information is not encoded equally well by the visual and haptic systems, because of competition from other, more salient, modality-specific object properties. Thus, in the haptic-visual cross-modal condition it might be more difficult to encode shape because of the more salient hardness and texture information, as noted above. This effect might be suppressed by the use of concurrent visual imagery in which shape information, common to vision and touch, might be brought to the fore. We should note, however, that when vision and touch are employed simultaneously, properties that are differently weighted in these modalities may be optimally combined on the basis of maximum likelihood estimates (see Ernst and Banks, 2002; Helbig and Ernst, 2007; Helbig et al., 2012; Takahashi and Watt, 2014).

Another explanation for cross-modal asymmetry could be differences in visual and haptic memory capacity. Haptic working memory capacity appears to be limited and variable, and may therefore be more error-prone than visual working memory (Bliss and Hämäläinen, 2005). Alternatively, haptic representations may simply decay faster than visual representations. Rather than a progressive decline over time, the haptic decay function appears to occur entirely in a band of 15–30 s post-stimulus (Kiphart et al., 1992). Consistent with this, a more recent study showed no decline in performance at 15 s (Craddock and Lawson, 2010) although longer intervals were not tested. Haptic-visual performance might therefore be lower because by the time visual recognition is tested, haptically encoded representations have substantially decayed. However, other cross-modal memory studies show that delays up to 30 s (Garvill and Molander, 1973; Woods et al., 2004) or even a week (Pensky et al., 2008) did not affect haptic-visual recognition more than visual-haptic recognition. Thus, an explanation in terms of a simple function of haptic memory properties is likely insufficient.

Cross-modal asymmetry is observed even in very young infants where it is ascribed to constraints imposed by different stages of motor development (Streri and Molina, 1994). But this explanation is also unsatisfactory since the asymmetry persists into maturity (Easton et al., 1997a,b; Bushnell and Baxt, 1999; Lacey and Campbell, 2006). Interestingly, implicit memory does not appear to be affected: cross-modal priming is symmetric (Easton et al., 1997a,b; Reales and Ballesteros, 1999) although verbal encoding strategies may have played a mitigating role in these studies. A recent study suggests that underlying neural activity is asymmetric between the two cross-modal conditions. Using a match-to-sample task, Kassuba et al. (2013) showed that bilateral lateral occipital complex (LOC), fusiform gyrus (FG), and anterior intraparietal sulcus (aIPS) selectively responded more strongly to crossmodal, compared to unimodal, object matching when haptic targets followed visual samples, and more strongly still when the haptic target and visual sample were congruent rather than incongruent; however, these regions showed no such increase for visual targets in either crossmodal or unimodal conditions. This asymmetric increase in activation in the visual-haptic condition may reflect multisensory binding of shape information and suggests that haptics – traditionally seen as the less reliable modality – has to integrate previously presented visual information more than vision has to integrate previous haptic information (Kassuba et al., 2013).

OBJECT CATEGORIZATION

Categorization facilitates recognition and is critical for much of higher-order cognition (Graf, 2010); hitherto, the emphasis in terms of perceptual categorization has been almost exclusively on the visual, rather than the haptic, modality. More recently, however, a series of studies has systematically compared visual and haptic categorization. Using multi-dimensional scaling analysis, these studies showed that visual and haptic similarity ratings and categorization result in perceptual spaces [i.e., topological representations of the perceived (dis)similarity along a given dimension] that are highly congruent between modalities for novel 3-D objects (Cooke et al., 2007), more realistic 3-D shell-like objects (Gaißert et al., 2008, 2010, 2011) and for natural objects, i.e., actual seashells (Gaißert and Wallraven, 2012). This was so in both unisensory and bisensory conditions (Cooke et al., 2007) and whether 2-D visual objects were compared to haptic 3-D objects (Gaißert et al., 2008, 2010) or passive viewing of 2-D objects was compared to interactive viewing and active haptic exploration of 3-D objects, i.e., such that visual and haptic exploration were more similar (Gaißert et al., 2010). These highly similar visual and haptic perceptual spaces both showed high fidelity to the physical object space [i.e., a topological representation of the actual (dis)similarity along a given dimension; Gaißert et al., 2008, 2010], retaining the category structure (the ordinal adjacency relationships within the category, i.e., the actual progression in variation along a given dimension, for example from roughest to smoothest; Cooke et al., 2007). The isomorphism between perceptual (in either modality) and physical spaces was, furthermore, task-independent, whether simple similarity rating (Gaißert et al., 2008), unconstrained (free sorting), semi-constrained (making exactly three groups) or constrained (matching to a prototype object) categorization (Gaißert et al., 2011). As in vision, haptics also exhibits categorical perception, i.e., discriminability increases sharply when objects belong to different categories and decreases when they belong to the same category (Gaißert et al., 2012).

However, visual and haptic categorization are not entirely alike and, consistent with differential perceptual salience (Klatzky et al., 1987), object properties are differentially weighted depending on the modality, whether they are controlled parametrically (Cooke

et al., 2007) or vary naturally (Gaißert and Wallraven, 2012). Shape was more important than texture for visual categorization whereas in haptic and bisensory categorization, shape and texture were approximately equally weighted (Cooke et al., 2007), although in this study shape and texture varied in ways that were intuitive to vision and haptics (broadly, width for shape and smoothness for texture). Using specially manufactured shell-like objects, Gaißert et al. (2010) varied three complex shape parameters that were not intuitive to either modality. While visual and haptic perceptual spaces and the physical object space were all highly similar, the shape dimensions were weighted differently: symmetry was more important than convolutions for vision while the reverse was true for haptics; aperture-tip distance was the least important factor for both modalities (Gaißert et al., 2010). For natural objects – seashells – that varied naturally in a number of properties, similarity ratings and categorization were still driven by global and local shape parameters rather than size, texture, weight etc. (Gaißert and Wallraven, 2012).

These studies suggest a close connection between vision and haptics in terms of similarity mechanisms for categorization but do not necessarily imply a shared representation because of the differential weighting of object properties in each modality. Nonetheless, there is symmetric cross-modal transfer of category information following either visual or haptic category learning, even for complex novel 3-D objects, and furthermore this transfer generalizes to new objects from these categories (Yildirim and Jacobs, 2013). A recent study shows that not only does category membership transfer cross-modally, as shown by Yildirim and Jacobs (2013), but so does category structure (Wallraven et al., 2013), i.e., the ordinal relationships and category boundaries (see Cooke et al., 2007) transcend modality. Crossmodal transfer of category structure is interesting because the ordering of each item within the category is (at least in the studies reviewed here) perceptually driven; thus it may be that a shared multisensory representation underlies cross-modal categorization, as has been suggested for cross-modal recognition (Lacey et al., 2009a; Lacey and Sathian, 2011).

Of course, perceptual similarity is not the only basis for categorization (Smith et al., 1998) and neither vision nor haptics appear to naturally recover categories on alternative bases that are more abstract or semantic. For example, Haag (2011) used realistically textured models of familiar animals that retained real-life size relations, and required visual and haptic categorization on the basis of size (big/small in real life), domesticity (wild/domestic), and predation (carnivore/herbivore). Errors increased as the basis of categorization moved from concrete (size) to abstract (predation) and were consistently greater in haptics than vision (Haag, 2011). Similarly, neither vision nor haptics naturally recovered the taxonomic relationships between the natural seashells used by Gaißert and Wallraven (2012): participants distinguished between concrete categories such as whether the shells used were flat or convoluted, rather than between abstract categories such as gastropods (e.g., sea-snail) vs. bivalves (e.g., oyster). If biological relationships were recovered at all, this was mainly contingent on shape similarities, although vision was better than haptics in this respect (Gaißert and Wallraven, 2012) as it was for the abstract categories studied by Haag (2011).

FACES: A SPECIAL CATEGORY

Faces are a special category of object that we encounter every day and at which we are especially expert, being able to differentiate large numbers of individuals (Maurer et al., 2002). We are also able to recognize faces under conditions that would impair recognition in other categories; for example, bad lighting or changes in viewpoint (Maurer et al., 2002) – though face recognition is impaired if the face is upside-down (Yin, 1969). An important distinction is made between configural and featural processing: the former refers to processing the spatial relationships between individual facial features as well as the shapes of the features themselves, while the latter refers to the piecemeal processing of individual face parts (Maurer et al., 2002; Dopjans et al., 2012). Although sighted humans obviously recognize faces almost exclusively through vision, live faces can also be identified haptically with high levels of accuracy (over 70%), whether they are learned through touch alone or using both vision and touch (Kilgour and Lederman, 2002). Interestingly, when participants had to haptically identify clay masks produced from live faces, accuracy was significantly lower than for live faces, suggesting that natural material cues and surface properties are important for haptic face recognition (Kilgour and Lederman, 2002). Visual experience may be necessary for haptic face recognition, since the congenitally blind were significantly less accurate than both the sighted and the late-blind (Wallraven and Dopjans, 2013). Nonetheless, haptic face recognition is not as good as visual recognition in the sighted either (Dopjans et al., 2012). This may be due to basic differences between visual and haptic processing. Haptic exploration of any object is almost exclusively sequential and serial (Lederman and Klatzky, 1987; Loomis et al., 1991) whilst visual processing is massively parallel (see Nassi and Callaway, 2009). In the context of face processing, therefore, haptics might be restricted to featural processing, in which individual features are processed independently and have to be assembled into a face context, which may account for lower haptic performance compared to visual configural encoding (Dopjans et al., 2012). When visual encoding was restricted, by using a participant-controlled moving window that only revealed a small portion of the face at a time, so that it was more like haptic sequential processing, visual and haptic performance were more equal (Dopjans et al., 2012), suggesting that any differences arise from different encoding strategies[1].

Despite these various differences in performance, visual and haptic face processing do have common aspects. For example, consistent with the shared perceptual spaces discussed above (e.g., Gaißert et al., 2008, 2010, 2011; Gaißert and Wallraven, 2012), there is evidence for similar "face-spaces" for vision and touch in which, again, different properties carry different weights depending on the modality (Wallraven, 2014). The evidence for a face-inversion effect – better recognition when faces are upright than inverted, an effect not seen for non-face categories – is clear for vision but less so for haptics. Kilgour and Lederman (2006) showed a clear haptic inversion effect for faces compared to non-face stimuli, whereas Dopjans et al. (2012) found an inversion

[1]For discussions of configural versus featural visual face processing, see Peterson and Rhodes (2003).

effect for unrestricted visual, but not for haptic or restricted visual, face encoding. In "face adaptation," a neutral face is perceived as having the opposite facial expression to a previously perceived face; for example, adaptation to a sad face leads to perception of a happy face upon subsequent presentation of a face with a neutral expression (e.g., Skinner and Benton, 2010). Such an effect is also seen in within-modal haptic adaptation to faces (Matsumiya, 2012) and transfers cross-modally both from vision to touch and *vice versa*, indicating that haptic face-related information and visual face processing share some common processing (Matsumiya, 2013).

Faces can also be recognized cross-modally between vision and touch (Kilgour and Lederman, 2002); this comes at a cost relative to within-modal recognition (Casey and Newell, 2007) although the cost decreases with familiarity (Casey and Newell, 2005). However, this disadvantage for cross-modal face recognition is unrelated to the encoding modality or to differences in encoding strategies, which suggests that, in contrast to object recognition (see below), vision and touch do not share a common face representation (Casey and Newell, 2007). On the other hand, visually presented faces disrupt identification of haptic faces when their facial expressions are incongruent and facilitate identification when they are congruent (Klatzky et al., 2011) which suggests a shared representation although response competition cannot be excluded as an explanation for these results. However, taken in conjunction with the finding that a visually prosopagnosic patient (i.e., a patient unable to recognize faces visually despite intact basic visual perception) was also unable to recognize faces haptically (Kilgour et al., 2004), a shared representation seems likely.

OBSTACLES TO EFFICIENT RECOGNITION

VIEW-DEPENDENCE

A change in the orientation of an object changes the related sensory input, e.g., retinal pattern, such that recognition is potentially impaired; an important goal of sensory systems is therefore to achieve perceptual constancy so that objects can be recognized independently of such changes. Visual object recognition is considered view-dependent if rotating an object away from its original orientation impairs subsequent recognition and view-independent if not (reviewed in Peissig and Tarr, 2007). During haptic exploration, the hands can contact an object from different sides simultaneously: intuitively, therefore, one might expect information about several different "views" to be acquired at the same time and that haptic recognition would be view-independent. However, numerous studies have now shown that this intuition is not correct and that haptic object recognition is also view-dependent (Newell et al., 2001; Lacey et al., 2007, 2009b; Ueda and Saiki, 2007, 2012; Craddock and Lawson, 2008, 2010; Lawson, 2009, 2011). The factors underlying haptic view-dependence are not currently known: even unlimited exploration time and orientation cuing do not reduce view-dependence (Lawson, 2011). It is interesting to examine how vision and touch are affected by different types of rotation. Visual recognition is differentially impaired by changes in orientation depending on the axis around which an object is rotated (Gauthier et al., 2002; Lacey et al., 2007). Recognition is slower and less accurate when objects are rotated about the x- and y-axes, i.e., in depth (**Figure 1**), than when rotated about the z-axis, i.e., in the picture plane, for both 2-D (Gauthier et al., 2002) and 3-D stimuli (Lacey et al., 2007). By contrast, haptic recognition is equally impaired by rotation about any axis (Lacey et al., 2007), suggesting that, although vision and haptics are both view-dependent, the basis for this is different in each modality. One possible explanation is that vision and haptics differ in whether or not a surface is occluded by rotation. In vision, a change in orientation can involve not only a transformation in perceptual shape but also occlusion of one or more surfaces – unless the observer physically changes position relative to the object (e.g., Pasqualotto et al., 2005; Pasqualotto and Newell, 2007). Compare, for example, **Figures 1A,C** – rotation about the x-axis means that the object is turned upside-down and that the former top surface becomes occluded. In haptic exploration, the hands are free to move over all surfaces of an object and to manipulate it into different orientations relative to the hand, thus in any given orientation, no surface is necessarily occluded, provided the object is small enough. If this is true, then no single axis of rotation should be more or less disruptive than another due to surface occlusion, so that haptic recognition only has to deal with a shape transformation. Further work is required to examine whether this explanation is, in fact, correct.

View-dependence mostly occurs when objects are unfamiliar. Increasing object familiarity reduces the disruptive effect of orientation changes and visual recognition tends to become view-independent (Tarr and Pinker, 1989; Bülthoff and Newell, 2006). An exception to this is when a familiar object is typically seen in one specific orientation known as a canonical view, for example the front view of a house (Palmer et al., 1981). View-independence may still occur for a limited range of orientations around the canonical view, but visual recognition is impaired for radically non-canonical views, for example, a teapot seen from directly above (Palmer et al., 1981; Tarr and Pinker, 1989; Bülthoff and Newell, 2006). Object familiarity also results in haptic view-independence and this remains so even where there is a change in the hand used to explore the object (Craddock and Lawson, 2009a). Haptic recognition also reverts to view-dependence for non-canonical orientations (Craddock and Lawson, 2008). However, vision and haptics differ in what constitutes a canonical view. The preferred view in vision is one in which the object is aligned at 45° to the observer (Palmer et al., 1981) while objects are generally aligned either parallel or orthogonal to the body midline in haptic canonical views (Woods et al., 2008). Canonical views may facilitate view-independent recognition either because they provide the most structural information about an object or because they most closely match a stored representation, but the end result is the same for both vision and haptics (Craddock and Lawson, 2008; Woods et al., 2008).

In contrast to within-modal recognition, visuo-haptic cross-modal recognition is view-independent even for unfamiliar objects that are highly similar (**Figure 1**), whether visual study is followed by haptic test or *vice versa* and whatever the axis of rotation (Lacey et al., 2007, 2010b; Ueda and Saiki, 2007, 2012). Haptic-visual, but not visual-haptic, cross-modal view-independence has been shown for familiar objects (Lawson, 2009). This asymmetry might

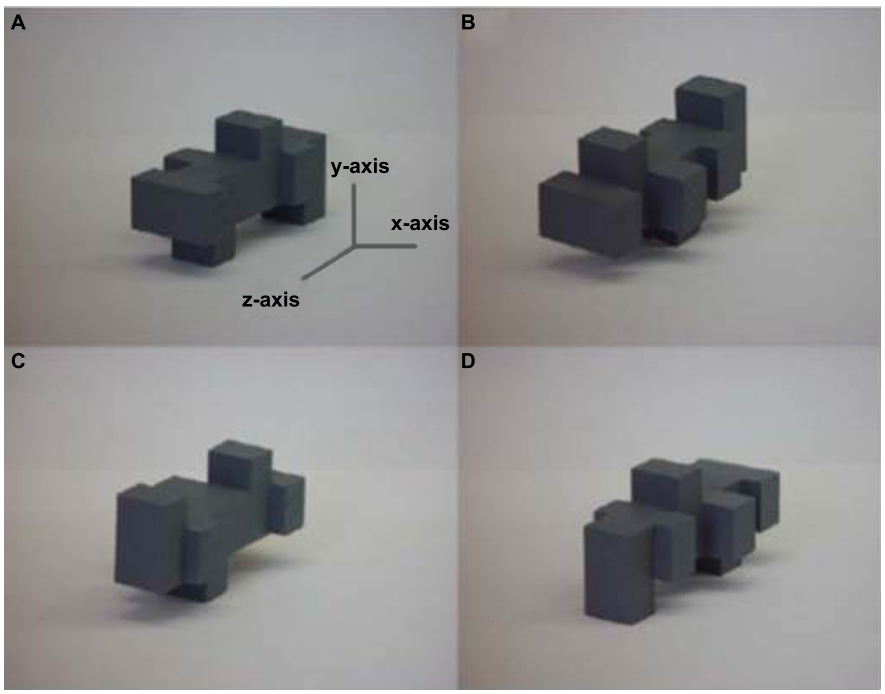


FIGURE 1 | Example 3-D unfamiliar object shown (A) in the original orientation and rotated 180° about the (B) z-axis, (C) x-axis, and (D) y-axis: rotation about the x- and y-axes are rotations in depth, rotation about the z-axis is a rotation in the picture-plane. Figure adapted from Lacey et al. (2007).

be due to the fact that the familiar objects used in this particular study were a mixture of scale models (e.g., bed, bath, and shark) and actual-size objects (e.g., jug, pencil); thus, some of these might have been more familiar visually than haptically, resulting in greater error when visually familiar objects had to be recognized by touch. Additional research on the potentially disruptive effects of differential familiarity is merited.

A strange finding is that knowledge of the test modality does not appear to help achieve view-independence. When participants knew the test modality, both visual and haptic within-modal recognition were view-dependent whereas cross-modal recognition was view-independent (Ueda and Saiki, 2007, 2012), but when the test modality was unknown both within- and cross-modal recognition were view-independent (Ueda and Saiki, 2007). At first glance this is puzzling: one would expect that knowledge of the test modality would confer an advantage. However, Ueda and Saiki (2012) showed that eye movements differed during encoding, with longer and more diffuse fixations when participants knew that they would be tested cross-modally (visual-haptic only) compared to within-modally. It is possible that, on the "principle of least commitment" (Marr, 1976), the same pattern of eye movements occurs when the test modality is not known (i.e., it is not possible to commit to an outcome), preserving as much information as possible and resulting in both within- and cross-modal view-independence. Further examination of eye movements during both cross-modal conditions would be valuable, as eye movements could serve as behavioral markers for the multisensory view-independent representation discussed next.

The simplest way in which cross-modal view-independence could arise is that the view-dependent visual and haptic unisensory representations are directly integrated into a view-independent multisensory representation (**Figure 2A**). An alternative explanation is that unisensory view-independence in vision and haptics is a precondition for cross-modal view-independence (**Figure 2B**). In a perceptual learning study, view-independence acquired by learning in one modality transferred completely and symmetrically to the other; thus, whether visual or haptic, within-modal view-independence relies on a single view-independent representation (Lacey et al., 2009b). Furthermore, both visual and haptic within-modal view-independence were acquired following cross-modal training (whether haptic-visual or visual-haptic); we therefore concluded that visuo-haptic view-independence is supported by a single multisensory representation that directly integrates the unisensory view-dependent representations (Lacey et al., 2009b; **Figure 2A**), similar to models that have been proposed for vision (Riesenhuber and Poggio, 1999). Thus, the same representation appears to support both cross-modal recognition and view-independence (whether within- or cross-modal).

SIZE-DEPENDENCE

In addition to achieving object constancy across orientation changes, the visual system also has to contend with variations in the size of the retinal image that arise from changes in object-observer distance: the same object can produce retinal images that vary in size depending on whether it is near to, or far from, the observer. Presumably, this is compensated by cues arising from

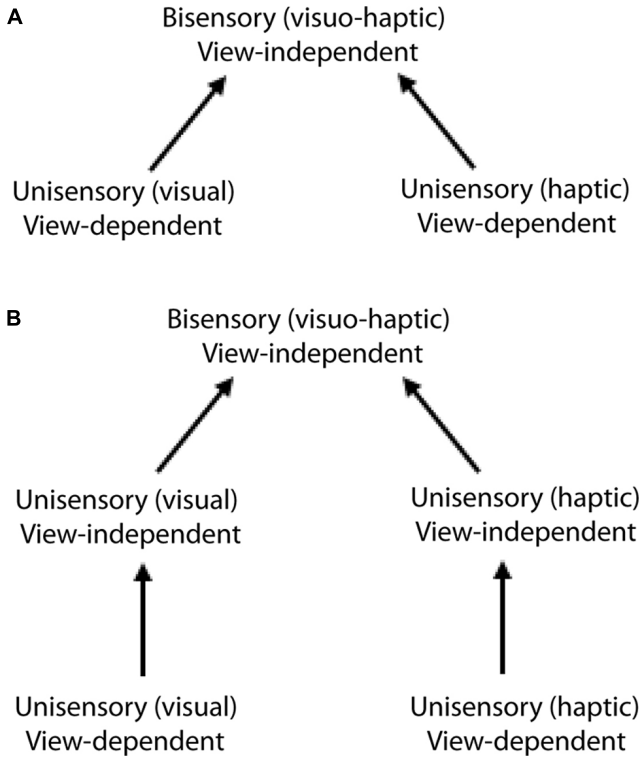


FIGURE 2 | Alternative models of visuo-haptic view-independence: (A) direct integration of the unisensory view-dependent representation into a multisensory view-independent representation; (B) bisensory view-independence gated by separate, unisensory view-independent representations. Evidence supports the direct integration model **(A)**. Figure adapted from Lacey et al. (2009b).

depth or motion perception, accounting for the fact that a change in size does not disrupt visual object identification (Biederman and Cooper, 1992; Uttl et al., 2007). However, size change does produce a cost in visual recognition for both unfamiliar (Jolicoeur, 1987) and familiar objects (Jolicoeur, 1987; Uttl et al., 2007). Interestingly, changes in retinal size due to movement of the observer result in better size-constancy than those due to movement of the object (Combe and Wexler, 2010).

Haptic size perception requires integration of both cutaneous (contact area and force) and proprioceptive (finger spread and position) information at initial contact (Berryman et al., 2006). Neither gripping an object tighter, which increases contact area, nor enlarging the spread of the fingers leads us to perceive a change in size (Berryman et al., 2006). Thus, in contrast to vision where perceived size varies with distance, in touch, physical size is perceived directly, i.e., haptic size equals physical size. It is intriguing then, that haptic (Craddock and Lawson, 2009b,c) and cross-modal (Craddock and Lawson, 2009c) recognition are apparently size-dependent and this merits further investigation. Further research should address whether haptic representations store a canonical size for familiar objects (as has recently been proposed for visual representations, Konkle and Oliva, 2011), deviations from which could impair recognition, and whether object constancy can be achieved across size changes in unfamiliar objects.

REPRESENTATIONS AND INDIVIDUAL DIFFERENCES

A crucial question for object recognition is what information is contained in the mental representations that support it. Visual shape, color, and texture are processed in different cerebral cortical areas (Cant and Goodale, 2007; Cant et al., 2009) but these structural (shape) and surface (color, texture, etc.) properties are integrated in visual object representations (Nicholson and Humphrey, 2003). Changing the color of an object or its part-color combinations between study and test impaired shape recognition, while altering the background color against which objects were presented did not (Nicholson and Humphrey, 2003). This effect could therefore be isolated to the object representation, indicating that this contains both shape and color information (Nicholson and Humphrey, 2003). Visual and haptic within-modal object discrimination are similarly impaired by a change in surface texture (Lacey et al., 2010b), showing firstly that haptic representations also integrate structural and surface properties and secondly that information about surface properties in visual representations is not limited to modality-specific properties like color. In order to investigate whether surface properties are integrated into the multisensory representation underlying cross-modal object discrimination, we tested object discrimination across changes in orientation (thus requiring access to the view-independent multisensory representation discussed above), texture or both. In line with earlier findings (Lacey et al., 2007; Ueda and Saiki, 2007, 2012), cross-modal object discrimination was view-independent when texture did not change; but if texture did change, performance was reduced to chance levels, whether orientation also changed or not (Lacey et al., 2010b). However, some participants were more affected by the texture changes than others. We wondered whether this arose from individual differences in the nature of object representations, which can be conveniently indexed by preferences for different kinds of imagery.

Two kinds of visual imagery have been described: "object imagery" (involving pictorial images that are vivid and detailed, dealing with the literal appearance of objects in terms of shape, color, brightness, etc.) and "spatial imagery" (involving schematic images more concerned with the spatial relations of objects, their component parts, and spatial transformations; Kozhevnikov et al., 2002, 2005; Blajenkova et al., 2006). An experimentally important difference is that object imagery includes surface property information while spatial imagery does not. To establish whether object and spatial imagery differences occur in touch as well as vision, we required participants to discriminate shape across changes in texture, and texture across changes in shape (**Figure 3**), in both visual and haptic within-modal conditions. We found that spatial imagers could discriminate shape despite changes in texture but not *vice versa*, presumably because their images tend not to encode surface properties. By contrast, object imagers could discriminate texture despite changes in shape, but not the reverse (Lacey et al., 2011), indicating that texture, a surface property, is integrated into their shape representations. Importantly, visual and haptic performance was not significantly different on either

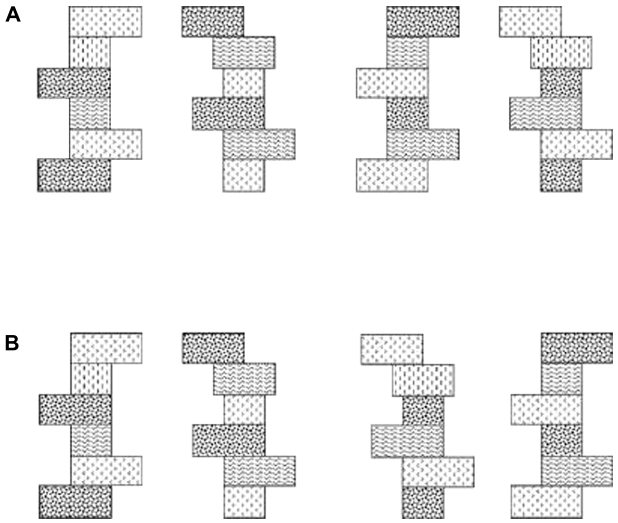


FIGURE 3 | (A) Schematic example of Shapes 1 and 2 with (left pair) original texture schemes and (right pair) the texture schemes exchanged. **(B)** Example of Textures 1 and 2 with (left pair) original shapes and (right pair) the shapes exchanged. Figure adapted from Lacey et al. (2011).

task and performance largely reflected both self-reports of imagery preference and scores on the Object and Spatial Imagery Questionnaire (OSIQ: Blajenkova et al., 2006). Thus, the object-spatial imagery continuum characterizes haptics as well as vision, and individual differences in imagery preference along this continuum affect the extent to which surface properties are integrated into object representations (Lacey et al., 2011). Further analysis of the texture-change condition in our earlier study (Lacey et al., 2010b) showed that performance was indeed related to imagery preference: both object and spatial imagers showed cross-modal view-independence but object imagers were impaired by texture changes whereas spatial imagers were not (Lacey et al., 2011). In addition, the extent of the impairment was correlated with OSIQ scores such that greater preference for object imagery was associated with greater impairment by texture changes; surface properties are therefore likely only integrated into the multisensory representation by object imagers (Lacey et al., 2011). Moreover, spatial imagery preference correlated with the accuracy of cross-modal object recognition (Lacey et al., 2007). It appears, then, that the multisensory representation has some features that are stable across individuals, like view-independence, and some that vary across individuals, such as integration of surface property information and individual differences in imagery preference.

THE NEURAL BASIS OF VISUO-HAPTIC OBJECT PROCESSING

SEGREGATED VENTRAL "WHAT" AND DORSAL "WHERE/HOW" PATHWAYS

At the macro-level, visual object processing divides along a ventral pathway concerned with object identity and perception for recognition, and a dorsal pathway dealing with object location and perception for action, e.g., reaching and grasping, (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992). Similar ventral and dorsal pathways have been proposed for the auditory (e.g., De Santis et al., 2007a) and somatosensory domains (Dijkerman and de Haan, 2007), with divergence of the "what" and "where/how" pathways in a similar timeframe (∼200 ms after stimulus onset) (De Santis et al., 2007a,b), and thus are probably common aspects of functional architecture across modalities.

In the case of touch, an early functional magnetic resonance (fMRI) study found that haptic object recognition activated frontal cortical areas as well as inferior parietal cortex, while a haptic object location task activated superior parietal regions (Reed et al., 2005). A later study from our laboratory (Sathian et al., 2011) compared perception of haptic texture and location, reasoning that texture would be a better marker of haptic object identity, given the salience of texture to touch (Klatzky et al., 1987). This study found that, while both visual and haptic location judgments involved a similar dorsal pathway comprising large sectors of the IPS and frontal eye fields (FEFs) bilaterally, haptic texture perception engaged extensive areas of the parietal operculum (OP), which contains higher-order (i.e., non-primary), ventral regions of somatosensory cortex. In addition, shared cortical processing of texture across vision and touch was found in parts of extrastriate (i.e., non-primary) visual cortex and ventral premotor cortex (Sathian et al., 2011). For both texture and location, several of these bisensory areas showed correlations of activation magnitude between the visual and haptic tasks, indicating some commonality of cortical processing across modalities (Sathian et al., 2011). Another group extended these findings by showing that early visual cortex showed activation magnitudes that not only scaled with the interdot spacing of dot-patterns, but were also modulated by the presence of matching haptic input (Eck et al., 2013).

MULTISENSORY PROCESSING OF OBJECT SHAPE

Cortical areas in both the ventral and dorsal pathways previously identified as specialized for various aspects of visual processing are also functionally involved during the corresponding haptic tasks (for reviews see Amedi et al., 2005; Sathian and Lacey, 2007; Lacey and Sathian, 2011). In the human visual pathway even early visual areas (which project to both dorsal and ventral streams) have been found to respond to changes in haptic shape, suggesting that haptic shape perception might involve the entire ventral stream (Snow et al., 2014). If true, this might reflect cortical pathways between primary somatosensory and visual cortices previously demonstrated in the macaque (Négyessy et al., 2006); however, as with other studies (see below), it is not possible to exclude visual imagery as an explanation for the findings of Snow et al. (2014). The majority of research on visuo-haptic processing of object shape has concentrated on higher-level visual areas, in particular the LOC, an object-selective region in the ventral visual pathway (Malach et al., 1995), a sub-region of which also responds selectively to objects in both vision and touch (Amedi et al., 2001, 2002; Stilla and Sathian, 2008). The LOC responds to both haptic 3-D (Amedi et al., 2001; Zhang et al., 2004; Stilla and Sathian, 2008) and tactile 2-D stimuli (Stoesz et al., 2003; Prather et al., 2004) but does not respond during auditory object recognition cued by object-specific sounds (Amedi et al., 2002). However, when participants listened to the impact sounds made by rods and balls made of either metal or wood and categorized these

sounds by the shape of the object that made them, the material of the object, or by using all the acoustic information available, the LOC was more activated when these sounds were categorized by shape than by material (James et al., 2011). Here again though, participants could have solved this matching task using visual imagery: we return to the potential role of visual imagery in a later section.

The LOC does, however, respond to auditory shape information created by a visual-auditory sensory substitution device (Amedi et al., 2007) using a specific algorithm to convert visual information into an auditory stream or "soundscape" in which the visual horizontal axis is represented by auditory duration and stereo panning, the visual vertical axis by variations in tone frequency, and pixel brightness by variations in tone loudness. Although it requires extensive training, both sighted and blind humans can learn to recognize objects by extracting shape information from such soundscapes (Amedi et al., 2007). However, the LOC only responds to soundscapes created according to the algorithm – and which therefore represent shape in a principled way – and not when participants learn soundscapes that are merely arbitrarily associated with particular objects (Amedi et al., 2007). Thus, the LOC can be regarded as processing geometric shape information independently of the sensory modality used to acquire it.

Apart from the LOC, multisensory (visuo-haptic) responses have also been observed in several parietal regions: in particular, the aIPS is involved in perception of both the shape and location of objects, with co-activation of the LOC for shape and the FEF for location (Stilla and Sathian, 2008; Sathian et al., 2011; see also Saito et al., 2003). The postcentral sulcus (PCS; Stilla and Sathian, 2008), corresponding to Brodmann's area 2 of primary somatosensory cortex (S1; Grefkes et al., 2001), also shows visuo-haptic shape-selectivity. This area is normally considered exclusively somatosensory but the bisensory responses observed by Stilla and Sathian (2008) are consistent with earlier neurophysiological studies that suggested visual responsiveness in parts of S1 (Iriki et al., 1996; Zhou and Fuster, 1997).

Multisensory responses in the LOC and elsewhere might reflect visuo-haptic integration in neurons that process both visual and haptic input; alternatively, they might arise from separate inputs to discrete but interdigitated unisensory neuronal populations. Tal and Amedi (2009) sought to distinguish between these using fMRI adaptation (fMR-A). This technique utilizes the repetition suppression effect, i.e., when the same stimulus is repeated, the blood-oxygen level dependent (BOLD) signal is attenuated. Since repetition suppression can be observed in single neurons, fMR-A can reveal neuronal selectivity profiles (see Grill-Spector et al., 2006; Krekelberg et al., 2006 for reviews). When stimuli that had been presented visually were presented again haptically, there was a robust cross-modal adaptation effect not only in the LOC and the aIPS, but also in bilateral precentral sulcus (preCS) corresponding to ventral premotor cortex, and the right anterior insula, suggesting that these areas were integrating multisensory inputs at the neuronal level. However, a separate preCS site and posterior parts of the IPS did not show cross-modal adaptation, suggesting that their multisensory responses arise from separate unisensory populations. Because fMR-A effects may not necessarily reflect neuronal selectivity (Mur et al., 2010), it will be necessary to confirm the findings of Tal and Amedi (2009) with converging evidence using other methods.

It is critical to determine whether haptic or tactile involvement in supposedly visual cortical areas is functionally relevant, i.e., whether it is actually necessary for task performance. Although research along these lines is still relatively sparse, two lines of evidence indicate that this is indeed the case. Firstly, case studies indicate that the LOC is necessary for both haptic and visual shape perception. A lesion to the left occipito-temporal cortex, which likely included the LOC, resulted in both tactile and visual agnosia even though somatosensory cortex and basic somatosensory function were intact (Feinberg et al., 1986). Another patient with bilateral LOC lesions was unable to learn new objects either visually or haptically (James et al., 2006b). These case studies are consistent with the existence of a shared multisensory representation in the LOC.

Transcranial magnetic stimulation (TMS) is a technique used to temporarily deactivate specific, functionally defined, cortical areas, i.e., to create "virtual lesions" (Sack, 2006). TMS over a parieto-occipital region previously shown to be active during tactile grating orientation discrimination (Sathian et al., 1997) interfered with performance of this task (Zangaladze et al., 1999) indicating that it was functionally, rather than epiphenomenally, involved. This area is the probable human homolog of macaque area V6 (Pitzalis et al., 2006). Repetitive TMS (rTMS) over the left LOC impaired visual object, but not scene, categorization (Mullin and Steeves, 2011), similarly suggesting that this area is necessary for object processing. rTMS over the left aIPS impaired visual-haptic, but not haptic-visual, shape matching using the right hand (Buelte et al., 2008), but shape matching with the left hand during rTMS over the right aIPS was unaffected in either cross-modal condition. The reason for this discrepancy is unclear, and emphasizes that the precise roles of the IPS and LOC in multisensory shape processing have yet to be fully worked out.

CATEGORY-SPECIFIC REPRESENTATIONS

There has been rather limited neural study of cross-modal category-selective representations. Using multivoxel pattern analysis of fMRI data, Pietrini et al. (2004) demonstrated that selectivity for particular categories of man-made objects was correlated across vision and touch in a region of inferotemporal cortex. In the case of face perception, fMRI studies, in contrast to the behavioral studies reviewed above, tend to favor separate, rather than shared representations. For example, visual and haptic face-selectivity in ventral and inferior temporal cortex are in largely separate voxel populations (Pietrini et al., 2004). Haptic face recognition activates the left FG, whereas visual face recognition activates the right FG (Kilgour et al., 2005); furthermore, activity in the left FG increases during haptic processing of familiar, compared to unfamiliar, faces while the right FG remains relatively inactive (James et al., 2006a). A further difference in FG face responses is that imagery of visually presented faces activates the left FG more than the right FG (Ishai et al., 2002)[2]; this raises the possibility that haptic face perception

[2]Note that, although these studies mainly refer to the fusiform gyrus, this is not the only cortical region involved in face processing, nor are faces necessarily the only

involves visual imagery mechanisms. Although one study found that haptic face recognition ability and imagery vividness ratings were uncorrelated (Kilgour and Lederman, 2002), the implication of visual imagery in haptic face perception is very consonant with our findings in haptic shape perception discussed below (Deshpande et al., 2010; Lacey et al., 2010a) especially as vividness ratings do not particularly index imagery ability (reviewed in Lacey and Lawson, 2013). Further studies are needed to resolve the neural basis of multisensory face perception, and its differences from multisensory object perception.

VIEW- AND SIZE-INDEPENDENCE

The cortical locus of the multisensory view-independent representation is currently not known. Evidence for visual view-independence in the LOC is mixed: as might be expected, unfamiliar objects produce view-dependent LOC responses (Gauthier et al., 2002) and familiar objects produce view-independent responses (Valyear et al., 2006; Eger et al., 2008a; Pourtois et al., 2009). By contrast, one study found view-dependence in the LOC even for familiar objects, although in this study there was position-independence (Grill-Spector et al., 1999), whereas another found view-independence for both familiar and unfamiliar objects (James et al., 2002a). A recent TMS study of 2-D shape suggests that the LOC is functionally involved in view-independent recognition (Silvanto et al., 2010) but only two rotations, 20 and 70°, were tested and TMS effects were only seen for the 20° rotation; further work is required to substantiate this finding. Responses in the FG are also variable with the left FG less sensitive to orientation changes than the right FG (Andresen et al., 2009; Harvey and Burgund, 2012). A study of face viewpoint-selectivity showed a gradient of decreasing orientation sensitivity, from view-dependence in early visual cortex to partial view-independence in later areas including LOC (Axelrod and Yovel, 2012); this sensitivity gradient may also apply to non-face objects.

Various parietal regions show visual view-dependent responses, e.g., the IPS (James et al., 2002a) and a parieto-occipital area (Valyear et al., 2006). Superior parietal cortex is view-dependent during mental rotation but not visual object recognition (Gauthier et al., 2002; Wilson and Farah, 2006). As these regions are in the dorsal pathway, concerned with object location and perception for action, view-dependent responses in these regions are not surprising (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992). Actions such as reaching and grasping adapt to changes in object orientation and consistent with this, lateral parieto-occipital cortex shows view-dependent responses for graspable, but not for non-graspable objects (Rice et al., 2007).

To date, we are not aware of neuroimaging studies of haptic or cross-modal processing of stimuli across changes in orientation. James et al. (2002b) varied object orientation, but this study concentrated on haptic-to-visual priming rather than the cross-modal response to same vs. different orientations *per se*. Additionally, there is much work to be done on the effect of orientation changes when shape information is derived from the auditory soundscapes produced by sensory substitution devices (SSDs) and also when the options for haptically interacting with an object are altered by a change in orientation. Similarly, there is no neuroimaging work on haptic and multisensory processing of stimuli across changes in size. However, visual size-independence has been consistently observed in the LOC (Grill-Spector et al., 1999; Ewbank et al., 2005; Eger et al., 2008a,b), with anterior regions showing more size-independence than posterior regions (Sawamura et al., 2005; Eger et al., 2008b).

A MODEL OF VISUO-HAPTIC MULTISENSORY OBJECT REPRESENTATION

Haptic activation of the LOC might arise from direct somatosensory input. Activity in somatosensory cortex propagates to the LOC as early as 150 ms after stimulus onset during tactile discrimination of simple shapes, a timeframe consistent with "bottom-up" projections to LOC (Lucan et al., 2010; Adhikari et al., 2014). Similarly, in a tactile microspatial discrimination task, LOC activity was consistent with feedforward propagation in a beta-band oscillatory network (Adhikari et al., 2014). In addition, a patient with bilateral ventral occipito-temporal lesions, but with sparing of the dorsal part of the LOC that likely included the multisensory sub-region, showed visual agnosia but intact haptic object recognition (Allen and Humphreys, 2009). Haptic object recognition was associated with activation of the intact dorsal part of the LOC, suggesting that somatosensory input could directly activate this region (Allen and Humphreys, 2009).

Alternatively, haptic perception might evoke visual imagery of the felt object resulting in "top-down" activation of the LOC (Sathian et al., 1997) and consistent with this hypothesis, many studies show LOC activity during visual imagery. During auditorily cued mental imagery of familiar object shape, both blind and sighted participants show left LOC activation, where shape information would arise mainly from haptic experience for the blind and mainly from visual experience for the sighted (De Volder et al., 2001). The left LOC is also active when geometric and material object properties are retrieved from memory (Newman et al., 2005) and haptic shape-selective activation magnitudes in the right LOC were highly correlated with ratings of visual imagery vividness (Zhang et al., 2004). A counter-argument is that imagery plays a relatively minor role because LOC activity was substantially lower during visual imagery compared to haptic shape perception (Amedi et al., 2001). However, this study could not verify that participants engaged in imagery throughout the imaging session, so that lower imagery-related activity might have resulted from non-compliance (or irregular compliance) with the task. It has also been argued that visual imagery cannot explain haptically evoked LOC activity because early- as well as late-blind individuals show shape-related LOC activation via both touch (reviewed in Pascual-Leone et al., 2005; Sathian, 2005; Sathian and Lacey, 2007) and hearing using SSDs (Arno et al., 2001; Renier et al., 2004, 2005; Amedi et al., 2007). But this argument, while true for the early blind, does not rule out a visual imagery explanation in the sighted, given the extensive evidence for cross-modal plasticity following visual deprivation (reviewed in Pascual-Leone et al., 2005; Sathian, 2005; Sathian and Lacey, 2007).

category processed in that, or other, regions; this issue remains controversial (see Harel et al., 2013, for a review).

In this section we describe a model of visuo-haptic multisensory object representation (Lacey et al., 2009a) and review the evidence for this model from studies designed to explicitly test the visual imagery hypothesis discussed above (Deshpande et al., 2010; Lacey et al., 2010a, 2014). In this model, object representations in the LOC can be flexibly accessed either bottom-up or top-down, depending on object familiarity, and independently of the input modality. There is no stored representation for unfamiliar objects so that during haptic recognition, an unfamiliar object has to be explored in its entirety in order to compute global shape and to relate component parts to one another. This, we propose, occurs in a bottom-up pathway from somatosensory cortex to the LOC, with involvement of the IPS in computing part relationships and thence global shape, facilitated by spatial imagery processes. For familiar objects, global shape can be inferred more easily, perhaps from distinctive features or one diagnostic part, and we suggest that haptic exploration rapidly acquires enough information to trigger a stored visual image and generate a hypothesis about its identity, as has been proposed for vision (e.g., Bar, 2007). This occurs in a top-down pathway from prefrontal cortex to LOC, involving primarily object imagery processes (though spatial imagery may still have a role in processing familiar objects, for example, in view-independent recognition).

We tested this model using analyses of inter-task correlations of activation magnitude between visual object imagery and haptic shape perception (Lacey et al., 2010a) and analyses of effective connectivity (Deshpande et al., 2010), reasoning that reliance on similar processes across tasks would lead to correlations of activation magnitude across participants, as well as similar patterns of effective connectivity across tasks. In contrast to previous studies, we ensured that participants engaged in visual imagery throughout each scan by using an object imagery task and recording responses. Participants also performed a haptic shape discrimination task using either familiar or unfamiliar objects. We found that object familiarity modulated inter-task correlations as predicted by our model. There were eleven regions common to visual object imagery and haptic perception of familiar shape, six of which (including bilateral LOC) showed inter-task correlations of activation magnitude. By contrast, object imagery and haptic perception of unfamiliar shape shared only four regions, only one of which (an IPS region) showed an inter-task correlation (Lacey et al., 2010a). More recently, we examined the relation between haptic shape perception and spatial imagery, using a spatial imagery task in which participants memorized a 4 × 4 lettered grid and, in response to auditory letter strings, constructed novel shapes within the imagined grid from component parts (Lacey et al., 2014); the haptic shape tasks were the same

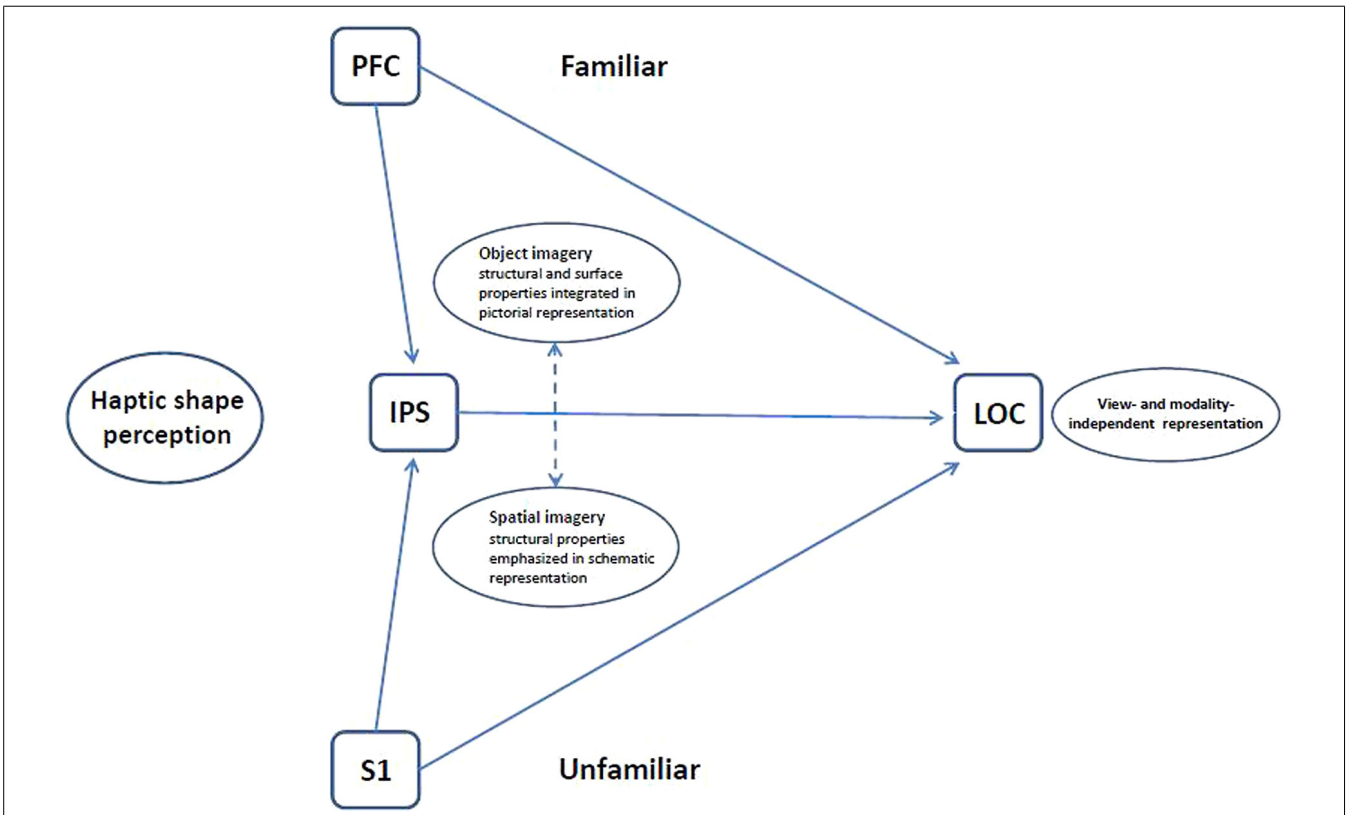


FIGURE 4 | Schematic model of haptic object representation in LOC modulated by object familiarity and imagery type. For unfamiliar more than familiar objects, the LOC is driven bottom-up from somatosensory cortex (S1) with support from spatial imagery processes in the IPS. For familiar more than unfamiliar objects, the LOC is driven top-down from prefrontal cortex (PFC) via object imagery processes. The LOC thus houses an object representation that is flexibly accessible, both bottom-up and top-down, and which is modality- and view-independent (Lacey et al., 2007, 2009b, 2011).

as in Lacey et al. (2010a). Contrary to the model, relatively few regions showed inter-task correlations between spatial imagery and haptic perception of either familiar or unfamiliar shape, with parietal foci featuring in both sets of correlations. This suggests that spatial imagery is relevant to haptic shape perception regardless of object familiarity, whereas our earlier finding suggested that object imagery is more strongly associated with haptic perception of familiar, than unfamiliar, shape (Lacey et al., 2010a). However, it is also possible that the parietal foci showing inter-task correlations between spatial imagery and haptic shape perception reflect spatial processing more generally, rather than spatial imagery *per se* (Lacey et al., 2014; and see Jäncke et al., 2001), or generic imagery processes, e.g., image generation, common to both object and spatial imagery (Lacey et al., 2014; and see Mechelli et al., 2004).

In our study of spatial imagery (Lacey et al., 2014), we also conducted effective connectivity analyses, based on the inferred neuronal activity derived from deconvolving the hemodynamic response out of the observed BOLD signals (Sathian et al., 2013). In order to make direct comparisons between the neural networks underlying object and spatial imagery in haptic shape perception, we re-analyzed our earlier data (Deshpande et al., 2010) using the newer effective connectivity methods. These analyses supported the broad architecture of the model, showing that the spatial imagery network shared much more commonality with the network associated with unfamiliar, compared to familiar, shape perception, while the object imagery network shared much more commonality with familiar, than unfamiliar, shape perception (Lacey et al., 2014). More specifically, the model proposes that the component parts of an unfamiliar object are explored in their entirety and assembled into a representation of global shape via spatial imagery processes (Lacey et al., 2009a). Consistent with this, in the parts of the network that were common to spatial imagery and unfamiliar haptic shape perception, the LOC is driven by parietal foci, with complex cross-talk between posterior parietal and somatosensory foci. These findings fit with the notion of bottom-up pathways from somatosensory cortex and a role for cortex in and around the IPS in spatial imagery (Lacey et al., 2014). The IPS and somatosensory interactions were absent from the sparse network that was shared by spatial imagery and haptic perception of familiar shape. By contrast, the relationship between object imagery and familiar shape perception is characterized by top-down pathways from prefrontal areas reflecting the involvement of object imagery, according to our model (Lacey et al., 2009a). The re-analyzed data supported this, showing the LOC driven bilaterally by the left inferior frontal gyrus in the network shared by object imagery and haptic perception of familiar shape, while these pathways were absent from the extremely sparse network common to object imagery and unfamiliar haptic shape perception (Lacey et al., 2014).

Figure 4 shows the current version of our model for haptic shape perception in which the LOC is driven bottom-up from primary somatosensory cortex as well as top-down via object imagery processes from prefrontal cortex, with additional input from the IPS involving spatial imagery processes. We propose that the bottom-up route is more important for haptic perception of unfamiliar than familiar objects, whereas the converse is true of the top-down route – more important for haptic perception of familiar than unfamiliar objects. It will be interesting to explore the impact of individual preferences for object vs. spatial imagery on these processes and paths.

SUMMARY

The research reviewed here illustrates how deeply interconnected the visual and haptic modalities are in object processing, from highly similar and transferable perceptual spaces underlying categorization, through shared representations in cross-modal and view-independent recognition and commonalities in imagery preferences, to multisensory neural substrates and complex interactions between bottom-up and top-down processes as well as between object and spatial imagery. Much, however, remains to be done in order to provide a detailed account of visuo-haptic multisensory behavior and its underlying mechanisms and how this understanding can be put to use, for example in the service of neurorehabilitation, particularly for those with sensory deprivation of various sorts.

ACKNOWLEDGMENTS

Support to K. Sathian from the National Eye Institute at the NIH, the National Science Foundation, and the Veterans Administration is gratefully acknowledged.

REFERENCES

Adhikari, B. M., Sathian, K., Epstein, C. M., Lamichhane, B., and Dhamala, M. (2014). Oscillatory activity in neocortical networks during tactile discrimination near the limit of spatial acuity. *Neuroimage* 91, 300–310. doi: 10.1016/j.neuroimage.2014.01.007

Allen, H. A., and Humphreys, G. W. (2009). Direct tactile stimulation of dorsal occipito-temporal cortex in a visual agnosic. *Curr. Biol.* 19, 1044–1049. doi: 10.1016/j.cub.2009.04.057

Amedi, A., Jacobson, G., Hendler, T., Malach, R., and Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cereb. Cortex* 12, 1202–1212. doi: 10.1093/cercor/12.11.1202

Amedi, A., Malach, R., Hendler, T., Peled, S., and Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330. doi: 10.1038/85201

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., and Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res.* 166, 559–571. doi: 10.1007/s00221-005-2396-5

Andresen, D. R., Vinberg, J., and Grill-Spector, K. (2009). The representation of object viewpoint in human visual cortex. *Neuroimage* 45, 522–536. doi: 10.1016/j.neuroimage.2008.11.009

Arno, P., De Volder, A. G., Vanlierde, A., Wanet-Defalque, M.-C., Streel, E., Robert, A., et al. (2001). Occipital activation by pattern recognition in the early blind using auditory substitution for vision. *Neuroimage* 13, 632–645. doi: 10.1006/nimg.2000.0731

Axelrod, V., and Yovel, G. (2012). Hierarchical processing of face viewpoint in human visual cortex. *J. Neurosci.* 32, 2442–2452. doi: 10.1523/JNEUROSCI.4770-11.2012

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289. doi: 10.1016/j.tics.2007.05.005

Berryman, L. J., Yau, J. M., and Hsiao, S. S. (2006). Representation of object size in the somatosensory system. *J. Neurophysiol.* 96, 27–39. doi: 10.1152/jn.01190.2005

Biederman, I., and Cooper, E. E. (1992). Size invariance in visual object priming. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 121–133. doi: 10.1037/0096-1523.18.1.121

Blajenkova, O., Kozhevnikov, M., and Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Appl. Cogn. Psychol.* 20, 239–263. doi: 10.1002/acp.1182

Bliss, I., and Hämäläinen, H. (2005). Different working memory capacity in normal young adults for visual and tactile letter recognition task. *Scand. J. Psychol.* 46, 247–251. doi: 10.1111/j.1467-9450.2005.00454.x

Buelte, D., Meister, I. G., Staedtgen, M., Dambeck, N., Sparing, R., Grefkes, C., et al. (2008). The role of the anterior intraparietal sulcus in crossmodal processing of object features in humans: an rTMS study. *Brain Res.* 1217, 110–118. doi: 10.1016/j.brainres.2008.03.075

Bülthoff, I., and Newell, F. N. (2006). The role of familiarity in the recognition of static and dynamic objects. *Prog. Brain Res.* 154, 315–325. doi: 10.1016/S0079-6123(06)54017-8

Bushnell, E. W., and Baxt, C. (1999). Children's haptic and cross-modal recognition with familiar and unfamiliar objects. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1867–1881. doi: 10.1037/0096-1523.25.6.1867

Cant, J. S., Arnott, S. R., and Goodale, M. A. (2009). fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Exp. Brain Res.* 192, 391–405. doi: 10.1007/s00221-008-1573-8

Cant, J. S., and Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex* 17, 713–731. doi: 10.1093/cercor/bhk022

Casey, S. J., and Newell, F. N. (2005). The role of long-term and short-term familiarity in visual and haptic face recognition. *Exp. Brain Res.* 166, 583–591. doi: 10.1007/s00221-005-2398-3

Casey, S. J., and Newell, F. N. (2007). Are representations of unfamiliar faces independent of encoding modality? *Neuropsychologia* 45, 506–513. doi: 10.1016/j.neuropsychologia.2006.02.011

Combe, E., and Wexler, M. (2010). Observer movement and size constancy. *Psychol. Sci.* 21, 667–675. doi: 10.1177/0956797610367753

Cooke, T., Jäkel, F., Wallraven, C., and Bülthoff, H. H. (2007). Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia* 45, 484–495. doi: 10.1016/j.neuropsychologia.2006.02.009

Craddock, M., and Lawson, R. (2008). Repetition priming and the haptic recognition of familiar and unfamiliar objects. *Percept. Psychophys.* 70, 1350–1365. doi: 10.3758/PP.70.7.1350

Craddock, M., and Lawson, R. (2009a). Do left and right matter for haptic recognition of familiar objects? *Perception* 38, 1355–1376. doi: 10.1068/p6312

Craddock, M., and Lawson, R. (2009b). The effect of size changes on haptic object recognition. *Atten. Percept. Psychophys.* 71, 910–923. doi: 10.3758/APP.71.4.910

Craddock, M., and Lawson, R. (2009c). Size-sensitive perceptual representations underlie visual and haptic object recognition. *PLoS ONE* 4:e8009. doi: 10.1371/journal.pone.0008009

Craddock, M., and Lawson, R. (2010). The effects of temporal delay and orientation on haptic object recognition. *Atten. Percept. Psychophys.* 72, 1975–1980. doi: 10.3758/APP.72.7.1975

De Santis, L., Clarke, S., and Murray, M. M. (2007a). Automatic and intrinsic auditory 'what' and 'where' processing in humans revealed by electrical neuroimaging. *Cereb. Cortex* 17, 9–17. doi: 10.1093/cercor/bhj119

De Santis, L., Spierer, L., Clarke, S., and Murray, M. M. (2007b). Getting in touch: segregated somatosensory what and where pathways in humans revealed by electrical neuroimaging. *Neuroimage* 37, 890–903. doi: 10.1016/j.neuroimage.2007.05.052

Deshpande, G., Hu, X., Lacey, S., Stilla, R., and Sathian, K. (2010). Object familiarity modulates effective connectivity during haptic shape perception. *Neuroimage* 49, 1991–2000. doi: 10.1016/j.neuroimage.2009.08.052

De Volder, A. G., Toyama, H., Kimura, Y., Kiyosawa, M., Nakano, H., Vanlierde, A., et al. (2001). Auditory triggered mental imagery of shape involves visual association areas in early blind humans. *Neuroimage* 14, 129–139. doi: 10.1006/nimg.2001.0782

Dijkerman, H. C., and de Haan, E. H. F. (2007). Somatosensory processes subserving perception and action. *Behav. Brain Sci.* 30, 189–239. doi: 10.1017/S0140525X07001392

Dopjans, L., Bülthoff, H. H., and Wallraven, C. (2012). Serial exploration of faces: comparing vision and touch. *J. Vis.* 12, 6. doi: 10.1167/12.1.6

Easton, R. D., Greene, A. J., and Srinivas, K. (1997a). Transfer between vision and haptics: memory for 2-D patterns and 3-D objects. *Psychon. Bull. Rev.* 4, 403–410. doi: 10.3758/BF03210801

Easton, R. D., Srinivas, K., and Greene, A. J. (1997b). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 153–163. doi: 10.1037/0278-7393.23.1.153

Eck, J., Kaas, A. L., and Goebel, R. (2013). Crossmodal interactions of haptic and visual texture information in early sensory cortex. *Neuroimage* 75, 123–135. doi: 10.1016/j.neuroimage.2013.02.075

Eger, E., Ashburner, J., Haynes, J-D., Dolan, R. J., and Rees, G. (2008a). fMRI activity patterns in human LOC carry information about object exemplars within category. *J. Cogn. Neurosci.* 20, 356–370. doi: 10.1162/jocn.2008.20019

Eger, E., Kell, C. A., and Kleinschmidt, A. (2008b). Graded size-sensitivity of object-exemplar-evoked activity patterns within human LOC regions. *J. Neurophysiol.* 100, 2038–2047. doi: 10.1152/jn.90305.2008

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Ewbank, M. P., Schluppeck, D., and Andrews, T. J. (2005). fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *Neuroimage* 28, 268–279. doi: 10.1016/j.neuroimage.2005.06.036

Feinberg, T. E., Rothi, L. J., and Heilman, K. M. (1986). Multimodal agnosia after unilateral left hemisphere lesion. *Neurology* 36, 864–867. doi: 10.1212/WNL.36.6.864

Gaißert, N., Bülthoff, H. H., and Wallraven, C. (2011). Similarity and categorization: from vision to touch. *Acta Psychol.* 138, 219–230. doi: 10.1016/j.actpsy.2011.06.007

Gaißert, N., and Wallraven, C. (2012). Categorizing natural objects: a comparison of the visual and haptic modalities. *Exp. Brain Res.* 216, 123–134. doi: 10.1007/s00221-011-2916-4

Gaißert, N., Wallraven, C., and Bülthoff, H. H. (2008). "Analyzing perceptual representations of complex, parametrically-defined shapes using MDS," in *Proceedings of the Sixth International Conference EuroHaptics 2008, Lecture Notes in Computer Science – Haptics: Perception, Devices and Scenarios, Vol. 5024* (Heidelberg: Springer Berlin Heidelberg), 265–274.

Gaißert, N., Wallraven, C., and Bülthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *J. Vis.* 10, 2. doi: 10.1167/10.11.2

Gaißert, N., Waterkamp, S., Fleming, R. W., and Bülthoff, I. (2012). Haptic categorical perception of shape. *PLoS ONE* 7:e43062. doi: 10.1371/journal.pone.0043062

Gallace, A. (2013). "Somesthetic mental imagery," in *Multisensory Imagery*, eds S. Lacey and R. Lawson (New York: Springer), 29–50. doi: 10.1007/978-1-4614-5879-1_3

Gallace, A., and Spence, C. (2009). The cognitive and neural correlates of tactile memory. *Psychol. Bull.* 135, 380–406. doi: 10.1037/a0015325

Garvill, J., and Molander, B. (1973). Effects of standard modality, comparison modality and retention interval on matching of form. *Scand. J. Psychol.* 14, 203–206. doi: 10.1111/j.1467-9450.1973.tb00111.x

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (2002). BOLD activity during mental rotation and view-dependent object recognition. *Neuron* 34, 161–171. doi: 10.1016/S0896-6273(02)00622-0

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Graf, M. (2010). "Categorization and object shape," in *Towards a Theory of Thinking: Building Blocks for a Conceptual Framework*, eds B. M. Glatzeder, V. Goel, and A. von Müller (Berlin: Springer-Verlag), 73–101. doi: 10.1007/978-3-642-03129-8_6

Grefkes, C., Geyer, S., Schormann, T., Roland, P., and Zilles, K. (2001). Human somatosensory area 2: observer-independent cytoarchitectonic mapping, interindividual variability, and population map. *Neuroimage* 14, 617–631. doi: 10.1006/nimg.2001.0858

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10, 14–23. doi: 10.1016/j.tics.2005.11.006

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203. doi: 10.1016/S0896-6273(00)80832-6

Haag, S. (2011). Effects of vision and haptics on categorizing common objects. *Cogn. Process.* 12, 33–39. doi: 10.1007/s10339-010-0369-5

Harel, A., Kravitz, D., and Baker, C. I. (2013). Beyond perceptual expertise: revisiting the neural substrates of expert object recognition. *Front. Hum. Neurosci.* 7:885. doi: 10.3389/fnhum.2013.00885

Harvey, D. Y., and Burgund, E. D. (2012). Neural adaptation across viewpoint and exemplar in fusiform cortex. *Brain Cogn.* 80, 33–44. doi: 10.1016/j.bandc.2012.04.009

Helbig, H. B., and Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Exp. Brain Res.* 179, 595–606. doi: 10.1007/s00221-006-0814-y

Helbig, H. B., Ernst, M. O., Ricciardi, E., Pietrini, P., Thielscher, A., Mayer, K. M., et al. (2012). The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage* 60, 1063–1072. doi: 10.1016/j.neuroimage.2011.09.072

Iriki, A., Tanaka, M., and Iwamura, Y. (1996). Attention-induced neuronal activity in the monkey somatosensory cortex revealed by pupillometrics. *Neurosci. Res.* 25, 173–181. doi: 10.1016/0168-0102(96)01043-7

Ishai, A., Haxby, J. V., and Ungerleider, L. G. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage* 17, 1729–1741. doi: 10.1006/nimg.2002.1330

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., and Goodale, M. A. (2002a). Differential effects of view on object-driven activation in dorsal and ventral streams. *Neuron* 35, 793–801. doi: 10.1016/S0896-6273(02)00803-6

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., and Goodale, M. A. (2002b). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40, 1706–1714. doi: 10.1016/S0028-3932(02)00017-9

James, T. W., Servos, P., Kilgour, A. R., Huh, E., and Lederman, S. (2006a). The influence of familiarity on brain activation during haptic exploration of 3-D facemasks. *Neurosci. Lett.* 397, 269–273. doi: 10.1016/j.neulet.2005.12.052

James, T. W., James, K. H., Humphrey, G. K., and Goodale, M. A. (2006b). "Do visual and tactile object representations share the same neural substrate?" in *Touch and Blindness: Psychology and Neuroscience*, eds M. A. Heller and S. Ballesteros (Mahwah, NJ: Lawrence Erlbaum Associates), 139–155.

James, T. W., Stevenson, R. W., Kim, S., VanDerKlok, R. M., and James, K. H. (2011). Shape from sound: evidence for a shape operator in the lateral occipital cortex. *Neuropsychologia* 49, 1807–1815. doi: 10.1016/j.neuropsychologia.2011.03.004

Jäncke, L., Kleinschmidt, A., Mirzazade, S., Shah, N. J., and Freund, H.-J. (2001). The role of the inferior parietal cortex in linking the tactile perception and manual construction of object shapes. *Cereb. Cortex* 11, 114–121. doi: 10.1093/cercor/11.2.114

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Mem. Cogn.* 15, 531–543. doi: 10.3758/BF03198388

Jones, B. (1981). "The developmental significance of cross-modal matching," in *Intersensory Perception and Sensory Integration,* eds R. D. Walk and H. L. Jr. Pick, (New York: Plenum Press), 108–136.

Kassuba, T., Klinge, C., Hölig, C., Röder, B., and Siebner, H. R. (2013). Vision holds a greater share in visuo-haptic object recognition than touch. *Neuroimage* 65, 59–68. doi: 10.1016/j.neuroimage.2012.09.054

Kilgour, A. R., de Gelder, B., and Lederman, S. (2004). Haptic face recognition and prosopagnosia. *Neuropsychologia* 42, 707–712. doi: 10.1016/j.neuropsychologia.2003.11.021

Kilgour, A. R., Kitada, R., Servos, P., James, T. W., and Lederman, S. J. (2005). Haptic face identification activates ventral occipital and temporal areas: an fMRI study. *Brain Cogn.* 59, 246–257. doi: 10.1016/j.bandc.2005.07.004

Kilgour, A. R., and Lederman, S. (2002). Face recognition by hand. *Percept. Psychophys.* 64, 339–352. doi: 10.3758/BF03194708

Kilgour, A. R., and Lederman, S. (2006). A haptic face-inversion effect. *Perception* 35, 921–931. doi: 10.1068/p5341

Kiphart, M. J., Hughes, J. L., Simmons, J. P., and Cross, H. A. (1992). Short-term haptic memory for complex objects. *Bull. Psychon. Soc.* 30, 212–214. doi: 10.3758/BF03330444

Klatzky, R. L., Abramowicz, A., Hamilton, C., and Lederman, S. J. (2011). Irrelevant visual faces influence haptic identification of facial expressions of emotion. *Atten. Percept. Psychophys.* 73, 521–530. doi: 10.3758/s13414-010-0038-x

Klatzky, R. J., and Lederman, S. J. (1992). Stages of manual exploration in haptic object identification. *Percept. Psychophys.* 52, 661–670. doi: 10.3758/BF03211702

Klatzky, R. L., and Lederman, S. J. (1995). Identifying objects from a haptic glance. *Percept. Psychophys.* 57, 1111–1123. doi: 10.3758/BF03208368

Klatzky, R. L., Lederman, S. J., and Metzger, V. A. (1985). Identifying objects by touch: an 'expert system'. *Percept. Psychophys.* 37, 299–302. doi: 10.3758/BF03211351

Klatzky, R. L., Lederman, S. J., and Reed, C. L. (1987). There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *J. Exp. Psychol. Gen.* 116, 356–369. doi: 10.1037/0096-3445.116.4.356

Konkle, T., and Oliva, A. (2011). Canonical visual size for real-world objects. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 23–37. doi: 10.1037/a0020413

Kozhevnikov, M., Hegarty, M., and Mayer, R. E. (2002). Revising the visualiser-verbaliser dimension: evidence for two types of visualisers. *Cogn. Instr.* 20, 47–77. doi: 10.1207/S1532690XCI2001_3

Kozhevnikov, M., Kosslyn, S. M., and Shephard, J. (2005). Spatial versus object visualisers: a new characterisation of cognitive style. *Mem. Cogn.* 33, 710–726. doi: 10.3758/BF03195337

Krekelberg, B., Boynton, G. M., and van Wezel, R. J. A. (2006). Adaptation: from single cells to BOLD signals. *Trends Neurosci.* 29, 250–256. doi: 10.1016/j.tins.2006.02.008

Lacey, S., and Campbell, C. (2006). Mental representation in visual/haptic cross-modal memory: evidence from interference effects. *Q. J. Exp. Psychol.* 59, 361–376. doi: 10.1080/17470210500173232

Lacey, S., Flueckiger, P., Stilla, R., Lava, M., and Sathian, K. (2010a). Object familiarity modulates the relationship between visual object imagery and haptic shape perception. *Neuroimage* 49, 1977–1990. doi: 10.1016/j.neuroimage.2009.10.081

Lacey, S., Hall, J., and Sathian, K. (2010b). Are surface properties integrated into visuo-haptic object representations? *Eur. J. Neurosci.* 31, 1882–1888. doi: 10.1111/j.1460-9568.2010.07204.x

Lacey, S., and Lawson, R. (2013). "Imagery questionnaires: vividness and beyond," in *Multisensory Imagery,* eds S. Lacey and R. Lawson (New York: Springer), 271–282.

Lacey, S., Lin, J. B., and Sathian, K. (2011). Object and spatial imagery dimensions in visuo-haptic representations. *Exp. Brain Res.* 213, 267–273. doi: 10.1007/s00221-011-2623-1

Lacey, S., Peters, A., and Sathian, K. (2007). Cross-modal object representation is viewpoint-independent. *PLoS ONE* 2:e890. doi: 10.1371/journal.pone0000890

Lacey, S., and Sathian, K. (2011). Multisensory object representation: insights from studies of vision and touch. *Prog. Brain Res.* 191, 165–176. doi: 10.1016/B978-0-444-53752-2.00006-0

Lacey, S., Stilla, R., Sreenivasan, K., Deshpande, G., and Sathian, K. (2014). Spatial imagery in haptic shape perception. *Neuropsychologia* doi: 10.1016/j.neuropsycholopia.2014.05.008 [Epub ahead of print].

Lacey, S., Tal, N., Amedi, A., and Sathian, K. (2009a). A putative model of multisensory object representation. *Brain Topogr.* 21, 269–274. doi: 10.1007/s10548-009-0087-4

Lacey, S., Pappas, M., Kreps, A., Lee, K., and Sathian, K. (2009b). Perceptual learning of view-independence in visuo-haptic object representations. *Exp. Brain Res.* 198, 329–337. doi: 10.1007/s00221-009-1856-8

Lawson, R. (2009). A comparison of the effects of depth rotation on visual and haptic three-dimensional object recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 911–930. doi: 10.1037/a0015025

Lawson, R. (2011). An investigation into the cause of orientation-sensitivity in haptic object recognition. *Seeing Perceiving* 24, 293–314. doi: 10.1163/187847511X579052

Lawson, R. (2014). Recognizing familiar objects by hand and foot: haptic shape perception generalizes to inputs from unusual locations and untrained body parts. *Atten. Percept. Psychophys.* 76, 541–558. doi: 10.3758/s13414-013-0559-1

Lederman, S. J., and Klatzky, R. L. (1987). Hand movements: a window into haptic object recognition. *Cogn. Psychol.* 19, 342–368. doi: 10.1016/0010-0285(87)90008-9

Loomis, J., Klatzky, R. L., and Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. *Perception* 20, 167–177. doi: 10.1068/p200167

Lucan, J. N., Foxe, J. J., Gomez-Ramirez, M., Sathian, K., and Molholm, S. (2010). Tactile shape discrimination recruits human lateral occipital complex during early perceptual processing. *Hum. Brain Mapp.* 31, 1813–1821. doi: 10.1002/hbm.20983

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance

imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8135–8139. doi: 10.1073/pnas.92.18.8135

Marr, D. (1976). Early processing of visual information. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 275, 483–524. doi: 10.1098/rstb.1976.0090

Matsumiya, K. (2012). Haptic face aftereffect. *Iperception* 3, 97–100.

Matsumiya, K. (2013). Seeing a haptically explored face: visual facial-expression aftereffect from haptic adaptation to a face. *Psychol. Sci.* 24, 2088–2098. doi: 10.1177/0956797613486981

Maurer, D., Le Grand, R., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4

Mechelli, A., Price, C. J., Friston, K. J., and Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cereb. Cortex* 14, 1256–1265. doi: 10.1093/cercor/bhh087

Mullin, C. R., and Steeves, J. K. E. (2011). TMS to the lateral occipital cortex disrupts object processing but facilitates scene processing. *J. Cogn. Neurosci.* 23, 4174–4184. doi: 10.1162/jocn_a_00095

Mur, M., Ruff, D. A., Bodurka, J., Bandettini, P. A., and Kriegeskorte, N. (2010). Face-identity change activation outside the face system: "release from adaptation" may not always indicate neuronal selectivity. *Cereb. Cortex* 20, 2027–2042. doi: 10.1093/cercor/bhp272

Nabeta, T., and Kawahara, J. (2006). Congruency effect of presentation modality on false recognition of haptic and visual objects. *Memory* 14, 307–315. doi: 10.1080/09658210500277398

Nassi, J. J., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372. doi: 10.1038/nrn2619

Négyessy, L., Nepusz, T., Kocsis, L., and Bazsó, F. (2006). Prediction of the main cortical areas and connections involved in the tactile function of the visual cortex by network analysis. *Eur. J. Neurosci.* 23, 1919–1930. doi: 10.1111/j.1460-9568.2006.04678.x

Newell, F. N., Ernst, M. O., Tjan, B. S., and Bülthoff, H. H. (2001). View dependence in visual and haptic object recognition. *Psychol. Sci.* 12, 37–42. doi: 10.1111/1467-9280.00307

Newman, S. D., Klatzky, R. L., Lederman, S. J., and Just, M. A. (2005). Imagining material versus geometric properties of objects: an fMRI study. *Cogn. Brain Res.* 23, 235–246. doi: 10.1016/j.cogbrainres.2004.10.020

Nicholson, K. G., and Humphrey, G. K. (2003). The effect of colour congruency on shape discriminations of novel objects. *Perception* 32, 339–353. doi: 10.1002/hbm.20983

Palmer, S. E., Rosch, E., and Chase, P. (1981). "Canonical perspective and the perception of objects," in *Attention and Performance IX*, eds J. Long and A. Baddeley (Hillsdale, NJ: Lawrence Erlbaum Associates), 135–151.

Pascual-Leone, A., Amedi, A., Fregni, F., and Merabet, L. B. (2005). The plastic human brain. *Annu. Rev. Neurosci.* 28, 377–401. doi: 10.1146/annurev.neuro.27.070203.144216

Pascual-Leone, A., and Hamilton, R. H. (2001). The metamodal organization of the brain. *Prog. Brain Res.* 134, 427–445. doi: 10.1016/S0079-6123(01)34028-1

Pasqualotto, A., Finucane, C., and Newell, F. N. (2005). Visual and haptic representations of scenes are updated with observer movement. *Exp. Brain Res.* 166, 481–488. doi: 10.1007/s00221-005-2388-5

Pasqualotto, A., and Newell, F. N. (2007). The role of visual experience on the representation and updating of novel haptic scenes. *Brain Cogn.* 65, 184–194. doi: 10.1016/j.bandc.2007.07.009

Peissig, J. J., and Tarr, M. J. (2007). Visual object recognition: do we know more now than we did 20 years ago? *Annu. Rev. Psychol.* 58, 75–96. doi: 10.1146/annurev.psych.58.102904.190114

Pensky, A. E. C., Johnson, K. A., Haag, S., and Homa, D. (2008). Delayed memory for visual-haptic exploration of objects. *Psychon. Bull. Rev.* 15, 574–580. doi: 10.3758/PBR.15.3.574

Peterson, M. A., and Rhodes, G. (2003). *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes.* Oxford: Oxford University Press.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., et al. (2004). Beyond sensory images: object-based representation in the human ventral pathway. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5658–5663. doi: 10.1073/pnas.0400707101

Pitzalis, S., Galletti, C., Huang, R. S., Patria, F., Committeri, G., Galati, G., et al. (2006). Wide-field retinotopy defines human cortical visual area V6. *J. Neurosci.* 26, 7962–7973. doi: 10.1523/JNEUROSCI.0178-06.2006

Pourtois, G., Schwarz, S., Spiridon, M., Martuzzi, R., and Vuilleumier, P. (2009). Object representations for multiple visual categories overlap in lateral occipital and medial fusiform cortex. *Cereb. Cortex* 19, 1806–1819. doi: 10.1093/cercor/bhn210

Prather, S. C., Votaw, J. R., and Sathian, K. (2004). Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia* 42, 1079–1087. doi: 10.1016/j.neuropsychologia.2003.12.013

Reales, J. M., and Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: cross-modal priming depends on structural descriptions. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 644–663. doi: 10.1037/0278-7393.25.3.644

Reed, C. L., Klatzky, R. L., and Halgren, E. (2005). What vs. where in touch: an fMRI study. *Neuroimage* 25, 718–726. doi: 10.1016/j.neuroimage.2004.11.044

Renier, L., Collignon, O., Tranduy, D., Poirier, C., Vanlierde, A., Veraart, C., et al. (2004). Visual cortex activation in early blind and sighted subjects using an auditory visual substitution device to perceive depth. *Neuroimage* 22, S1.

Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., et al. (2005). Cross modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage* 26, 573–580. doi: 10.1016/j.neuroimage.2005.01.047

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, T87–T93. doi: 10.1016/j.neuroimage.2007.03.032

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Sack, A. T. (2006). Transcranial magnetic stimulation, causal structure-function mapping and networks of functional relevance. *Curr. Opin. Neurobiol.* 16, 593–599. doi: 10.1016/j.conb.2006.06.016

Saito, D. N., Okada, T., Morita, Y., Yonekura, Y., and Sadato, N. (2003). Tactile-visual cross-modal shape matching: a functional MRI study. *Cogn. Brain Res.* 17, 14–25. doi: 10.1016/S0926-6410(03)00076-4

Sathian, K. (2005). Visual cortical activity during tactile perception in the sighted and the visually deprived. *Dev. Psychobiol.* 46, 279–286. doi: 10.1002/dev.20056

Sathian, K., Deshpande, G., and Stilla, R. (2013). Neural changes with tactile learning reflect decision-level reweighting of perceptual readout. *J. Neurosci.* 33, 5387–5398. doi: 10.1523/JNEUROSCI.3482-12.2013

Sathian, K., and Lacey, S. (2007). Journeying beyond classical somatosensory cortex. *Can. J. Exp. Psychol.* 61, 254–264. doi: 10.1037/cjep2007026

Sathian, K., Lacey, S., Stilla, R., Gibson, G., Deshpande, G., Hu, X., et al. (2011). Dual pathways for somatosensory information processing. *Neuroimage* 57, 462–475. doi: 10.1016/j.neuroimage.2011.05.001

Sathian, K., Zangaladze, A., Hoffman, J. M., and Grafton, S. T. (1997). Feeling with the mind's eye. *Neuroreport* 8, 3877–3881. doi: 10.1097/00001756-199712220-00008

Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., and Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J. Neurosci.* 25, 4294–4306. doi: 10.1523/JNEUROSCI.0377-05.2005

Silvanto, J., Schwarzkopf, D. S., Gilaie-Dotan, S., and Rees, G. (2010). Differing causal roles for lateral occipital complex and occipital face area in invariant shape recognition. *Eur. J. Neurosci.* 32, 165–171. doi: 10.1111/j.1460-9568.2010.07278.x

Skinner, A. L., and Benton, C. P. (2010). Anti-expression aftereffects reveal prototype-referenced coding of facial expressions. *Psychol. Sci.* 21, 1248–1253. doi: 10.1177/0956797610380702

Smith, E. E., Patalano, A. L., and Jonides, J. (1998). Alternative strategies of categorization. *Cognition* 65, 167–196. doi: 10.1016/S0010-0277(97)00043-7

Snow, J. C., Strother, L., and Humphreys, G. W. (2014). Haptic shape processing in visual cortex. *J. Cogn. Neurosci.* 26, 1154–1167. doi: 10.1162/jocn_a_00548

Stilla, R., and Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Hum. Brain Mapp.* 29, 1123–1138. doi: 10.1002/hbm.20456

Stoesz, M., Zhang, M., Weisser, V. D., Prather, S. C., Mao, H., and Sathian, K. (2003). Neural networks active during tactile form perception: common and differential activity during macrospatial and microspatial tasks. *Int. J. Psychophysiol.* 50, 41–49. doi: 10.1016/S0167-8760(03)00123-5

Streri, A., and Molina, M. (1994). "Constraints on intermodal transfer between touch and vision in infancy," in *The Development of Intersensory Perception: Comparative Perspectives*, eds D. J. Lewkowicz and R. Lickliter (Hove: Lawrence Erlbaum Associates), 285–307.

Takahashi, C., and Watt, S. J. (2014). Visual-haptic integration with pliers and tongs: signal 'weights' take account of changes in haptic sensitivity caused by different tools. *Front. Psychol.* 5:109. doi: 10.3389/fpsyg.2014.00109

Tal, N., and Amedi, A. (2009). Multisensory visual-tactile object related network in humans: insights gained using a novel crossmodal adaptation approach. *Exp. Brain Res.* 198, 165–182. doi: 10.1007/s00221-009-1949-4

Tarr, M. J., and Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cogn. Psychol.* 21, 233–282. doi: 10.1016/0010-0285(89)90009-1

Ueda, Y., and Saiki, J. (2007). Viewpoint independence in visual and haptic object recognition. *Jpn. J. Psychon. Sci.* 26, 11–19.

Ueda, Y., and Saiki, J. (2012). Characteristics of eye movements in 3-D object learning: comparison between within-modal and cross-modal object recognition. *Perception* 41, 1289–1298. doi: 10.1068/p7257

Ungerleider, L. G., and Mishkin, M. (1982). "Two cortical visual systems," in *Analysis of Visual Behavior*, eds D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Cambridge, MA: MIT Press), 549–586.

Uttl, B., Graf, P., and Siegenthaler, A. L. (2007). Influence of object size on baseline identification, priming, and explicit memory. *Scand. J. Psychol.* 48, 281–288. doi: 10.1111/j.1467-9450.2007.00571.x

Valyear, K. F., Culham, J. C., Sharif, N., Westwood, D., and Goodale, M. A. (2006). A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal streams: a human fMRI study. *Neuropsychologia* 44, 218–228. doi: 10.1016/j.neuropsychologia.2005.05.004

Wallraven, C. (2014). Touching on face space: comparing visual, and haptic processing of face shapes. *Psychon. Bull. Rev.* doi: 10.3758/s13423-013-0577-y [Epub ahead of print].

Wallraven, C., Bülthoff, H. H., Waterkamp, S., van Dam, L., and Gaiβert, N. (2013). The eyes grasp, the hands see: metric category knowledge transfers between vision and touch. *Psychon. Bull. Rev.* doi: 10.3758/s13423-013-0563-4 [Epub ahead of print].

Wallraven, C., and Dopjans, L. (2013). Visual experience is necessary for efficient haptic face recognition. *Neuroreport* 24, 254–258. doi: 10.1097/WNR.0b013e32835f00c0

Wilson, K., and Farah, M. J. (2006). Distinct patterns of viewpoint-dependent BOLD activity during common-object recognition and mental rotation. *Perception* 35, 1351–1366. doi: 10.1068/p5571

Woods, A. T., Moore, A., and Newell, F. N. (2008). Canonical views in haptic object perception. *Perception* 37, 1867–1878. doi: 10.1068/p6038

Woods, A. T., O'Modhrain, S., and Newell, F. N. (2004). The effect of temporal delay and spatial differences on crossmodal object recognition. *Cogn. Affect. Behav. Neurosci.* 4, 260–269. doi: 10.3758/CABN.4.2.260

Yildirim, I., and Jacobs, R. A. (2013). Transfer of object category knowledge across visual and haptic modalities: experimental and computational studies. *Cognition* 126, 135–148. doi: 10.1016/j.cognition.2012.08.005

Yin, R. K. (1969). Looking at upside-down faces. *J. Exp. Psychol.* 81, 141–145. doi: 10.1037/h0027474

Zangaladze, A., Epstein, C. M., Grafton, S. T., and Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature* 401, 587–590. doi: 10.1038/44139

Zhang, M., Weisser, V. D., Stilla, R., Prather, S. C., and Sathian, K. (2004). Multisensory cortical processing of object shape and its relation to mental imagery. *Cogn. Affect. Behav. Neurosci.* 4, 251–259. doi: 10.3758/CABN.4.2.251

Zhou, Y.-D., and Fuster, J. M. (1997). Neuronal activity of somatosensory cortex in a cross-modal (visuo-haptic) memory task. *Exp. Brain Res.* 116, 551–555. doi: 10.1007/PL00005783

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: 17 February 2014; accepted: 23 June 2014; published online: 17 July 2014.
*Citation: Lacey S and Sathian K (2014) Visuo-haptic multisensory object recognition, categorization, and representation. Front. Psychol. **5**:730. doi: 10.3389/fpsyg.2014.00730*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Lacey and Sathian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 02 April 2014
doi: 10.3389/fpsyg.2014.00274

Short-term plasticity of visuo-haptic object recognition

Tanja Kassuba[1,2,3]*, Corinna Klinge[2,4], Cordula Hölig[2,5], Brigitte Röder[5] and Hartwig R. Siebner[1,2,3]

[1] Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark
[2] NeuroImageNord/Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3] Department of Neurology, Christian-Albrechts-University, Kiel, Germany
[4] Department of Psychiatry, Warneford Hospital, Oxford, UK
[5] Biological Psychology and Neuropsychology, University of Hamburg, Hamburg, Germany

Edited by:
Chris Fields, Retired, USA

Reviewed by:
Ryo Kitada, National Institute for Physiological Sciences, Japan
Jacqueline Clare Snow, University of Nevada, USA

***Correspondence:**
Tanja Kassuba, Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
e-mail: kassuba@princeton.edu

Functional magnetic resonance imaging (fMRI) studies have provided ample evidence for the involvement of the lateral occipital cortex (LO), fusiform gyrus (FG), and intraparietal sulcus (IPS) in visuo-haptic object integration. Here we applied 30 min of sham (non-effective) or real offline 1 Hz repetitive transcranial magnetic stimulation (rTMS) to perturb neural processing in left LO immediately before subjects performed a visuo-haptic delayed-match-to-sample task during fMRI. In this task, subjects had to match sample (S1) and target (S2) objects presented sequentially within or across vision and/or haptics in both directions (visual-haptic or haptic-visual) and decide whether or not S1 and S2 were the same objects. Real rTMS transiently decreased activity at the site of stimulation and remote regions such as the right LO and bilateral FG during haptic S1 processing. Without affecting behavior, the same stimulation gave rise to relative increases in activation during S2 processing in the right LO, left FG, bilateral IPS, and other regions previously associated with object recognition. Critically, the modality of S2 determined which regions were recruited after rTMS. Relative to sham rTMS, real rTMS induced increased activations during crossmodal congruent matching in the left FG for haptic S2 and the temporal pole for visual S2. In addition, we found stronger activations for incongruent than congruent matching in the right anterior parahippocampus and middle frontal gyrus for crossmodal matching of haptic S2 and in the left FG and bilateral IPS for unimodal matching of visual S2, only after real but not sham rTMS. The results imply that a focal perturbation of the left LO triggers modality-specific interactions between the stimulated left LO and other key regions of object processing possibly to maintain unimpaired object recognition. This suggests that visual and haptic processing engage partially distinct brain networks during visuo-haptic object matching.

Keywords: multisensory interactions, visual perception, haptic perception, object recognition, repetitive transcranial magnetic stimulation, functional magnetic resonance imaging

INTRODUCTION

An object's geometrical structure (shape) and surface can be extracted by both using vision and haptics. Integrating shape information across senses can facilitate object recognition (Stein and Stanford, 2008). In vision, the lateral occipital complex (LOC), consisting of subregions in the lateral occipital cortex (LO) and in the fusiform gyrus (FG) (Malach et al., 2002), has long been known to show a preferential response to images of objects as opposed to their scrambled counterparts or other textures (Malach et al., 1995; Grill-Spector et al., 1999; Kourtzi and Kanwisher, 2001). Subsequent neuroimaging studies and studies using transcranial magnetic stimulation (TMS) have linked object- or shape-specific brain responses in the LOC to individual performance during visual object recognition (Grill-Spector et al., 2000; Bar et al., 2001; Ellison and Cowey, 2006; Williams et al., 2007; Pitcher et al., 2009). The functional relevance of the LOC has been further substantiated by patients with lesions in the occipito-temporal cortex suffering from visual agnosia, that is, a severe deficit in visually recognizing objects despite otherwise intact intelligence (Goodale et al., 1991; Karnath et al., 2009). Object-specific responses in the LOC, particularly in the left LO, have also been found when comparing brain responses during the haptic exploration of objects and texture stimuli (Amedi et al., 2001, 2002; Kassuba et al., 2011) or when testing for haptic shape adaptation (Snow et al., 2013). Accordingly, lesions in occipito-temporal cortex can lead to haptic object agnosia (Morin et al., 1984; Feinberg et al., 1986) but see (Snow et al., 2012). Since both vision and haptics provide shape information, it has been proposed that the left LO comprises multisensory representations of object shape that are accessed by the different senses (Amedi et al., 2002). Accordingly, the left LO is typically not recruited by auditory object stimuli which do not provide any shape information unless subjects have learned to extract shape information from soundscapes produced by visual-to-auditory sensory substitution devices (Amedi et al., 2002, 2007).

Previous studies had neglected potential intrinsic differences in the relative contributions of vision and haptics to visuo-haptic shape or object recognition. Since vision provides information

about several object features in parallel and even if the object is outside the reaching space, there might be an overall dominance of vision in object recognition, at least if objects have to be recognized predominantly based on their shape. In line with this notion, we have recently found an asymmetry in the processing of crossmodal information during visual and haptic object recognition (Kassuba et al., 2013a). Using a visuo-haptic delayed-match-to-sample task during functional magnetic resonance imaging (fMRI), the direction of delayed matching (visual-haptic vs. haptic-visual) influenced the activation profiles in bilateral LO, FG, anterior (aIPS) and posterior intraparietal sulcus (pIPS), that is, in regions which have previously been associated with visuo-haptic object integration (Grefkes et al., 2002; Saito et al., 2003; Stilla and Sathian, 2008; Kassuba et al., 2011; for review see Lacey and Sathian, 2011). Only when a haptic target was matched to a previously presented visual sample but not in the reverse order (i.e., when a visual target was matched to a haptic sample) we found activation profiles in these regions suggesting multisensory interactions. In line with the maximum likelihood account of multisensory integration (Ernst and Banks, 2002; Helbig and Ernst, 2007), we attributed this asymmetry to the fact that haptic exploration is less efficient than vision when recognizing objects based on their shape (given highly reliable input from both modalities) and gains more from integrating additional crossmodal information than vision.

To further explore the role of left LO in visuo-haptic object integration, we here examined how repetitive TMS (rTMS) of the left LO affects crossmodal object matching. Specifically, we applied real or sham (non-effective) offline 1 Hz rTMS to the left LO immediately before subjects performed a visuo-haptic delayed-match-to-sample task during fMRI. The published results reported above (Kassuba et al., 2013a) present the results after sham rTMS, the current paper focuses on how these multisensory interaction effects were modulated by real rTMS. During fMRI, a visual or haptic sample object (S1) and a visual or haptic target object (S2) were presented sequentially, and subjects had to indicate whether the identity of both objects was the same (congruent) or not (incongruent). Thus, the event of matching (processing S2 and matching it to previously presented S1) was manipulated by three orthogonal factors: (1) S1 and S2 were from the same (unimodal) or different modalities (crossmodal), (2) their identity was congruent or incongruent, and (3) S2 was presented either in the visual or the haptic modality. We assumed that crossmodal integration occurs only when the visual and haptic object inputs are semantically congruent (Laurienti et al., 2004). Multisensory interactions were defined as an increased activation during crossmodal vs. unimodal matching (crossmodal matching effects, cf. Grefkes et al., 2002) that were stronger for congruent than incongruent object pairs (crossmodal matching by semantic congruency interaction; Kassuba et al., 2013a,b). The rTMS-induced changes in task-related activity were investigated with blood-oxygenated-level-dependent (BOLD) fMRI. We hypothesized that real compared to sham rTMS of the left LO would trigger compensatory increases in activity not only at the site of stimulation but additionally in remote key regions of visuo-haptic object integration such as the right LO, bilateral FG, and IPS. Based on previous work (Kassuba et al., 2013a), we predicted that real rTMS would particularly affect visuo-haptic interactions during matching of haptic as opposed to visual S2.

MATERIALS AND METHODS

SUBJECTS

The description of the subjects is reproduced from (Kassuba et al., 2013a: Participants, p. 60) and adjusted to include rTMS-specific information. Nineteen healthy right-handed volunteers took part in this study. In one female subject, real rTMS caused uncomfortable sensations on her skull and the experiment was aborted. Data acquisition was successfully completed in 18 participants (9 females, 22–33 years of age, average 25.72 $\pm$ 2.87). All subjects reported normal or corrected-to-normal vision, normal tactile and hearing ability, and none had a history of psychiatric or neurological disorders. Handedness was assessed with the short form of the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects were right-handed [Laterality Index $\geq$ 0.78; scaling adapted from Annett (1970)]. Written informed consent was obtained from each subject prior to the experiment. The study protocol was approved by the local ethics committee (Ärztekammer Hamburg).

EXPERIMENTAL PROCEDURES

The description of the experimental procedures is reproduced from Kassuba et al. (2013a: Experimental design and procedure, p. 60) with slight changes in phrasing. All subjects took part in four experimental sessions which were conducted on separate days (**Figure 1**). First, subjects attended a behavioral training session. An epoch-related fMRI localizer session was performed a day later. The last two sessions each consisted of a 40 min run of event-related fMRI, preceded by either 30 min of sham or real rTMS. The order of the real and sham rTMS sessions was counterbalanced across subjects and separated by at least one week.

In the initial training session, subjects were trained outside the MRI scanner room to recognize 24 object stimuli by viewing photographs and by haptic exploration with an appropriate speed (without ever viewing the real objects themselves). The training was repeated until the object stimuli were identified with an accuracy of 100% (0-1 repetitions per subject and object). In addition, subjects were familiarized with the visual and haptic texture stimuli used in the localizer fMRI session to be presented on the next day. We ran this training to avoid confounding effects due to differences in familiarity and recognition performance between the two modalities.

EPOCH-RELATED fMRI LOCALIZER

The description of the localizer is reproduced from (Kassuba et al., 2013a: Visuo-haptic fMRI localizer, p. 60) with slight changes in phrasing. The left LO (rTMS target region) and further regions of interest (ROIs) were identified by means of an fMRI localizer. The paradigm determined the convergence of brain activation during unimodal processing of visual and haptic object stimuli as compared to non-object control stimuli of the same modality. In different blocks, we presented visual, haptic or auditory object or corresponding texture stimuli, resulting in six different block conditions: visual-object,

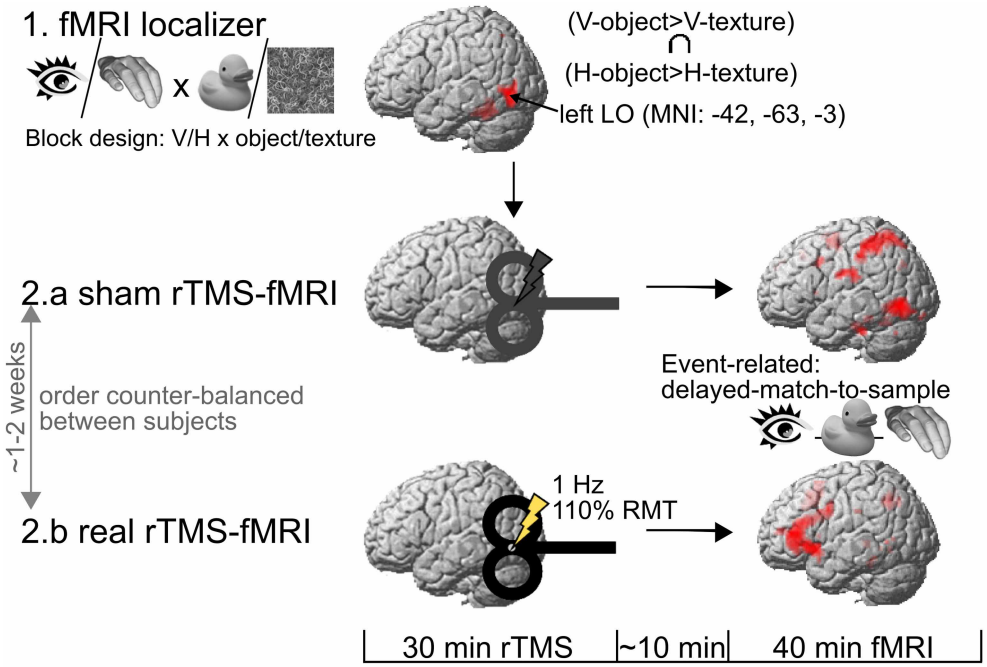


FIGURE 1 | Experimental procedure. One day after a behavioral training session (not illustrated here), subjects took part in an fMRI localizer in order to identify the rTMS target area in the left lateral occipital cortex (LO). The rTMS target area was defined by the convergence of visual (V) and haptic (H) processing of objects vs. textures in the left LO. Then, the subjects underwent two 40 min event-related fMRI sessions on separate days (at least one week apart), both preceded by an off-line session of either real or sham 1 Hz rTMS to the left LO. RMT, motor resting threshold.

haptic-object, auditory-object, visual-texture, haptic-texture, and auditory-texture. Within each block condition, subjects had to press a button whenever an identical stimulus was presented in two consecutive trials (1-back task, responses in 12.5% of trials). Each stimulation block lasted 30 s during which 8 stimuli from the respective condition were presented (2 s stimuli + 2 s inter-stimulus-interval). The subjects were informed 2.8 s before each block by a visual instruction (0.8 s) about the upcoming block and whether they would see (picture of an eye), touch (picture of a hand) or hear (picture of an ear) stimuli. Each stimulation block was followed by 11.5 s of rest, and each blocked condition was presented six times. The left LO was destined as the peak of the group mean BOLD response in the conjunction contrast (visual-object > visual-texture) ∩ (haptic-object > haptic-texture) at $p < 0.001$, uncorrected (MNI coordinates in mm: $x = -42$, $y = -63$, $z = -3$). The auditory stimuli were used in the context of a different research question (these results have been previously published in Kassuba et al., 2011).

EVENT-RELATED fMRI EXPERIMENT

The description of the event-related fMRI experiment is reproduced from (Kassuba et al., 2013a: Event-related fMRI experiment, pp. 60–62) with slight changes in phrasing. The main fMRI experiment entailed two experimental sessions that used an identical event-related fMRI paradigm (except for differences due to pseudorandomization of the conditions and stimuli). Each experiment started with a short practice session, consisting of a short recall of the initial training, and then subjects were familiarized with the subsequent fMRI task. Thereafter, real or sham 1 Hz rTMS was applied to the left LO for 30 min followed by the event-related fMRI experiment (for details on rTMS see Repetitive TMS).

Example trials of the event-related fMRI paradigm are shown in **Figure 2**. Each trial consisted of a sample object stimulus (S1) and a target object stimulus (S2) presented successively, and the subjects' task was to decide whether or not both stimuli referred to the same object (50% congruent and 50% incongruent). The object stimuli were presented either haptically (actively palpating an object) or visually (seeing a black-and-white photograph of an object; for a detailed description of the objects, see Object Stimuli), and S1 and S2 were both presented either within the same modality (unimodal) or across modalities (crossmodal). With respect to the event of matching (i.e., processing of S2 and relating it to S1), the experiment resulted in a 2 × 2 × 2 design. The first factor was the mode of sensory matching (unimodal or crossmodal). The second factor related to congruency in object identity between S1 and S2 (congruent or incongruent). The sensory modality of the S2 (visual or haptic) constituted the third experimental factor.

A visual instruction was presented before each stimulus which specified the type of upcoming stimulus (S1 or S2) and whether subjects would see or touch it. An exclamation mark announced an S1, a question mark an S2, a white font a visual stimulus, and a black font a haptic stimulus. The instruction was presented for 0.5 s. A short blank screen of 0.1 s separated instruction and stimulus presentation. S1 and S2 were both presented for 2 s.

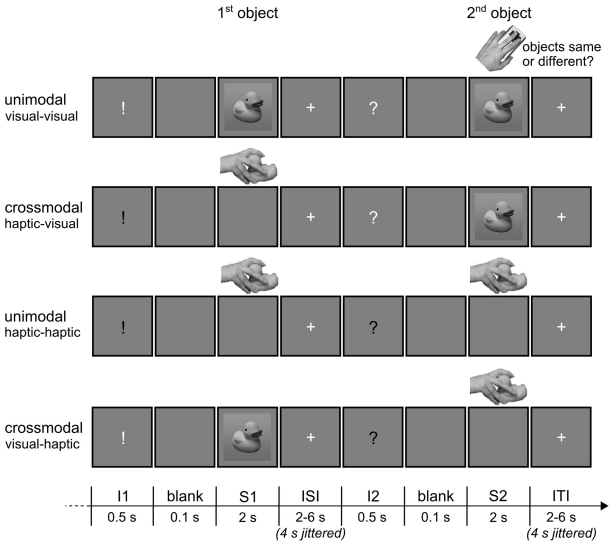


FIGURE 2 | Illustration of the event-related fMRI paradigm. Each trial consisted of a sample (S1) and a target object stimulus (S2), and subjects had to decide by button press whether the two objects were congruent (50%) or incongruent (50%). S1 and S2 were either haptic or visual stimuli, and both could be presented either within the same modality (unimodal) or across modalities (crossmodal). A white or black visually presented exclamation (I1) and question mark (I2) before the stimuli informed the subjects about the sensory modality S1 and S2, respectively, would be presented in. ISI, inter-stimulus-interval; ITI, inter-trial-interval. Reprinted from NeuroImage, 65, Kassuba et al., Vision holds a greater share in visuo-haptic object recognition than touch, p. 61, Copyright Elsevier Inc. (2013a), with permission from Elsevier.

Inter-stimulus- and inter-trial-intervals (i.e., the time between the offset of an S1 or S2 stimulus, respectively, and onset of the next visual instruction) were randomized between 2 and 6 s in length (in steps of 1 s). During the whole scanning session, the visual display showed a gray background (RGB 128/128/128) on which either the visual objects, the visual instructions, a white fixation cross (inter-stimulus- and inter-trial-interval) or nothing was presented (blank and presentation of haptic objects). Trials were presented pseudo-randomized such that the same objects would not repeat across successive trials. Moreover, the sensory modality combination was repeated maximally once across successive trials. Every object appeared once as S1 and once as S2 in each experimental condition. The combination of S1 and S2 objects in incongruent trials was randomized. Importantly, subjects did not know whether the S2 would be a visual or a haptic object until 0.6 s before its onset (i.e., when a visual instruction informed the subjects about the modality of S2). Thus, all trials with a visual S1 and all trials with a haptic S1 were identical, respectively, until shortly before the onset of S2.

A total of 192 trials (24 trials per condition) were presented during each fMRI experiment. The experiment was split into two runs lasting approximately for 20 min (96 pseudorandomized trials per run). Subjects lay supine in the scanner with their right hand on the right side of a custom-made board fixed by a vacuum-cushion onto their waists. The board was placed such that subjects were comfortably able to reach the placement area in the middle of the board with their forearm and hand without moving either the upper arm or neck muscles. Their left hand was placed beside the body and rested on the button box. Subjects were presented with a white fixation cross and instructed to wait for a visual instruction. When presented with a black sign, they were asked to move their right hand toward the placement area and explore the presented object. During visual and haptic stimulus presentations, the fixation cross disappeared. Subjects were trained to keep a pace of maximally 2 s for hand movements and exploration, and they were asked to repose their hand after the fixation cross reappeared. In case of a white sign, they were asked to look at the following visually presented object until the fixation cross reappeared. At the presentation of S2, subjects were instructed to indicate by button press as fast and accurately as possible whether both objects were the same or different. Responses were made with the middle and index finger of the left hand, with the finger-response assignment being counterbalanced across subjects. Visual stimuli were presented using Presentation (Neurobehavioral Systems, Albany, CA, USA) running on a Windows XP professional SP3 PC. Visual stimuli (objects subtended 8° × 8° and instructions 0.6° × 1.6° on a background of 23° × 12° of visual angle) were back-projected onto a screen using a LCD projector (PROxtraX, Sanyo, Munich, Germany) visible to the subjects through a mirror mounted on the MR head coil. Haptic stimuli were exchanged by the investigator, and the individual objects were always placed in the same viewpoint. The investigator was informed by auditory instructions one trial in advance about which object had to be placed and also about the start and ending of trials. Thus, the investigator was able to control that the haptic stimuli were palpated within the required time frame.

OBJECT STIMULI

The stimulus description is reproduced from (Kassuba et al., 2013a: Stimuli, p. 62) with slight changes in phrasing. Visual and haptic object stimuli were identical for the localizer and the experimental task [same as in Kassuba et al. (2011, 2013a)]. They were manipulable man-made hand-sized objects that the subjects palpated with their right hand. Object categories were restricted to tools, toys, and musical instruments. All objects were real-sized and composed of the same material as in the real world so that they were familiar to the subjects. Furthermore, the objects were deliberately chosen to have an original size such that the objects were easy to palpate and manipulate with one hand. Identical objects appeared in both sensory modalities. Visual object stimuli were black/white photographs taken from the objects used as haptic stimuli. The objects were photographed from the corresponding viewpoint as they were presented to the participants in the haptic condition, and centered on a 350 × 350 pixel sized square consisting of a vertical gray gradient going from RGB 108/108/108 to 148/148/148.

MRI DATA ACQUISITION

The study was carried out on a 3 Tesla MRI scanner with a 12-channel head coil (TRIO, Siemens, Erlangen, Germany). We acquired 38 transversal slices (216 mm FOV, 72 × 72 matrix, 3 mm thickness, no spacing) covering the whole brain using

a fast gradient echo T2*-weighted echo planar imaging (EPI) sequence (TR 2480 ms, TE 30 ms, 80° flip angle). High-resolution T1-weighted anatomical images were additionally acquired after the localizer fMRI scan using an MPRAGE (magnetization-prepared, rapid acquisition gradient echo) sequence (256 mm FOV, 256 × 192 matrix, 240 transversal slices, 1 mm thickness, 50% spacing, TR 2300 ms, TE 2.98 ms).

REPETITIVE TMS

Focal rTMS was applied off-line outside the MR scanner room using a figure-of-eight coil attached to a Magstim Rapid stimulator (Magstim Company, Dyfeld, UK). The coil was centered over the left LO using Brainsight frameless stereotaxy (Rogue Research, Montreal, Canada). The center of the eight-shaped coil targeted the MNI coordinates in mm: $x = -42, y = -63, z = -3$ as determined by the localizer (for details see Epoch-related fMRI Localizer). For each subject, the group peak LO coordinates were transformed into individual anatomical MRI native space coordinates, and the site of rTMS stimulation was verified and traced throughout the conditioning with the frameless stereotaxy device.

Subjects received continuous 1 Hz rTMS for 30 min (1800 stimuli). Stimulation intensity was set to 110% of the individual resting motor threshold (RMT) of the right first dorsal interosseous muscle. Mean stimulation intensity during real rTMS was 53.00 ± 7.50% of total stimulator output. The RMT was defined as the lowest stimulus intensity that evoked a motor evoked potential (MEP) of 50 μV in five out of ten stimuli given over the motor hot spot. Besides the stimulation intensity, the rTMS protocol was identical the protocol used by Siebner et al. (2003) which had resulted in a suppression of neuronal activity in the stimulated left dorsal premotor region that was measurable for at least 1 h after the end of stimulation. In the current study, stimulation intensity was increased to account for the greater scalp-cortex distance of the target region compared to primary motor cortex (Stokes et al., 2005). Repetitive TMS was well tolerated by all participants apart from one female subject who aborted the real rTMS session because of uncomfortable sensations on her skull. Four of the remaining 18 subjects displayed slight twitches in neck and jaw muscles during real rTMS. Repetitive TMS of the left LO did not produce phosphenes in any subject.

MEPs were recorded from the first dorsal interosseous muscle with Ag-AgCl electrodes attached to the skin using a tendon-belly montage. Electromyographic responses were amplified, filtered, and sampled using a D360 eight-channel amplifier (Digitimer, Welwyn Garden City, UK), a CED 1401 analog-to-digital converter (Cambridge Electronics Design, Cambride, UK), and a personal computer running Signal software (Cambridge Electronic Design). The sampling rate was 5 kHz, and signals were band-pass filtered between 5 and 1000 Hz.

An air-cooled figure-of-eight coil (double 70 mm cooled coil system; Magstim Company) was used for real rTMS. The coil was placed tangential to the skull with the handle pointing backward, parallel to the horizontal and the mid-sagittal plane (Ellison and Cowey, 2006). For sham rTMS, a non-charging standard figure-of-eight coil (double 70 mm coil; Magstim Company) was placed at the skull instead, and the charging coil was placed 90° tilted on top of the non-charging coil. In order to provide a comparable acoustic stimulus, intensity of the charging coil was increased for 15% of the total stimulator output. In analogy to the sham rTMS condition, the non-charging coil was placed 90° tilted on top of the charging coil during real rTMS in order to keep the real and sham rTMS conditions as similar as possible.

Repetitive TMS conditioning was performed offline before fMRI but after the short object recognition and task training session. On average, it took 10 ± 2 min from the end of rTMS until fMRI data acquisition was started. This time was needed to move the subjects from the TMS lab to the MR scanner, bed them, set up the board for haptic stimulus presentation, and localize the FOV. Since previous neuroimaging studies have shown that 1Hz rTMS conditioning can produce effects on regional neuronal activity that last for up to 1 h after the end of stimulation (Lee et al., 2003; Siebner et al., 2003), fMRI lasted 40 min and was, thus, within the time limits for capturing reorganizational effects.

BEHAVIORAL DATA ANALYSIS

The description of the behavioral data analysis is reproduced from (Kassuba et al., 2013a: Behavioral analysis, p. 62) and adjusted to include rTMS-specific analysis steps. For each subject and for each trial condition, mean RTs relative to the onset of S2, and response accuracies were calculated. Only correct responses were considered for further analyses (trials excluded due to errors: 0–5 per subject/condition, overall Median = 0; $M \pm SD$ sham rTMS session 0.69 ± 0.84 trials, real rTMS session 0.81 ± 1.07 trials, $p = 0.32$). Haptic trials in which participants did not palpate the object, dropped the object, or made premature or late palpations, as well as palpations lasting longer than 2 s were excluded from analysis (sham rTMS session 0.04 ± 0.08 trials, real rTMS session 0.01 ± 0.03 trials, $p = 0.41$). Within each participant and condition, RTs that differed ±3 standard deviations from the preliminary mean were defined as outliers and excluded from further analyses (sham rTMS session 0.29 ± 0.16 trials, real rTMS session 0.29 ± 0.19 trials). Mean RTs of the adjusted data were entered into a repeated-measures ANOVA (PASW Statistics 18) with RTMS (real/sham), S2-MODALITY (visual/haptic), CONGRUENCY (congruent/incongruent), and SENSORY-MATCHING (unimodal/crossmodal) as within-subject factors. In order to capture transient effects of rTMS conditioning on behavior, RTs within each condition were divided into four time bins of about 10 min each (∼4-7 trials/bin). Additional ANOVAs with the factors TIME and RTMS were run for each S1-S2 condition. Each of these ANOVAs tested for a linear trend in the factor Time, and whether this trend interacted with rTMS. Statistical effects at $p < 0.05$ were considered significant. *Post-hoc* Bonferroni corrected paired *t*-tests were used to test for differences between single conditions.

FUNCTIONAL MRI DATA ANALYSIS

The basic steps of the fMRI analysis is reproduced from Kassuba et al. (2013a: Functional image analysis, pp. 62–63) with slight changes in phrasing and adjustments to include rTMS-specific analysis steps. Image processing and statistical analyses were performed using SPM8 (statistical parametric mapping 8; www.fil.ion.ucl.ac.uk/spm). The first five volumes of each time

series were discarded to account for T1 equilibrium effects. Data processing consisted of slice timing (correction for differences in slice acquisition time), realignment (rigid body motion correction) and unwarping (accounting for susceptibility by movement interactions), spatial normalization to MNI standard space as implemented in SPM8, thereby resampling to a voxel size of $3 \times 3 \times 3\,mm^3$, and smoothing with an 8 mm full-width at half-maximum isotropic Gaussian kernel.

Statistical analyses were carried out using a general linear model approach. The time jitter between the onsets of S1 and S2 allowed us to model the effects of rTMS on sample encoding (response to S1) and target matching (response to S2) independently. At the individual level (fixed effects), we defined separate regressors for the onsets of S1 and S2 in each session (i.e., after sham and real rTMS): two different S1 regressors (one for visual S1 and one for haptic S1; Vx and Hx) and eight different S2 regressors (one for each matching condition: V, visual; H, haptic; c, congruent; i, incongruent: VVc, HVc, VVi, HVi, HHc, VHc, HHi, VHi) for each rTMS condition. Only onsets of S1 and S2 in correct trials withstanding the same inclusion criteria as applied for RT analyses were included. An additional regressor modeled the onsets of S1 and S2 in all excluded trials (errors, improper haptic exploration, and outliers) combined over all conditions. All onset vectors were modeled by convolving delta functions with a canonical hemodynamic response function as implemented in SPM8 and their first derivative. Low frequency drifts in the BOLD signal were removed by a high-pass filter with a cut-off period of 128 s. On the group level, we evaluated effects of rTMS on sample encoding (onset S1), target matching (onset S2) as well as time dependent effects.

Sample encoding (onset S1)

In order to determine the modulation of visual (Vx) and haptic (Hx) S1 encoding by rTMS on the group level (random effects), a flexible factorial design with the within-subject factors MODALITY (Vx/Hx) and rTMS (r/s) was configured. The model also included the estimation of the subjects' constants in form of a SUBJECT factor, and accounted for a possible non-sphericity of the error term (dependences and possible unequal variances between conditions in the within-subject factors).

Target matching (onset S2)

Given the complexity of the design (rTMS × S2-MODALITY × CONGRUENCY × SENSORY-MATCHING: $2 \times 2 \times 2 \times 2$), we aggregated the S2 matching conditions (S2-MODALITY × CONGRUENCY × SENSORY-MATCHING) into one S2-Condition factor (SPM does not allow a specification of more than 3 factors in a factorial model). In order to evaluate the modulation of S2 processing in a random effects group analysis, we configured a flexible factorial design with the within-subject factors rTMS (r/s) and S2-CONDITION (VVc/HVc/VVi/HVi/HHc/VHc/HHi/VHi). The model also included the estimation of the subjects' constants in form of a SUBJECT factor, and accounted for a possible non-sphericity of the error term (dependences and possible unequal variances between conditions in the within-subject factors). Note that in order to evaluate S2 matching effects, we first calculated contrasts of interest for visual and haptic S2 conditions separately (e.g., crossmodal > unimodal × congruent > incongruent for haptic S2: [VHc - HHc] > [VHi - HHi], for visual S2: [HVc - VVc] > [HVi - VVi]). This enabled us to eliminate modality-specific confounding factors such as residual effects of the cue on S2 processing, eye movements or potential visual imagery and motor activations during haptic but not visual exploration. In a next step, we compared these modality-specific differential effects across modalities (instead of comparing visual and haptic S2 processing directly).

Time-dependent effects of rTMS

Time-dependent effects on the processing of S1 and matching of S2 were also investigated in order to capture transient effects of rTMS on task-related neuronal processing which gradually recovered during the ∼40 min fMRI session. In each session, each of the two S1 processing conditions (Vx, Hx) was divided into 10 time bins (5 time bins per run) of about 4 min each (∼7-10 trials/bin). In the single subject analysis, we defined a regressor for each time bin in each condition. For each condition, we defined contrasts that represented a linear or an exponential modulation over time (i.e., across successive time bins). The exponential function we modeled was $y = a + (b \cdot 2^{-x})$, where y is the BOLD signal and x is time. The beta images of these contrasts of all subjects in the real rTMS and the sham rTMS sessions were then entered into a random effects flexible factorial model [cf. Sample Encoding (Onset S1)] in order to compare time-dependent effects between real and sham rTMS sessions on the group level.

We applied the same approach to the analysis of S2 responses. Here, each S2 matching condition was divided into four time bins of about 10 min each (∼4-7 trials/bin) and fitted to a linear function. A division into more than four time bins was not reasonable given the limited number of trials. Given only four time bins for the S2 matching conditions, non-linear time-dependent effects were not modeled here.

Regions of interest

The description of the regions of interest is reproduced from Kassuba et al. (2013a: Functional image analysis, p. 63) with slight changes in phrasing and adjustments. We report voxel-wise family wise error rate (FWE) corrected p-values as obtained from small volume correction in visuo-haptic regions of interest (ROIs; $p < 0.05$). Four brain regions were predefined as ROIs: LO, FG, aIPS, and pIPS. The ROIs in left and right LO and FG were delineated from the localizer. Images of the localizer data were preprocessed and analyzed as reported previously (Kassuba et al., 2011). Converging object-specific processing across vision and haptics was calculated with a conjunction of the respective object > texture contrasts within each modality. Only voxels that showed an absolute increase during object processing vs. baseline fixation were included. Small volume correction was based on spheres of 8 mm radius centered at the group-based peak coordinates obtained from the conjunction contrast thresholded at $p < 0.001$, uncorrected: $x = -42$, $y = -63$, $z = -3$ for the left LO (rTMS target), $x = 48$, $y = -69$, $z = -9$ for the right LO, $x = -36$, $y = -39$, $z = -21$ for the left FG, and $x = 36$, $y = -45$, $z = -27$ for the right FG.

Four additional ROIs in the left and right aIPS and pIPS were derived from previous studies applying a crossmodal matching task. Correction was based on spheres of 8 mm radius centered at group-based peak coordinates reported by the previous studies. Talairach coordinates (Talairach and Tournoux, 1988) from previous studies were transformed into MNI standard space (mm) as implemented in SPM8 using a MATLAB code provided by BrainMap (http://brainmap.org/icbm2tal/index.html; Lancaster et al., 2007). The spherical ROIs were centered over the stereotactic coordinates $x = -42$, $y = -40$, $z = 40$ for the left aIPS (Grefkes et al., 2002), $x = -28$, $y = -65$, $z = 49$ for the left pIPS (Saito et al., 2003), and $x = 31$, $y = -62$, $z = 50$ for the right pIPS (Saito et al., 2003). We also included the right hemispheric homolog of the left aIPS as a region of interest ($x = 42$, $y = -40$, $z = 40$). Whole-brain voxel-wise FWE correction was applied for all other voxels in the brain ($p < 0.05$). Activations derived from the whole-brain analyses were anatomically labeled using the probabilistic stereotaxic cytoarchitectonic atlas implemented in the SPM Anatomy Toolbox version 1.8 (Eickhoff et al., 2005), adjusted based on anatomical landmarks in the average structural T1-weighted image of all subjects. Percent signal changes used for visualization of the results were extracted using the SPM toolbox rfxplot (Gläscher, 2009).

RESULTS

BEHAVIORAL PERFORMANCE

Task performance after sham rTMS has been reported in a previous paper (Kassuba et al., 2013a). In short, RTs were longer for incongruent than for congruent trials [$F_{(1,17)} = 31.43$, $p < 0.001$], indicating that incongruent matching was in general more demanding than congruent matching. RTs decreased linearly during the fMRI session in all conditions [$F_{(1,17)} = 14.37$, $p < 0.01$]. Response accuracies were nearly perfect irrespectively of condition (on average 96.76 ± 0.97% correct). Neither response accuracies nor RTs (time-dependent and time-independent effects) were affected by rTMS conditioning ($p > 0.10$, see **Figure 3** and Supplementary Table S1).

FUNCTIONAL MRI

The fMRI results after sham rTMS have been reported in a previous paper (Kassuba et al., 2013a).

Sample encoding (response to S1)

Bilateral LO, FG, aIPS, and pIPS were all activated during visual and haptic S1 encoding both after sham and real rTMS [$t_{(51)} \geq 5.30$, $p < 0.001$, corrected]. This mean response to S1 was increased in an inferior portion of bilateral FG [left: −33, −46, −23, $t_{(51)} = 2.86$, $p = 0.052$, corrected; right: 33, −43, −23, $t_{(51)} = 3.46$, $p < 0.05$, corrected; see **Figures 4A,B**] after real as opposed to sham rTMS but otherwise did not differ between the two sessions ($p > 0.01$, uncorrected).

Real TMS affected the activity at the site of stimulation (left LO) mainly during S1 encoding and in a time-dependent fashion. After real rTMS, the BOLD response at the left LO to haptic S1 was initially attenuated and exponentially recovered until ∼30 min post rTMS [−42, −67, −11; $t_{(51)} = 3.49$, $p < 0.05$, corrected; see **Figure 4C**]. The regional BOLD response to haptic S1 stimuli displayed opposite temporal dynamics after sham rTMS with a higher initial level of S1-induced activity which quickly attenuated during continuous task performance. Relative to sham rTMS, real rTMS additionally caused a transient attenuation of haptic S1 processing in the right LO [45, −73, −5; $t_{(51)} = 3.37$], a superior portion of bilateral FG [left: −36, −46, −20, $t_{(51)} = 3.74$; right: 36, −43, −20, $t_{(51)} = 3.16$], and bilateral posterior superior temporal sulcus and adjacent middle temporal gyrus [pSTS/MTG; left: −66, −40, 1; $t_{(51)} = 6.10$; right: 54, −40, −8, $t_{(51)} = 5.44$; all $p < 0.05$, corrected; see **Figures 4A,C**]. Similar but weaker ($p < 0.05$, uncorrected) transient decreases in activation were found for visual S1 encoding as well. The effects for haptic S1 were not significant different from the effects for visual S1 ($p > 0.05$, corrected).

Target matching (response to S2)

Effects of real rTMS on crossmodal congruent matching. We expected rTMS to evoke the strongest reorganizational effects for crossmodal matching of semantically congruent stimulus pairs (i.e., in the crossmodal matching by semantic congruency interaction contrast as indication for multisensory interactions). After sham rTMS, we had found such multisensory interaction effects in bilateral LO, FG, aIPS, and pIPS which were more pronounced for haptic than visual S2 (Kassuba et al., 2013a). Based on these findings, we proposed that multisensory interactions are more likely for haptic than visual object recognition, and we, therefore, expected stronger effects of real rTMS for the matching of haptic as opposed to visual S2. After real rTMS, we found comparable multisensory interaction effects in our ROIs that were stronger pronounced for haptic as opposed to visual S2 conditions (see **Figure 5** and Supplementary Tables S2–S4). We did not observe any significant effects of rTMS on multisensory interactions (rTMS x crossmodal > unimodal × congruent > incongruent) nor on crossmodal matching effects (rTMS × crossmodal > unimodal), neither for visual nor haptic S2.

However, real rTMS altered the temporal dynamics of event-related activity during crossmodal matching compared to sham rTMS. Several regions in left temporal cortex showed initial increases in activations after real rTMS during crossmodal matching of congruent onjects (see **Table 1**). These effects of real rTMS were transient and decreased gradually during the fMRI session, resulting in a negative linear modulation of the BOLD response. For congruent crossmodal matching of haptic S2 (VHc), the left FG showed an initial relative enhancement of the BOLD response to S2 after real rTMS with a subsequent linear decay over time. In contrast, for congruent crossmodal matching of visual S2 (HVc), the left temporal pole and pSTS/MTG displayed an initial increase in S2-related activation after real rTMS (see **Table 1**). Direct comparisons between the two modalities (r-VHc > s-VHc × time vs. r-HVc > s-HVc × time) showed that these effects were modality specific. No consistent effects of real rTMS were found during unimodal matching in these regions. Yet, the effects found for crossmodal matching did not differ significantly from the effects for unimodal matching.

Effects of real rTMS on incongruent matching. Longer response latencies suggested that matching of incongruent objects was behaviorally more challenging than matching of congruent

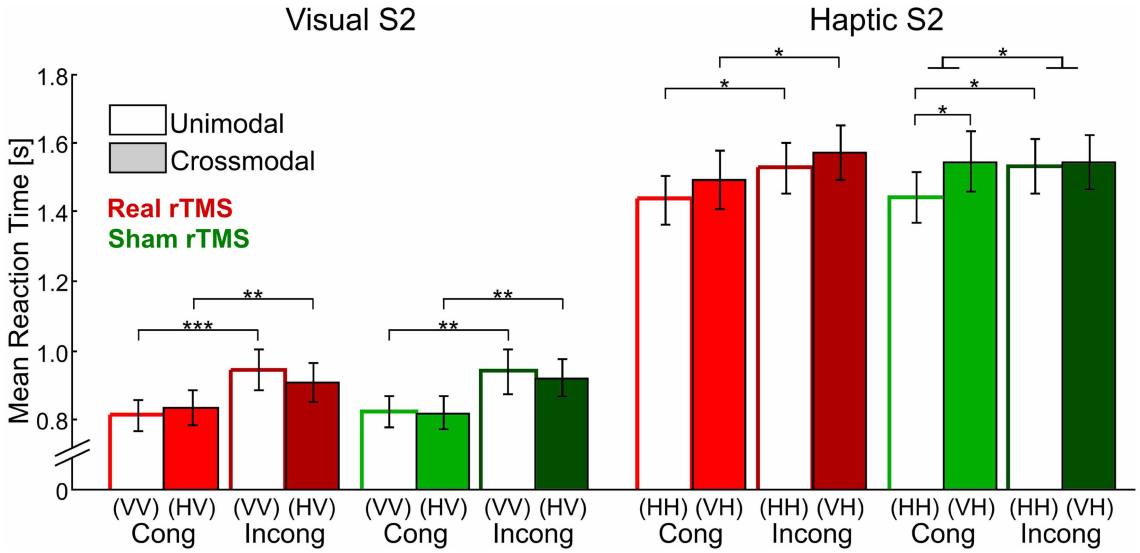


FIGURE 3 | Mean reaction times for visual and haptic S2 for all four delayed matching conditions (unimodal/crossmodal × congruent/incongruent) and after real (red bars) and sham rTMS (green bars) conditions. Error bars indicate the standard error of the mean. Reaction times were recorded from S2 onset onwards. Repetitive TMS did not have any effects on reaction times. Sample-target (S1–S2) conditions: V, visual; H, haptic. Cong, congruent; Incong, incongruent. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, Bonferroni corrected.

objects (see **Figure 3**). Since behavioral performance was not impaired by rTMS, we next asked whether we could find reorganizational effects on the neuronal level related to incongruent matching, that is, triggered by task difficulty. We found rTMS-induced increases in activations related to matching of incongruent objects for both haptic and visual S2. These effects were found transiently for crossmodal matching of haptic S2 and lastingly (i.e., temporally stable for the whole duration if the experiment) for unimodal matching of visual S2 (see **Table 2**). When a haptic S2 was matched to an incongruent visual S1 (r-VHi > s-VHi), real rTMS-induced transient increases in activation were found in bilateral parahippocampus, right LO, bilateral pSTS/MTG, IPS, and in the right middle and adjacent superior frontal gyrus. On the other hand, when a visual S2 was matched to an incongruent visual S1, temporarily stable increases in activation were found in the left FG and pIPS. No other incongruent matching condition was affected by real rTMS.

Incongruency effects (incongruent > congruent) after real rTMS. The time-dependent effects in the right anterior parahippocampus and middle frontal gyrus and adjacent precentral gyrus found for crossmodal matching of haptic S2 were significantly more pronounced for incongruent than congruent conditions (real > sham × VHi > VHc × time, see **Table 3** and **Figure 6**). Thus in these regions, real rTMS conditioning induced incongruency effects, that is, stronger activations during incongruent than congruent matching, that were not evident after sham rTMS. Such rTMS by incongruency interactions (real > sham × incongruent > congruent) were found for unimodal visual (VV) matching as well. For unimodal visual matching, temporarily stable rTMS-induced incongruency effects were found the left superior medial gyrus extending to the right hemisphere, left FG, and bilateral pIPS (see **Table 3** and **Figure 7**). A direct comparison of visual and haptic S2 conditions showed that these time-dependent ([r-VHi > s-VHi × time] > [r-HVi > s-HVi × time]) and time-independent effects ([r-VVi > s-VVi] > [r-HHi > s-HHi]) were modality-specific. Unimodal matching of haptic S2 and crossmodal matching of visual S2 did not show real rTMS-induced incongruency effects.

Exclusion of subjects with low LO activations in the localizer

One concern with respect to the null findings regarding multisensory interactions could be that we used the peak coordinates from the localizer group analysis as rTMS target instead of individual peaks. Yet theoretically, the group peak coordinates represent the peak responses across subjects, and indeed, the Eucledian distance between individual peaks and the group peak were smaller than 1 cm in all subjects. However, 5 out of the 18 subjects showed very weak activations in the localizer contrast and peaks in the left LO could only be localized at very low thresholds ($p > 0.05$, uncorrected). In these subjects, the group peak coordinates provided a more objective guide for placing the TMS coil. To test whether these subjects had biased our results, we repeated our analyses without these 5 subjects. There were still no significant effects of rTMS on multisensory interactions.

DISCUSSION

We probed short-term plasticity of visuo-haptic object recognition by conditioning neuronal processing in left LO with low-frequency offline rTMS. Compared to sham rTMS, real rTMS led to a dynamic redistribution of brain activity during visuo-haptic object matching. Changes in task-related activity were not only triggered in the stimulated left and contralateral

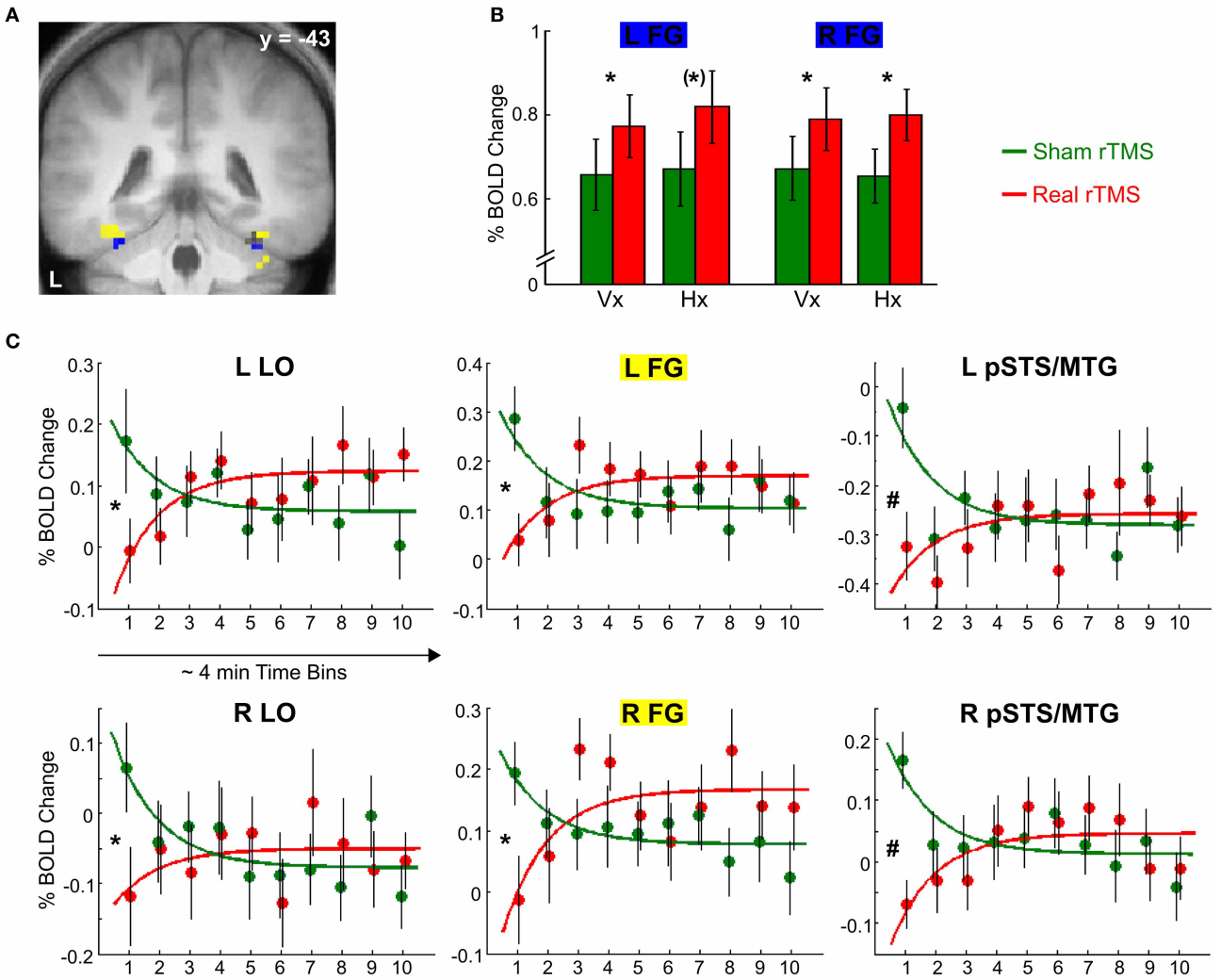


FIGURE 4 | Effects of real rTMS on S1 processing. (A) Activation map showing temporally stable increases in activation (blue) and transient decreases in activation (yellow) in bilateral FG after real compared to sham rTMS ($p < 0.01$, uncorrected). **(B)** Temporarily stable increases in activation in bilateral FG (blue portion in **(A)**, MNI coordinates *x, y, z*; left: −33, −46, −23; right: 33, −43, −23) to both visual S1 (Vx) and haptic S1 (Hx) after real (red) relative to sham rTMS (green). **(C)** Transient rTMS-induced decreases in activation during haptic S1 encoding. Regional activity in bilateral LO, FG, and pSTS/MTG showed an interaction of exponential time-dependent effects by rTMS condition when haptic S1 were processed: Whereas regional activity was initially decreased and exponentially increased over time after real rTMS (red), the reversed pattern was found after sham rTMS (green). Similar but weaker effects were found for visual S1 processing ($p < 0.05$, uncorrected). Each time bin represents ~4 min and 7–10 trials. FG, fusiform gyrus [yellow portion in **(A)**, left: −36, −46, −20; right: 36, −43, −20]; LO, lateral occipital cortex (left, i.e., rTMS target area: −42, −67, −11; right: 45, −73, −5); pSTS/MTG, posterior superior temporal sulcus /middle temporal gyrus (left: −66, −40, 1; right: 54, −40, −8). L, left; R, right. $^{*}p < 0.05$, small volume corrected, (*)$p = 0.052$, small volume corrected, $^{\#}p < 0.05$, whole brain corrected.

LO but also in remote temporal and parietal regions previously associated with object recognition. While LO, FG, aIPS, pIPS have been implicated in visuo-haptic object recognition (Amedi et al., 2001; Grefkes et al., 2002; Saito et al., 2003; Kassuba et al., 2011), the pSTS/MTG seems to participate in audio-visual and audio-haptic object recognition (Beauchamp et al., 2004, 2008; Kassuba et al., 2011, 2013b), and the temporal pole appears to support semantic memory (Martin and Chao, 2001; Rogers et al., 2006). Since behavioral performance was not impaired, the real rTMS-induced changes in task-related brain activity likely indicate compensatory processes preserving behavior after neuronal challenge. Importantly, the pattern of real rTMS-induced changes in regional activity differed as a function of the stage of the delayed-match-to-sample task (S1 encoding vs. S2 matching) and the target modality.

Since various previous studies have implicated the left LO in visuo-haptic integration of object information (Lacey and Sathian, 2011), we predicted that rTMS of the left LO would particularly affect multisensory interactions as defined by crossmodal matching by semantic congruency interactions and particularly for haptic S2 conditions (Kassuba et al., 2013a,b). Contrary to our expectations, rTMS had no impact on crossmodal matching

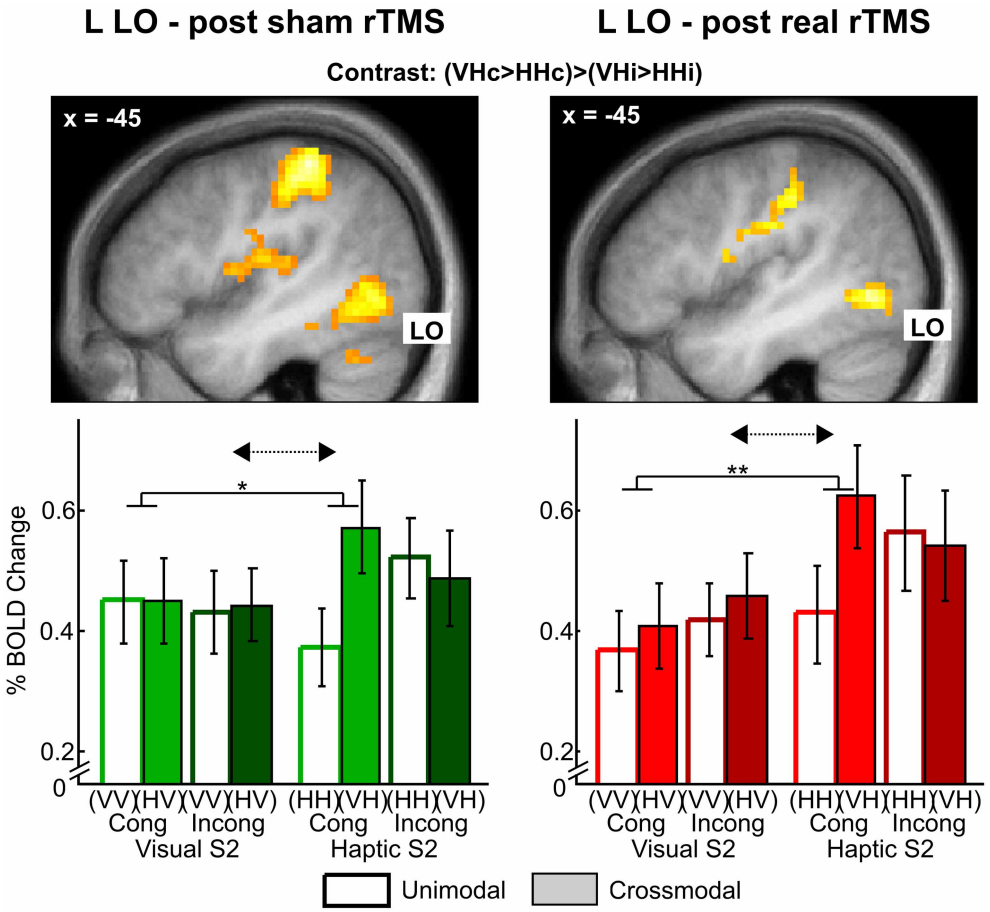


FIGURE 5 | Stronger multisensory interaction effects for haptic vs. visual S2 in left lateral occipital cortex (LO) after sham (left) and real rTMS (right). Top: Activation maps showing a crossmodal matching by semantic congruency interaction effect (crossmodal > unimodal × congruent > incongruent) for haptic S2. For illustrative purposes, the statistical maps are thresholded at $p < 0.01$, uncorrected, and overlaid on the average structural T1-weighted image of all subjects. **Bottom:** Percent signal change and error bars indicating the standard error of the mean for each condition at the left LO (MNI coordinates: *x*, *y*, *z* = −45, −70, −5). Stronger visuo-haptic interaction effects for haptic as opposed to visual S2 (solid lines: congruent crossmodal > unimodal × haptic S2 > visual S2, $^{*}p < 0.05$, $^{**}p < 0.01$, corrected; dashed arrows: crossmodal > unimodal x congruent > incongruent × haptic S2 > visual S2, $p \leq 0.063$, corrected) were not affected by rTMS ($p > 0.01$, uncorrected). Sample-target (S1-S2) conditions: V, visual; H, haptic; Cong, congruent; Incong, incongruent. For further significant results see Supplementary Tables S2–S4. The results after sham rTMS have been previously published in Kassuba et al. (2013a).

Table 1 | Linear time-dependent effects of rTMS on regional activity during crossmodal congruent matching.

Region	*x*	*y*	*z*	t_{peak}	p_{corr} (p_{uncorr})	
HAPTIC S2 (VHc): POST-REAL rTMS > POST-SHAM rTMS × Time						**>VISUAL S2**
L FG	−42	−34	−20	3.03	0.024[§]	0.066[§]
R anterior parahippocampus	24	−1	−29	3.76	(<0.001)	(<0.001)
R precentral gyrus	39	−7	52	3.95	(<0.001)	0.032
VISUAL S2 (HVc): POST-REAL rTMS > POST-SHAM rTMS × Time						**>HAPTIC S2**
L temporal pole	−51	−4	−29	5.85	0.001	0.004
L pSTS/MTG	−66	−40	−2	4.87	0.034	(<0.001)

Coordinates are denoted by x, y, z in mm (MNI space) and indicate the peak voxel. The last column shows the direct comparison of the respective effect to the corresponding effect of the other S2 modality condition. Strength of activation is expressed in t- and p-values corrected for the whole brain and uncorrected p-values in parentheses, respectively, at peak voxel (df = 119), [§]small volume corrected. FG, fusiform gyrus; pSTS/MTG, posterior superior temporal sulcus/middle temporal gyrus. L, left; R, right.

Table 2 | Linear time-dependent effects of rTMS on regional activity during crossmodal incongruent matching.

Region	*x*	*y*	*z*	t_{peak}	p_{corr} (p_{uncorr})	
TEMPORALLY STABLE EFFECTS: POST-REAL rTMS > POST-SHAM rTMS						
Haptic S2—crossmodal: VHi						
No significant results						
Haptic S2—unimodal: HHi						
No significant results						
Visual S2—crossmodal: HVi						
No significant results						
Visual S2—unimodal: VVi						**>haptic S2**
L FG	−33	−43	−26	2.70	0.051[§]	0.289[§]
L pIPS	−30	−64	55	2.91	0.034[§]	0.192[§]
LINEAR TIME DEPENDENT EFFECTS: POST-REAL rTMS > POST-SHAM rTMS						
Haptic S2—crossmodal: VHi × time > visual S2						**>visual S2**
L parahippocampus / FG	−36	−16	−20	5.69	0.001	(<0.001)
R middle frontal gyrus	39	−4	61	5.46	0.004	0.014
R superior frontal gyrus	27	5	64	5.13	0.013	(<0.001)
R anterior parahippocampus	24	−1	−26	5.28	0.007	
L pSTS/MTG	−48	−73	13	5.14	0.013	(<0.001)
R pSTS/MTG	42	−70	10	4.66	(<0.001)	(<0.001)
R LO	48	−73	−2	3.56	0.006[§]	0.024[§]
L aIPS	−36	−37	37	3.55	0.006[§]	0.002[§]
L pIPS	−21	−67	49	3.65	0.005[§]	0.001[§]
R pIPS	27	−61	46	2.99	0.026[§]	0.017[§]
Haptic S2—unimodal: HHi × time						
No significant results						
Visual S2—crossmodal: VHi × time						
No significant results						
Visual S2—unimodal: VVi × time						
No significant results						

Coordinates are denoted by x, y, z in mm (MNI space) and indicate the peak voxel. The last column shows the direct comparison of the respective effect to the corresponding effect of the other S2 modality condition. Strength of activation is expressed in t- and p-values corrected for the whole brain and uncorrected p-values in parentheses, respectively, at peak voxel (df = 255 for temporally stable effects, df = 119 for time-dependent effects), [§]small volume corrected. aIPS, anterior intraparietal sulcus; FG, fusiform gyrus; LO, lateral occipital cortex; pIPS, posterior IPS; pSTS/MTG, posterior superior temporal sulcus/middle temporal gyrus. L, left; R, right.

effects (crossmodal > unimodal) regardless of whether or not semantic congruency was considered and neither for visual nor haptic S2.

ATTENUATED RESPONSE TO S1 BUT NOT S2 AT THE SITE OF STIMULATION (LEFT LO)

However, in accordance with a suppressive effect on regional neuronal activity (Gerschlager et al., 2001; Siebner et al., 2003) focal 1 Hz rTMS of the left LO temporarily decreased the neural response to S1 in the stimulated region. This decrease in activity was primarily observed during haptic S1 processing in left LO with only a weak trend of deactivation for visual S1. The suppressive effect of rTMS on haptic processing involved the whole LOC and pSTS/MTG bilaterally, indicating a spread of the suppressive effect of rTMS to other posterior cortical areas presumably via cortico-cortical connections. Together, the findings show that rTMS to the left LO selectively suppressed haptic processing of S1 but not S2 in the stimulated LO. This context-dependent effect on haptic processing suggests that 1 Hz rTMS primarily suppressed regional neural activity in the left LO related to more exploratory haptic processing (S1) without affecting a more comparative processing (S2) of objects in a delayed match-to-sample context.

Using the current design (Kassuba et al., 2013a) or an analogous design with auditory and haptic stimuli (Kassuba et al., 2013b), we have previously reported a dissociation between S1 and S2 processing related to an adaptation of the BOLD response due to the repeated presentation of objects with the same identity over the duration of the experiment. Only S1 encoding but not S2 matching showed reduced responses as a function of how often an object had been already presented throughout the experiment. We speculate that S1 encoding and S2 matching represent distinct functional states, the former might be more bottom-up driven while the latter might be more top-down dependent. As a consequence, left LO conditioning leads to different reorganizational changes.

Table 3 | Real rTMS induced incongruency effects (incongruent > congruent × real rTMS > sham rTMS).

Region	*x*	*y*	*z*	t_{peak}	p_{corr} (p_{uncorr})	
TEMPORALLY STABLE EFFECTS: POST-REAL rTMS > POST-SHAM rTMS						
Haptic S2—crossmodal: VHi > VHc						
No significant results						
Haptic S2—unimodal: HHi > HHc						
No significant results						
Visual S2—crossmodal: HVi > VVc						
No significant results						
Visual S2—unimodal: VVi > VVc						**>haptic S2**
L superior medial gyrus	3	32	52	4.92	0.014	(<0.001)
L FG	−36	−37	−26	4.55	<0.001[§]	0.039[§]
L pIPS	−30	−61	55	3.81	0.002[§]	0.092[§]
R pIPS	27	−58	52	3.73	0.003[§]	0.050[§]
LINEAR TIME DEPENDENT EFFECTS: POST-REAL rTMS > POST-SHAM rTMS						
Haptic S2—crossmodal: VHi > VHc × time > visual S2						**>visual S2**
R anterior parahippocampus	24	−1	−29	5.60	0.002	(<0.001)
R middle frontal gyrus	39	−1	58	5.61	0.002	0.004
Haptic S2—unimodal: HHi > HHc × time						
No significant results						
Visual S2—crossmodal: VHi > VHc × time						
No significant results						
Visual S2—unimodal: VVi > VVc × time						
No significant results						

Coordinates are denoted by x, y, z in mm (MNI space) and indicate the peak voxel. The last column shows the direct comparison of the respective effect to the corresponding effect of the other S2 modality condition. Strength of activation is expressed in t- and p-values corrected for the whole brain and uncorrected p-values in parentheses, respectively, at peak voxel (df = 255 for temporally stable effects, df = 119 for time-dependent effects), § small volume corrected. FG, fusiform gyrus, pIPS, posterior intraparietal sulcus. L, left; R, right.

INCREASED RESPONSES TO S2 IN REMOTE REGIONS

While processing of S2 was unchanged at the site of stimulation, transient increases in activation emerged in remote regions after real relative to sham rTMS in congruent crossmodal matching trials, that is, when object concepts were most likely integrated across the senses (Laurienti et al., 2004). These putatively compensatory increases in activation were found in temporal regions such as the left temporal pole and pSTS/MTG for crossmodal matching of visual S2 (HVc) and the left FG and right anterior parahippocampus for crossmodal matching of haptic S2 (VHc) and were specific for the respective S2 modality. It has been previously proposed that the temporal cortex integrates object information (e.g., object motion, shape, use-associated motor movements) with increasing convergence and abstraction along the posterior to anterior axis (Martin and Chao, 2001; Martin, 2007). For instance, studies that used dynamic visual and auditory object stimuli suggested that the pSTS/MTG is tuned to features of motion associated with different objects (Beauchamp et al., 2002, 2004). We have previously shown that the same left FG region as found here shows object-specific responses independent of whether objects were seen, heard, or touched, suggesting more abstract or conceptual representations of object information (Kassuba et al., 2011, 2013b; see also Martin, 2007). Patient studies suggest that the anterior temporal pole is critical for semantic memory (Rogers et al., 2006) and particularly for retrieving object information about unique entities (Damasio, 1989; Damasio et al., 1996). We, therefore, propose that in the presence of a functional perturbation of the left LO, regions of a semantic object recognition network are increasingly activated when the same objects are matched across vision and haptics. These enhanced activations might reflect a compensatory strategy involving semantic memory. Critically, retrieving haptic object information and matching it to the same object processed visually activated different nodes of this putative network than retrieving visual object information and matching it to the same object processed haptically.

One likely explanation for the null findings with respect to real rTMS effects on multisensory interactions is that the delayed-match-to-sample task was not challenging enough. Even after real rTMS, task accuracy was nearly perfect (≥95%). We found real rTMS-induced increases in activations in LOC and IPS related to matching of incongruent objects, which was behaviorally more difficult than matching of congruent objects. Some of these increased brain activations were specifically stronger during incongruent than congruent matching (incongruency effect), only after real but not sham rTMS. Again, these rTMS-induced incongruency effects differed based on the S2 modality: The effects were limited to the first 30 min post rTMS in the right anterior parahippocampus for crossmodal haptic matching of haptic S2 (VH) but remained stable throughout the session in left FG and bilateral pIPS for unimodal matching of visual S2 (VV). Even though left LO rTMS had no effects on multisensory

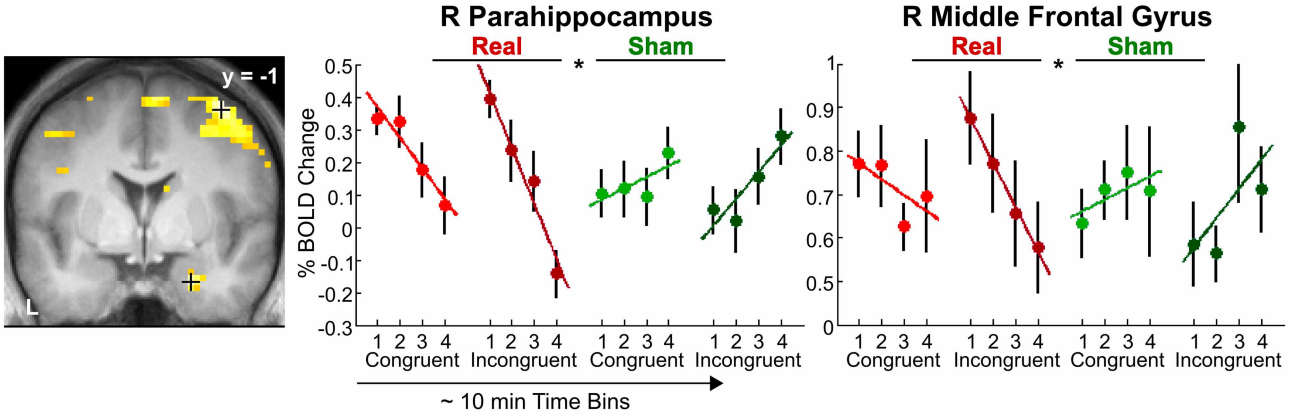


FIGURE 6 | Transient incongruency effects for crossmodal matching of haptic S2 (VH) evoked by rTMS conditioning. Regional activity in the right parahippocampus (MNI coordinates: *x*, *y*, *z* = 24, −1, 29) and right middle frontal gyrus (39, −1, 58) showed an interaction of linear time-dependent effects by rTMS condition that was stronger for incongruent than congruent trials: Whereas regional activity was initially increased and linearly decreased over time after real rTMS (red, real-congruent; dark red, real-incongruent), no significant linear time-dependent increases in activations (or rather decreases) were found after sham rTMS (green, sham-congruent; dark green, sham-incongruent), and these differential effects were stronger for incongruent than congruent conditions (real > sham × incongruent > congruent × time). For illustrative purposes, the statistical maps are thresholded at $p < 0.001$, uncorrected, and overlaid on the average structural T1-weighted image of all subjects. Each time bin represents ~10 min and 4–7 trials. L, left; R, right. $^*p < 0.05$, corrected.

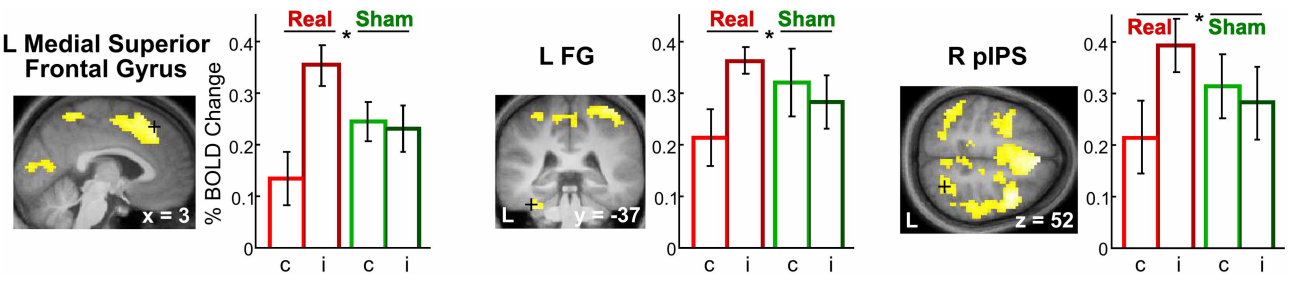


FIGURE 7 | Incongruency effects for unimodal visual matching (VV) evoked by rTMS conditioning. Only after real but not sham rTMS, regional activity in the L superior medial gyrus (MNI coordinates: *x*, *y*, *z* = 3, 32, 52) **(left)**, L FG (−36, −37, −26) **(middle)**, and R pIPS (27, −58, 52) **(right)** was increased when a visual S2 was matched to an incongruent (i) as compared to a congruent (c) visual S1 (real > sham × incongruent > congruent). The same effects were found in the left pIPS. For illustrative purposes, the statistical maps are thresholded at $p < 0.001$, uncorrected, and overlaid on the average structural T1-weighted image of all subjects. Bars represent percent signal change and error bars the standard error of the mean for each VV condition after real and sham rTMS (red, real-congruent; dark red, real-incongruent; green, sham-congruent; dark green, sham-incongruent). FG, fusiform gyrus; pIPS, posterior intraparietal sulcus; L, left; R, right. $^*p < 0.05$, corrected.

interactions, these results suggest a functional relevance of left LO for evaluating visual and haptic object information.

Since S1 and S2 were presented sequentially, incongruency effects (incongruent > congruent) could also be interpreted as repetition suppression or fMRI-adaptation (fMRI-A) effects (i.e., decreased activity in the congruent condition due to the repeated presentation of objects with the same identity). Thus, incongruency effects found for crossmodal matching could be interpreted as crossmodal adaptation and might indicate multisensory integration (cf. Tal and Amedi, 2009; Doehrmann et al., 2010; Van Atteveldt et al., 2010). However, we argue that the task demands in our paradigm have overruled general effects of stimulus habituation (for a detailed discussion of this issue, see Kassuba et al., 2013a,b). First, the stimulus onset asynchronies between S1 and S2 in the present study were rather long and favored a semantic encoding of S1. Second, while other studies showing adaptation effects typically used a task orthogonal to the effect of interest such as a detection task (Doehrmann et al., 2010; Van Atteveldt et al., 2010; Snow et al., 2013) or passive recognition (Tal and Amedi, 2009), our task required an explicit semantic decision on the identity of S1 and S2. In addition, using this delayed-match-to-sample paradigm, we did not find any general adaptation of the BOLD response to S2 due to repeated presentations of the same objects throughout the experiment (independent of matching condition), neither when using visual and haptic stimuli (Kassuba et al., 2013a), nor when using auditory and haptic stimuli (Kassuba et al., 2013b). Consistent with our findings, other studies employing longer delays in visuo-haptic priming (James et al., 2002) or using a delayed-match-to-sample task (Grefkes et al., 2002) have found enhanced instead of decreased

BOLD responses in LO and IPS to crossmodal matching. Thus, the transient rTMS-induced incongruency effects for crossmodal matching of haptic S2 most likely reflect an increased response to incongruent stimuli after real rTMS. We speculate that this increase is due to compensatory activations that help to maintain task performance in the behaviorally most challenging condition.

METHODOLOGICAL CONSIDERATIONS

The null effects of rTMS with respect to behavioral performance and multisensory interactions have to be interpreted in light of the stimulus paradigm and applied rTMS stimulation protocol. In addition to semantic congruency, temporal and spatial coherence are important factors for multisensory integration (Stein and Stanford, 2008). We presented crossmodal stimuli sequentially (instead of simultaneously) and in different positions with respect to the subjects' egocentric spaces (visual: mirror on head coil, haptic: on the subjects waist). The delayed-match-to-sample task enabled us to identify a differential contribution of vision and haptics to visuo-haptic interactions and guaranteed that objects were processed conceptually. Therefore, our paradigm rather probed visuo-haptic interactions in higher-order object recognition than basic visuo-haptic integration. In addition, behavioral performance was at ceiling. Thus, the delayed-match-to-sample task might have not been sensitive enough to identify rTMS effects on multisensory interactions, behaviorally or neurally.

Previous studies in which LO TMS had been found to impair visual object processing have used different tasks and applied TMS "online" (i.e., while participants performed the task). For example, Ellison and Cowey (2006) used discrimination tasks with simultaneously presented shapes and applied a high-frequency five-pulse train at stimulus onset. In the study by Pitcher et al. (2009), subjects performed a delayed-match-to-sample task as well, although with shorter presentation times (500ms S1 + 500 ms mask + 500 ms S2) than in the present study and TMS was applied to the right LO. In that study, the online administration of a 10 Hz TMS train was aligned with S2. These studies have applied TMS online during the task and not as a conditioning offline protocol as we did in the present study.

It is important to recall that the effects of online and offline rTMS are not the same (Siebner and Rothwell, 2003; Siebner et al., 2009b). With its prolonged effects on cortical excitability, offline rTMS induces a complex reorganization and re-weighting of the involvement of cortical structures in task relevant networks (Siebner et al., 2009a). The system may adapt to the rTMS-induced changes to maintain functional homeostasis. Effects of rTMS conditioning on behavior are typically reported in the first 15 min post rTMS (cf. Rounis et al., 2006; O'Shea et al., 2007; Mancini et al., 2011), while effects on neuronal activity can be measured up to 1 h post rTMS (Siebner et al., 2003). Our fMRI measurement started on average 10 min post rTMS. There are previous studies that found changes in neuronal activity at the stimulated region and in remote regions after 1 Hz rTMS conditioning without affecting behavior later than 10 min post rTMS (Lee et al., 2003; O'Shea et al., 2007). Therefore, the lack of behavioral impairment but task-related changes in cortical activity found in the current study could be interpreted as functional reorganization preserving behavior after neuronal challenge.

It is possible that we would have found behavioral effects if fMRI had started earlier within the first 10 min post rTMS. The rTMS-related effects might have been stronger if we had used individual activations from the localizer as rTMS target regions instead of the peak response from the group analysis. However, individual peak responses were close to the group peak. Further, results did not change when we excluded 5 subjects from the analyses that showed only weak visuo-haptic convergence in the left LO during the localizer.

CONCLUSIONS

The fact that we found distinct effects for different S2 matching conditions supports the idea that these reflect compensatory mechanisms provoked by task demands rather than mere transsynaptic spreading of rTMS conditioning. Together, the results support the notion that the left LO is functionally relevant for both visual and haptic object recognition but to a different extent. Our data suggest that visuo-haptic object recognition involves a network of regions comprising the bilateral LO, FG, aIPS, pIPS, pSTS/MTG, and anterior temporal regions, which can be flexibly recruited if the system is challenged. How compensatory processing is allocated depends on the target modality (visual vs. haptic) and task demands (S1 encoding vs. S2 matching).

AUTHOR CONTRIBUTIONS

Tanja Kassuba and Hartwig R. Siebner conceived the experiment. Tanja Kassuba, Cordula Hölig, and Hartwig R. Siebner designed the experiment. Tanja Kassuba and Corinna Klinge collected the data. Tanja Kassuba analyzed the data. All authors interpreted the results and wrote the paper.

ACKNOWLEDGMENTS

The authors would like to thank Gesine Müller, Katrin Müller, Kathrin Wendt, and Karolina Müller for their support with acquiring the MRI data and Mareike Menz for helpful suggestions on the data analysis. The study was funded by BMBF (grant 01GW0562 to Hartwig R. Siebner). Tanja Kassuba was supported by the Janggen-Pöhn Stiftung, the Jubiläumsstiftung der BLKB, and by a postdoctoral fellowship from the Swiss National Science Foundation. Hartwig R. Siebner was supported by a Grant of Excellence on the control of actions "ContAct" from the Lundbeck Foundation (grant no. R59 A5399).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00274/abstract

REFERENCES

Amedi, A., Jacobson, G., Hendler, T., Malach, R., and Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cereb. Cortex* 12, 1202–1212. doi: 10.1093/cercor/12.11.1202

Amedi, A., Malach, R., Hendler, T., Peled, S., and Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330. doi: 10.1038/85201

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689. doi: 10.1038/nn1912

Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321. doi: 10.1111/j.2044-8295.1970.tb01248.x

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., et al. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29, 529–535. doi: 10.1016/S0896-6273(01)00224-0

Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/S0896-6273(04)00070-4

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159. doi: 10.1016/S0896-6273(02)00642-6

Beauchamp, M. S., Yasar, N. E., Frye, R. E., and Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage* 41, 1011–1020. doi: 10.1016/j.neuroimage.2008.03.015

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33, 25–62. doi: 10.1016/0010-0277(89)90005-X

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. doi: 10.1038/380499a0

Doehrmann, O., Weigelt, S., Altmann, C. F., Kaiser, J., and Naumer, M. J. (2010). Audiovisual functional magnetic resonance imaging adaptation reveals multisensory integration effects in object-related sensory cortices. *J. Neurosci.* 30, 3370–3379. doi: 10.1523/JNEUROSCI.5074-09.2010

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Ellison, A., and Cowey, A. (2006). TMS can reveal contrasting functions of the dorsal and ventral visual processing streams. *Exp. Brain Res.* 175, 618–625. doi: 10.1007/s00221-006-0582-8

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a

Feinberg, T. E., Rothi, L. J., and Heilman, K. M. (1986). Multimodal agnosia after unilateral left hemisphere lesion. *Neurology* 36, 864–867. doi: 10.1212/WNL.36.6.864

Gerschlager, W., Siebner, H. R., and Rothwell, J. C. (2001). Decreased corticospinal excitability after subthreshold 1 Hz rTMS over lateral premotor cortex. *Neurology* 57, 449–455. doi: 10.1212/WNL.57.3.449

Gläscher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics* 7, 73–82. doi: 10.1007/s12021-008-9042-x

Goodale, M. A., Milner, A. D., Jakobson, L. S., and Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature* 349, 154–156. doi: 10.1038/349154a0

Grefkes, C., Weiss, P. H., Zilles, K., and Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron* 35, 173–184. doi: 10.1016/S0896-6273(02)00741-9

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187–203. doi: 10.1016/S0896-6273(00)80832-6

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. doi: 10.1038/77754

Helbig, H. B., and Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Exp. Brain Res.* 179, 595–606. doi: 10.1007/s00221-006-0814-y

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., and Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40, 1706–1714. doi: 10.1016/S0028-3932(02)00017-9

Karnath, H. O., Ruter, J., Mandler, A., and Himmelbach, M. (2009). The anatomy of object recognition–visual form agnosia caused by medial occipitotemporal stroke. *J. Neurosci.* 29, 5854–5862. doi: 10.1523/JNEUROSCI.5192-08.2009

Kassuba, T., Klinge, C., Hölig, C., Menz, M. M., Ptito, M., Röder, B., et al. (2011). The left fusiform gyrus hosts trisensory representations of manipulable objects. *Neuroimage* 56, 1566–1577. doi: 10.1016/j.neuroimage.2011.02.032

Kassuba, T., Klinge, C., Holig, C., Roder, B., and Siebner, H. R. (2013a). Vision holds a greater share in visuo-haptic object recognition than touch. *Neuroimage* 65, 59–68. doi: 10.1016/j.neuroimage.2012.09.054

Kassuba, T., Menz, M. M., Roder, B., and Siebner, H. R. (2013b). Multisensory interactions between auditory and haptic object recognition. *Cereb. Cortex* 23, 1097–1107. doi: 10.1093/cercor/bhs076

Kourtzi, Z., and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science* 293, 1506–1509. doi: 10.1126/science.1061133

Lacey, S., and Sathian, K. (2011). Multisensory object representation: insights from studies of vision and touch. *Prog. Brain Res.* 191, 165–176. doi: 10.1016/B978-0-444-53752-2.00006-0

Lancaster, J. L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., et al. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* 28, 1194–1205. doi: 10.1002/hbm.20345

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., and Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158, 405–414. doi: 10.1007/s00221-004-1913-2

Lee, L., Siebner, H. R., Rowe, J. B., Rizzo, V., Rothwell, J. C., Frackowiak, R. S., et al. (2003). Acute remapping within the motor system induced by low-frequency repetitive transcranial magnetic stimulation. *J. Neurosci.* 23, 5308–5318. Available online at: http://www.jneurosci.org/content/23/12/5308.long.

Malach, R., Levy, I., and Hasson, U. (2002). The topography of high-order human object areas. *Trends Cogn. Sci.* 6, 176–184. doi: 10.1016/S1364-6613(02)01870-3

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8135–8139. doi: 10.1073/pnas.92.18.8135

Mancini, F., Bolognini, N., Bricolo, E., and Vallar, G. (2011). Cross-modal processing in the occipito-temporal cortex: a TMS study of the Muller-Lyer illusion. *J. Cogn. Neurosci.* 23, 1987–1997. doi: 10.1162/jocn.2010.21561

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143

Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201. doi: 10.1016/S0959-4388(00)00196-3

Morin, P., Rivrain, Y., Eustache, F., Lambert, J., and Courtheoux, P. (1984). [Visual and tactile agnosia]. *Rev. Neurol. (Paris).* 140, 271–277.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

O'Shea, J., Johansen-Berg, H., Trief, D., Gobel, S., and Rushworth, M. F. (2007). Functionally specific reorganization in human premotor cortex. *Neuron* 54, 479–490. doi: 10.1016/j.neuron.2007.04.021

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., and Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curr. Biol.* 19, 319–324. doi: 10.1016/j.cub.2009.01.007

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., et al. (2006). Anterior temporal cortex and semantic memory: reconciling findings from neuropsychology and functional imaging. *Cogn. Affect. Behav. Neurosci.* 6, 201–213. doi: 10.3758/CABN.6.3.201

Rounis, E., Stephan, K. E., Lee, L., Siebner, H. R., Pesenti, A., Friston, K. J., et al. (2006). Acute changes in frontoparietal activity after repetitive transcranial magnetic stimulation over the dorsolateral prefrontal cortex in a cued reaction time task. *J. Neurosci.* 26, 9629–9638. doi: 10.1523/JNEUROSCI.2657-06.2006

Saito, D. N., Okada, T., Morita, Y., Yonekura, Y., and Sadato, N. (2003). Tactile-visual cross-modal shape matching: a functional MRI study. *Brain Res. Cogn. Brain Res.* 17, 14–25. doi: 10.1016/S0926-6410(03)00076-4

Siebner, H. R., Bergmann, T. O., Bestmann, S., Massimini, M., Johansen-Berg, H., Mochizuki, H., et al. (2009a). Consensus paper: combining transcranial stimulation with neuroimaging. *Brain Stimul.* 2, 58–80. doi: 10.1016/j.brs.2008.11.002

Siebner, H. R., Filipovic, S. R., Rowe, J. B., Cordivari, C., Gerschlager, W., Rothwell, J. C., et al. (2003). Patients with focal arm dystonia have increased sensitivity to slow-frequency repetitive TMS of the dorsal premotor cortex. *Brain* 126, 2710–2725. doi: 10.1093/brain/awg282

Siebner, H. R., Hartwigsen, G., Kassuba, T., and Rothwell, J. C. (2009b). How does transcranial magnetic stimulation modify neuronal activity in

the brain? Implications for studies of cognition. *Cortex* 45, 1035–1042. doi: 10.1016/j.cortex.2009.02.007

Siebner, H. R., and Rothwell, J. (2003). Transcranial magnetic stimulation: new insights into representational cortical plasticity. *Exp. Brain Res.* 148, 1–16. doi: 10.1007/s00221-002-1234-2

Snow, J. C., Goodale, M. A., and Culham, J. C. (2012). "The lateral occipital complex is not necessary for haptic shape representation," in *Talk Presented at the Annual Meeting of the Society for Neuroscience* (New Orleans, LA).

Snow, J. C., Strother, L., and Humphreys, G. W. (2013). Haptic shape processing in visual cortex. *J. Cogn. Neurosci.* 26, 1–14. doi: 10.1162/jocn_a_00548

Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331

Stilla, R., and Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Hum. Brain Mapp.* 29, 1123–1138. doi: 10.1002/hbm.20456

Stokes, M. G., Chambers, C. D., Gould, I. C., Henderson, T. R., Janko, N. E., Allen, N. B., et al. (2005). Simple metric for scaling motor threshold based on scalp-cortex distance: application to studies using transcranial magnetic stimulation. *J. Neurophysiol.* 94, 4520–4527. doi: 10.1152/jn.00067.2005

Tal, N., and Amedi, A. (2009). Multisensory visual-tactile object related network in humans: insights gained using a novel crossmodal adaptation approach. *Exp. Brain Res.* 198, 165–182. doi: 10.1007/s00221-009-1949-4

Talairach, J., and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain.* New York, NY: Thieme.

Van Atteveldt, N. M., Blau, V. C., Blomert, L., and Goebel, R. (2010). fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC Neurosci.* 11:11. doi: 10.1186/1471-2202-11-11

Williams, M. A., Dang, S., and Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nat. Neurosci.* 10, 685–686. doi: 10.1038/nn1900

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 January 2014; accepted: 14 March 2014; published online: 02 April 2014.
Citation: Kassuba T, Klinge C, Hölig C, Röder B and Siebner HR (2014) Short-term plasticity of visuo-haptic object recognition. Front. Psychol. ***5****:274. doi: 10.3389/fpsyg.2014.00274*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Kassuba, Klinge, Hölig, Röder and Siebner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

REVIEW ARTICLE
published: 17 June 2014
doi: 10.3389/fpsyg.2014.00538

Cortical processing of object affordances for self and others' action

Monica Maranesi[1]*, Luca Bonini[1] and Leonardo Fogassi[2]

[1] *Brain Center for Social and Motor Cognition, Italian Institute of Technology, Parma, Italy*
[2] *Department of Neuroscience, University of Parma, Parma, Italy*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Patrizia Fattori, University of Bologna, Italy
Anna M. Borghi, University of Bologna, Italy

****Correspondence:***
Monica Maranesi, Brain Center for Social and Motor Cognition, Istituto Italiano di Tecnologia, Via Volturno 39, 43125 Parma, Italy
e-mail: monica.maranesi@iit.it

The perception of objects does not rely only on visual brain areas, but also involves cortical motor regions. In particular, different parietal and premotor areas host neurons discharging during both object observation and grasping. Most of these cells often show similar visual and motor selectivity for a specific object (or set of objects), suggesting that they might play a crucial role in representing the "potential motor act" afforded by the object. The existence of such a mechanism for the visuomotor transformation of object physical properties in the most appropriate motor plan for interacting with them has been convincingly demonstrated in humans as well. Interestingly, human studies have shown that visually presented objects can automatically trigger the representation of an action provided that they are located within the observer's reaching space (peripersonal space). The "affordance effect" also occurs when the presented object is outside the observer's peripersonal space, but inside the peripersonal space of an observed agent. These findings recently received direct support by single neuron studies in monkey, indicating that space-constrained processing of objects in the ventral premotor cortex might be relevant to represent objects as potential targets for one's own or others' action.

Keywords: perception, space, sensorimotor transformation, visual streams, grasping

INTRODUCTION

Perception and action have been considered for a long time as two serially organized steps of processing, with the former relying on sensory brain areas and the latter implemented by the motor cortex. In this view, cognition would emerge as an intermediate step of information processing performed by associative cortical areas. This classical "sandwich model" (Hurley, 1998), in which perception and action do never directly interact one with the other, has been challenged by a growing body of evidence in the last three decades (see Goodale and Milner, 1992; Rizzolatti and Matelli, 2003). These studies suggest that a crucial role in perception is played by cortical motor regions as well, especially when sensory information is required for acting. An intriguing synthesis of this view maintains that "perception is not something that happens to us, or in us: It is something we do" (Noë, 2004).

The tight link of perceptual processes with the motor ones has a particularly elegant exemplification in the concept of "affordance", coined by the psychologist James Gibson (1979). According to Gibson, affordances are all the motor possibilities that an object in the environment offers an individual: crucially, they depend on the motor capabilities of the observer but not on his/her intentions or needs. Among the different possible affordances of an object, the one that will prevail and will be more likely turned into an overtly executed action depends upon the contextual situation, the goals and intentions of the perceiver. For example, a cup might afford grasping of its handle or of its body if one expects it contains a hot or cold drink, respectively. In addition, it might also afford grasping of its top, if it is empty and the agent wants simply to move it away. In all these cases, two types of parallel processing of the object take place: its semantic description, provided by higher order cortical visual areas, and a pragmatic description, which includes the extraction of its various affordances and micro-affordances (Ellis and Tucker, 2000), and their possible translation into action (Jeannerod et al., 1995).

Which are the cortical regions involved in the processing of objects affordances? Goodale and Milner (1992) modified the Ungerleider and Mishkin's proposal of the two visual streams (1982), suggesting that the "ventral stream", linking primary visual cortex to the inferotemporal regions, is responsible for object recognition, while the "dorsal stream", ending in the posterior parietal region, plays a crucial role in the sensorimotor transformations for visually guided object-directed actions. Based on clinical, functional and anatomical data, Rizzolatti and Matelli (2003) proposed to further subdivide the dorsal stream into two distinct functional systems, formed by partially segregated cortical pathways: the dorso-dorsal (d-d) and the ventro-dorsal (v-d) stream. According to their proposal, the d-d stream would correspond to the dorsal stream as previously defined by Milner and Goodale, exploiting sensory information for the control of reaching movements in space, while the v-d stream would be specifically involved in sensorimotor transformation for grasping, space perception and action recognition. Thus, also within the originally defined dorsal stream, there is a subsystem, the v-d stream, which might play a role in perceptual functions.

FROM OBJECT AFFORDANCES TO SENSORIMOTOR TRANSFORMATIONS: PARALLEL PARIETO-FRONTAL CIRCUITS

Object grasping is one of the most frequently performed and highly specialized behavior in primates (Jeannerod et al., 1995; Macfarlane and Graziano, 2009). One of the most challenging aspects in the control of grasping is the configuration of the hand according to the object features during the reaching phase (Jeannerod et al., 1995). Jeannerod (1984) and Arbib (1985), independently, proposed the existence of two specific neural systems responsible for the reaching and grasping components of reach-to-grasp actions. In the last decades, several studies on both humans and monkeys have been carried out in order to identify and describe the cortical mechanisms underlying such a complex sensorimotor transformation. While most of these studies aimed at clarifying the role of areas of the v-d stream, particularly of the anterior intraparietal area (AIP) and ventral premotor area F5, recent findings shed new light on the possible involvement of areas belonging to the dorso-dorsal stream (parietal area V6A and dorsal premotor area F2) in the visuomotor transformations involved in grasping actions.

THE AIP-F5 CIRCUIT

From the early '90s, Sakata and colleagues have investigated monkey parietal cortex by means of a paradigm designed to study neuronal activity while the monkey had to observe and subsequently grasp objects of different size and shape (Taira et al., 1990; Sakata et al., 1995; Murata et al., 2000). This condition could be performed in the light or in the dark, in separate sessions. Moreover, the task also included a condition in which the monkey had to simply fixate the object, without performing any grasping movement. The authors were able to describe, as in their previous studies, two types of visually-modulated neurons: "visual-dominant" neurons, which discharged during grasping in the light but not in the dark, and "visual-motor" neurons, which fired also during grasping in the dark, although weaker compared with the same action performed in the light. Within both these two populations of neurons, they further subdivided neurons in "object-type" or "non-object-type," depending on whether or not they responded to object presentation during the fixation task. Interestingly, the discharge of many object-type neurons exhibited the same preference for a given object (or set of objects) during both object fixation and grasping. This finding suggests that object-type neurons play a crucial role in the visuomotor transformation of object affordances in the most appropriate hand shape for grasping. Their response and the preserved object selectivity, also during trials in which the monkey did not perform any action, further indicate that the neural mechanisms for the extraction of object affordances rely on the monkey motor possibilities, but not necessarily on its actual execution of a grasping action. Therefore, also the dorsal pathway (in particular the ventro-dorsal stream), appears to play a role in object perception.

Another study demonstrated a causal role of area AIP in computing object properties for adjusting the finger posture according to the size and shape of the target object (Gallese et al., 1994). In this study, muscimol (a GABA-agonist which inhibits neurons activity) was injected in monkey area AIP, showing that while the arm reaching component was unimpaired, the hand shaping for grasping objects, particularly the small ones, was clearly altered, and associated with a reduced movement speed. The affected grip could be subsequently corrected by the monkey based on tactile exploration of the target object, suggesting that the deficit specifically concerns the visuomotor transformation for hand grasping.

What is the anatomo-functional mechanism through which the perceptual description of an object accesses the motor representations necessary for turning it into the most appropriate hand shape? Anatomical studies based on tracers injections in AIP have shown that this area is linked to many others through monosynaptic connections. In particular, they showed that area AIP forms an anatomo-functional module with the ventral premotor area F5 (Luppino et al., 1999; Borra et al., 2008).

Neurophysiological studies showed that area F5 contains neurons discharging during specific goal-related motor acts (Rizzolatti et al., 1988). Moreover, similarly to area AIP, F5 visuomotor neurons discharge to the visual presentation of graspable objects, often with a clear selectivity for their size and shape (Murata et al., 1997; Raos et al., 2006). These neurons have been defined as "canonical" neurons (Rizzolatti and Fadiga, 1998). Interestingly, both during object fixation and grasping in the dark, F5 neurons maintained the same selectivity for a given object or set of objects (Raos et al., 2006), reflecting a visuomotor matching mechanism as the one previously described for area AIP. In contrast to area AIP, however, no F5 neurons were recorded discharging only during grasping in the light and also to object presentation. In addition, while AIP visual responses to objects appear to encode the geometrical features shared by the different objects (Murata et al., 2000), F5 visual responses reflect the parameters of hand configuration shared by different types of grip (Raos et al., 2006). In line with these findings, muscimol inactivation of the F5 sector buried in the bank of the arcuate sulcus (F5p—Belmalih et al., 2009), which is more tightly linked with area AIP than F5 convexity (Luppino et al., 1999; Borra et al., 2008), produced a markedly impaired shaping of the hand during grasping (Fogassi et al., 2001). In particular, monkey were unable to produce the fingers configuration appropriate for the size and shape of the to-be-grasped object and, similarly to what previously described following inactivation of area AIP, the monkey could accomplish object grasping only by means of tactile feedback obtained through hand-object exploration.

Human studies revealed the existence of a putative homolog of monkey's area AIP in the anterior portion of the intraparietal sulcus (aIPS—Culham et al., 2003; Frey et al., 2005), which becomes specifically active during visually guided grasping. Interestingly, studies using TMS applied to aIPS reported a disruption of goal-dependent kinematics during reach-to-grasp trials (Tunik et al., 2005). In particular, this study reported that, depending on which parameter had to be controlled in the ongoing trial (object size or orientation), TMS pulse delivered to aIPS specifically disrupted the online control of the correspondent parameters of hand kinematics. Importantly, this effect was selectively produced by stimulation of aIPS and not of other parietal regions. The anatomo-functional connectivity between AIP and ventral premotor (PMv, considered the human homolog of area F5) has

been demonstrated also in humans by a TMS study (Davare et al., 2010). These authors induced an AIP virtual lesion by means of repetitive TMS. At the same time, they studied with another (paired-pulse) combined TMS technique the possible facilitation exerted by the ventral premotor (PMv) on the primary motor (M1) cortex. The results clearly indicated that PMv-M1 interactions during grasping are driven by information about object properties provided by AIP, demonstrating the existence of a causal transfer of information on object features between the human parietal (AIP) and the premotor (PMv) nodes of the visuomotor transformation network.

AREA V6A-F2 CIRCUIT

The parieto-frontal circuit formed by area V6A (Galletti et al., 1999; Fattori et al., 2001, 2005), and dorsal premotor area F2vr (Raos et al., 2003) constitutes a subdivision of the dorsal visual pathway (Galletti et al., 2003), deemed to play a role in the encoding of the arm direction toward different locations in space. Surprisingly, recent studies have demonstrated that the neural code of this circuit is not limited to reaching movements.

Indeed, area V6A also contains neurons modulated by wrist orientation (Fattori et al., 2009) and by hand shape (Fattori et al., 2010) during object grasping. In addition, single V6A neurons have been described responding also to the visual presentation of real objects (Fattori et al., 2012). In this latter study, the authors tested single neurons responses to object presentation within two different task contexts, similar to those previously employed to test AIP and F5 visuomotor neurons, namely: a passive "object viewing task," in which the monkey had to passively fixate the visually presented object, and a "reach-to-grasp task," in which object presentation was followed by object grasping. Results showed that 60% of area V6A neurons discharged to the presentation of objects, regardless of the task context. In addition, about half of them showed a preferential discharge for a particular object or set of objects. Although AIP and V6A neurons appear to be similar in this respect, two important differences emerged from this comparison. First, a greater number of AIP than V6A neurons showed object selectivity (45 vs. 25%, respectively). Second, while AIP visual responses encoded the geometric features shared by the observed objects, both during passive fixation and grasping tasks (Murata et al., 2000), object coding by V6A neurons showed an interesting interaction with the task context: in the object viewing task, V6A neurons encoded objects geometric features, like those of AIP, while during the reach-to-grasp task V6A neurons' responses reflected the features of the grip used for grasping a certain set of object, regardless of their geometric similarity.

Further studies revealed that neuronal activity in area V6A can also specify object position with high specificity for the peripersonal (reachable) space not only during reaching tasks (Fattori et al., 2001, 2005; Hadjidimitrakis et al., 2013), but also during passive fixation tasks (Hadjidimitrakis et al., 2011). In particular, Hadjidimitrakis et al. (2013) investigated object position coding according to different reference frames. In this study, the monkey had to reach a spot of light located at different distances and lateralities from the body, with its hand starting at two different initial positions (near to or far from the body). Results showed that the majority of V6A neurons encoded reach targets mainly based on a body-centered frame of reference or combined with information relative to the hand position.

Taken together, these findings suggest that both object features and its spatial position are encoded by V6A neurons, very likely playing a role in turning perceptual representations of geometrical and spatial properties of objects into the appropriate motor plans for interacting with them. In this respect, V6A contribution appears to be quite similar to that of area AIP. However, differently from AIP, area V6A has no direct anatomical connections with areas of the ventral visual stream (Gamberini et al., 2009; Passarelli et al., 2011), suggesting that it might play a more relevant role in monitoring the ongoing visuomotor transformations during reaching-grasping movements. The rapid recovery from reaching and grasping deficits produced by V6A bilateral lesions (Battaglini et al., 2002) is in line with this view. Area V6A is also strongly connected with the dorsal premotor area F2 (Matelli et al., 1998), thus forming a parieto-frontal circuit similar to the AIP-F5 one. Area F2 has been shown to play a role in the encoding of object features (Raos et al., 2004), as well as in specifying object location relative to the monkey's peri- or extrapersonal space (Fogassi et al., 1999). In particular, Raos et al. (2004) have investigated the possible role of neurons in the ventral part of area F2 (F2vr) in encoding object within the peripersonal (reaching) space by employing the same paradigm previously used to test F5 visuomotor neurons. Interestingly, the results evidenced that several visually responsive F2vr visuomotor neurons displayed object-selective visual responses congruent with their selectivity shown during reaching-grasping execution. The presence of slightly similar visuomotor properties in areas V6A and AIP, on one side, and F2vr and F5, on the other, is in line with the evidence that these pairs of areas have some reciprocal anatomical connections (Borra et al., 2008; Gamberini et al., 2009; Gerbella et al., 2011), indicating that the ventral and dorsal aspects of the dorsal stream are not completely segregated. Indeed, these findings support the idea that the V6A-F2vr circuit can process both object intrinsic (shape and size) and extrinsic (spatial location) features, thus extending to areas belonging to the dorsal visual stream (Galletti et al., 2003; Rizzolatti and Matelli, 2003) the functions of encoding object features and of monitoring object-directed actions.

Although the homology between monkey and human posterior parietal areas remains not completely clear (Silver and Kastner, 2009), recent indirect evidence suggest that object features, as well as their location in space, might be processed along the dorsal pathway not only for motor purposes. For example, Gallivan et al. (2009) showed that a reach-related area in the superior parieto-occipital cortex in human was more activated for objects located in the peripersonal space, even when passively observed. Another study evidenced that posterior parietal cortex activated during visual processing of objects not only when no action planning was involved, but even when the subjects' attention was drawn away from the stimuli (Konen and Kastner, 2008). In the same study, the top stages of both ventral and dorsal streams showed considerable invariance of their activation in relation to changes in stimulus features such as size and viewpoint, which generally affects the lower stages of both streams. More interestingly, activations in both the ventral and the dorsal

stream during the presentation of three-dimensional shapes have been reported with fMRI even in anesthetized monkeys (Sereno et al., 2002). Together with an increasing number of studies (Xu and Chun, 2009; Zachariou et al., 2014) on cortical object processing, these findings suggest that object information in the dorsal pathways is not only processed with the purpose of guiding or monitoring sensorimotor transformations, but can also play some role in perceptual and cognitive functions.

VISUOMOTOR TRANSFORMATION OR SENSORIMOTOR ASSOCIATION? PRAGMATIC AND PERCEPTUAL FUNCTIONS IN OBJECT PROCESSING

What happens exactly in the brain when we observe a graspable object? One possibility is that, as described above, a graspable object is represented pictorially in visual brain areas and, simultaneously, its pragmatic description (visuomotor transformation) is activated in areas of the v-d stream. Alternatively, neurons discharging at the sight of a real object might simply reveal that a visuomotor association did occur, likely irrespective of the specific physical properties of the object itself. Based on this latter view, one would predict that both seeing the real object and an arbitrary cue signal (e.g., a colored spot of light) previously associated to a specific grip posture, might evoke the same visuomotor response.

A recent study provides interesting data that directly address this issue. Baumann et al. (2009) recorded single neurons in area AIP of monkeys performing a delayed grasping task. During this task, monkeys were presented with a handle (target object) in different orientations, and a colored LED (cue signal), which instructed the animal to subsequently perform a power or a precision grip. Results showed that AIP neurons could represent both the handle orientation and the instructed grip type immediately after the presentation of the visual stimulus, indicating that AIP neurons can process object features in a context-dependent fashion. A modified version of the task (cue separation task) enabled to study neuronal responses also when information on object orientation and the required grip type were separately presented. In particular, when the target object was presented first, visuomotor neurons became active regardless of the preference for power or precision grip that they exhibited in the delayed grasping task. In contrast, when the cue was presented first (and the object was not yet visible), this information was only weakly represented in area AIP, while it was strongly encoded thereafter, when the target object was revealed. Together with the data reviewed above (Sakata et al., 1995; Murata et al., 1996), these findings indicate that, besides transforming object properties into the appropriate grip type, AIP visuomotor neurons can also encode abstract information provided by any visual stimulus previously associated with a specific grip. However, both object- and context-driven transformations of visual information into an appropriate motor representation of a hand grip require that the object to be grasped be visible in front of the monkey. Thus, area AIP does not simply associate contextual visual stimuli with motor representations, but plays an active role in the processing of a pragmatic description of observed objects. Interestingly, even human fMRI studies showed that area AIP can activate during both the recognition and construction of three-dimensional shapes in the absence of visual guidance, but not during mental imagery of the same processes (Jancke et al., 2001), where overt sensory input and motor output are absent: this finding clearly supports the idea that the physical presence of the object is crucial for triggering area AIP neurons activity.

Do parallel processings of pictorial and pragmatic description of object features integrate or remain independent? Anatomical studies have demonstrated a rich pattern of connections linking temporal visual areas with inferior parietal regions belonging to the v-d stream (Borra et al., 2008, 2010). In addition, neurophysiological data on monkeys have revealed that a crucial aspect for both pictorial and pragmatic description of real objects—namely, their three-dimensional shape—is processed by both inferotemporal cortex (Janssen et al., 2000a,b) and area AIP (Srivastava et al., 2009; Verhoef et al., 2010). However, IT neurons' activity start shortly after the visual presentation while area AIP becomes active later on, leading some authors to suggest that the former plays a role in the formation of a perceptual decision and in the monkey behavioral choice; while the latter would reflect the three-dimensional features of the stimulus only after perceptual decision formation (Verhoef et al., 2010).

All the studies so far reviewed converge in indicating that (1) two cortical areas (IT and AIP) are involved in the parallel processing of the same information on objects (size, shape, etc), (2) they share some neuronal properties, and (3) are tightly interconnected one with the other. However, while part of the posterior parietal cortex, in particular area AIP, is devoted to extract object affordances for pragmatic purposes, the inferotemporal areas encode object features for object recognition. This latter conclusion somehow reminds a categorical, anatomo-functional distinction between perceptual and pragmatic functions of the "visual brain in action" (Milner and Goodale, 1993). However, it might be suggested that "*objects, as pictorially described by visual areas, are devoid of meaning. They gain meaning because of an association between their pictorial description (meaningless) and motor behavior (meaningful)*" (Rizzolatti and Gallese, 1997). Thus, in this view, although pragmatic and pictorial aspects of object processing might play partially distinct roles in mediating behavior within specific contexts, they would jointly contribute to our qualitative, phenomenological perceptual experience of the outside world. An interesting fMRI experiment on human subjects provides direct support to this claim. Grefkes et al. (2002) asked human volunteers to recognize whether an object was identical to another one previously assessed by the same subject. Objects were abstract three-dimensional solids differing one from the other only in size and shape (not weight, texture, etc.), and the two objects could be assessed and recognized either visually or by tactile manipulation. The results showed that human area AIP was specifically activated when cross-modal matching of visual and tactile object features was required, even when no specific motor act had to be performed on the perceived object, thus supporting the role of this area in the processing of multimodal information about object shape.

Noteworthy, the possible link between pragmatic and semantic cross-modal processing of object features is even more evident if one considers the network of areas connected with area AIP. On one side, AIP has reciprocal connections with a sector of

the secondary somatosensory cortex (Disbrow et al., 2003; Borra et al., 2008) which is particularly active during haptic exploration of objects (Krubitzer et al., 1995; Fitzgerald et al., 2004) and tactile object recognition (Reed et al., 2004). On the other, as already mentioned, AIP is connected with inferotemporal areas of the middle temporal gyrus, which convey semantic information on object identity (Borra et al., 2008). Thus, it is not surprising that cortical lesions involving AIP not only impair visually guided grasping (Gallese et al., 1994; Tunik et al., 2005), but also cause deficits in active tactile shape recognition, in the absence of (Valenza et al., 2001) or in association with (Reed and Caselli, 1994) tactile agnosia.

Taken together, all these data strongly indicate that AIP plays a crucial role in visuomotor transformation for visually- and somatosensory-guided manipulation of objects, but both pragmatic and pictorial information are involved in this process, likely contributing not only to the efficient organization of hand actions, but also to our phenomenological perceptual experience of objects.

SPACE-DEPENDENT CODING OF OBJECTS AFFORDING SELF OR OTHERS' ACTION

The studies so far reviewed demonstrate that seeing an object, such as an apple, simultaneously activates parallel neuronal representations of its pictorial features and motor affordances, providing a comprehensive perceptual experience of the object itself. However, several recent studies evidenced that affordances can be modulated by different contextual factors (Costantini et al., 2010, 2011a,b; Borghi et al., 2012; Ambrosini and Costantini, 2013; Kalenine et al., 2013; Van Elk et al., 2014) and, among these latter, one of the most crucial is represented by the space in which objects are located. Is an apple processed and perceived in the same way when it is at hand, on the table in front of me, as when it is out of reach, on the top of the apple tree?

According to Poincaré (1908), "it is in reference to our own body that we locate exterior objects, and the only special relations of these objects that we can picture to ourselves are their relations with our body." A similar idea has been expressed more recently by Gibson (1979), according to whom the abstract concept of space is only a conceptual achievement, while the perception of space is intimately linked with the guidance of our behavior in a crowded and cluttered environment. Thus, our capacity to act with our own body on the external world appears to be, theoretically, of crucial importance in establishing the way our brain process information on objects.

Although some previous behavioral studies in humans suggested that object affordances might not be influenced by the location in space of the observed object (Tucker and Ellis, 2001), recent behavioral (Costantini et al., 2010, 2011a; Ambrosini and Costantini, 2013) and TMS (Cardellicchio et al., 2013) studies suggest that the extraction of affordances and the recruitment of motor representations of graspable objects crucially depend on whether the object falls within the peripersonal, reachable space of the observer, in line with the classical philosophical and psychological models described above. While affordance effects are typically studied in relation to potential motor acts allowing one to approach and interact with an object, Anelli et al. (2013) demonstrated that potentially noxious objects (e.g., cactus, scorpio, broken bulb, etc.) induce an aversive affordance, which triggers in the observer's motor system the representation of escaping-avoidance reactions, particularly when the dangerous stimulus moves toward the observer's peripersonal space. Taken together, these findings support the idea that object processing is strictly related with the object spatial location, and that the peripersonal space is the most relevant source of information for affordance extraction.

According with the aforementioned concept of space, one would expect that the link between object affordances and the observer's peripersonal space relies on a pragmatic, rather than metric, reference frame. In other terms: is the physical distance of the object from the observer the crucial variable to gate affordance effect (metric representation) or does it depend on the observer's possibility to directly interact with the object (pragmatic representation)? The study by Costantini et al. (2010) addressed this issue by means of a behavioral paradigm exploiting the spatial alignment effect. In this study, subjects were visually presented with an object which could be located within or outside their peripersonal space, and the results evidenced the presence of an object affordance effect only when the object was located in the observer's peripersonal space. Crucially, if a transparent barrier was interposed between the subject and the object, although this latter was within the observer's peripersonal space (same metric distance), the affordance effect vanished as if the object were located in the extrapersonal space. Thus, the power of an object to automatically evoke potential motor acts appears to be strictly linked to the effective possibility of the onlooker to interact with it. Based on these findings, one would expect that seeing an object out-of-reach does not induce any activation of the observer's motor system, thus object perception should completely rely on posterior visual areas. In another behavioral study, Costantini et al. (2011b) replicated the finding that the affordance effect is evoked only when the object falls within the observer's peripersonal space, not when it is located in the extrapersonal space. However, they added a further interesting condition in which another individual (a virtual avatar) was sat close to the object presented in the extrapersonal space (see also Creem-Regehr et al., 2013): in this condition, the affordance effect was restored, showing that objects can afford suitable motor acts to interact with them when they are ready not only for the subject's hand, but also for another agent's hand. In line with this view, recent monkey (Ishida et al., 2010) and human (Brozzoli et al., 2013, 2014) studies showed that neuronal populations do exist in parietal and ventral premotor cortex encoding the spatial position of objects relative to both one's own body and the corresponding body part of an observed subject, suggesting the existence of a shared representation of the space near oneself and others.

CANONICAL AND CANONICAL-MIRROR NEURONS: MOTOR REPRESENTATIONS OF OBJECTS AND ACTIONS IN SPACE

The behavioral evidence so far reviewed suggest that the peripersonal space and social contexts in which an object is seen play a crucial role in affecting the likelihood that it will trigger potential motor representations in the observer's brain. However, the

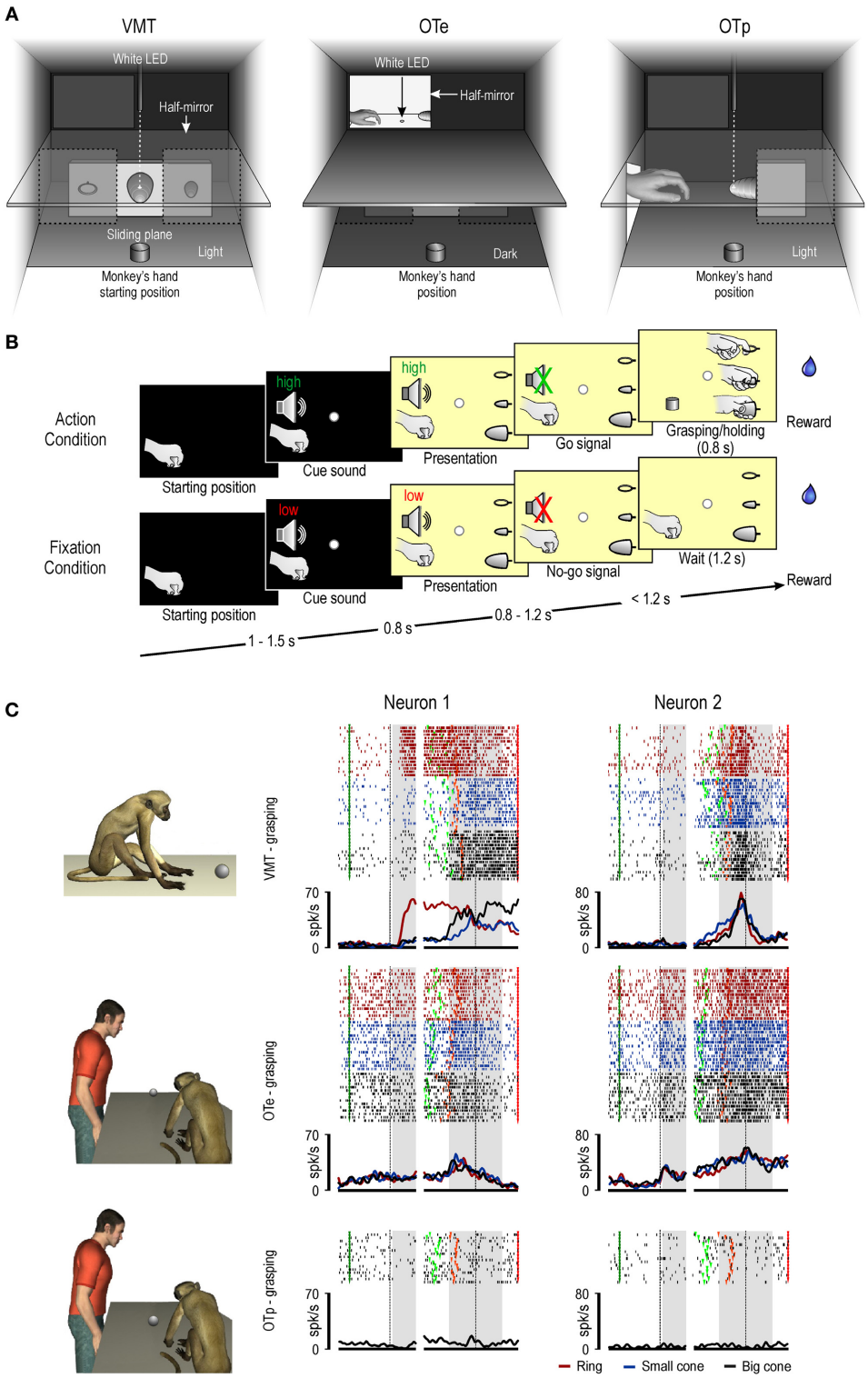


FIGURE 1 | (A) Box and apparatus (seen from the monkey's point of view) settled for carrying out the visuomotor task (VMT), the observation task in the monkey's extrapersonal (OTe) and peripersonal (OTp) space. **(B)** Task phases of Action and Fixation conditions. Each trial started when the monkey had its hand in the starting position. A fixation point was presented and the monkey was required to fixate it for the entire duration of the trial. One of two cue sounds was then presented: a high tone, associated with the action trials, and a low tone, associated with fixation trials. After 0.8 s the lower sector of the box was illuminated and one of the three objects became visible. Then, after a variable time lag (0.8–1.2 s), the sound ceased (go/no-go signal) and the monkey either reached, grasped, and pulled the object (Action condition) or remained still for 1.2 s (Fixation condition) in order to receive the reward. The sequence

(Continued)

FIGURE 1 | Continued
of events and temporal constraints of the OTe and OTp were the same as in the monkey VMT, and the monkey had to simply maintain fixation in order to get the reward. **(C)** Examples of canonical-mirror neurons recorded in all the task contexts. On the left, a schematic view of the experimental paradigm. Each panel shows, from top to bottom, rastergrams and the spike density function. The gap in the rastergrams and histograms is used to indicate that the activity on its left side has been aligned on object presentation (first dashed black vertical line) while that on its right side is aligned on the pulling onset (second dashed black vertical line) of the same trial. The gray shaded areas indicate the time windows used for statistical analysis of neuronal response to object presentation (on the left) and grasping (on the right). Markers: dark green, cue sound onset; light green, cue sound offset (go signal); orange, detachment of the hand from the starting position (reaching onset); red, reward delivery at the end of the trial.

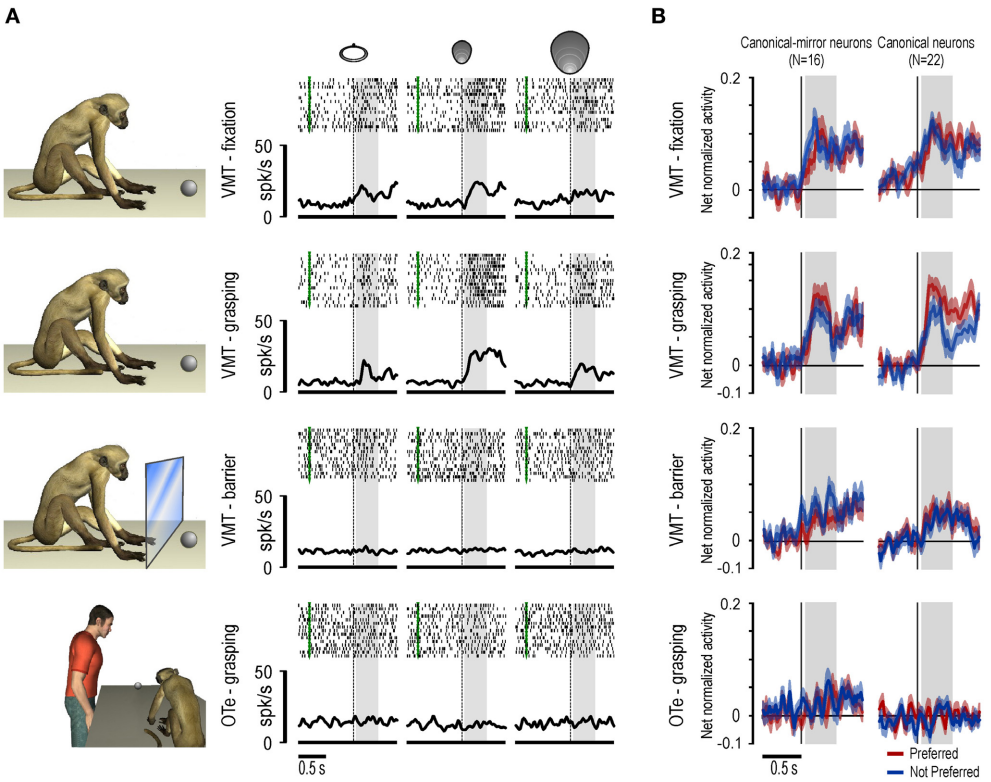


FIGURE 2 | (A) Example of a canonical neuron recorded during an additional control experiment in which the object was presented behind a transparent plastic barrier. Note that the response during object presentation in the VMT was abolished with the interposition of the barrier. Only the alignment and the time window related to object presentation are shown. Other conventions as in **Figure 1**. **(B)** Time course and intensity of the population activity of canonical-mirror and canonical neurons relative to the preferred (red) and not preferred (blue) target object. For each neuron, the preferred/not preferred object are those triggering the stronger/wicker response during grasping execution. The activity is aligned on the light onset during different tasks and conditions.

cortical mechanisms and neural bases underlying these processes need to be further investigated.

Before discussing recent data on these issues, it must be remembered that area F5 contains two main categories of visuomotor neurons, namely, canonical and mirror neurons. The neurons of these two categories show the same response during movement execution, while they differ in the type of visual stimulus triggering them. Canonical neurons, as previously described, respond only when the monkey observes an object, whereas mirror neurons activate only during observation of a motor act performed by another individual. In a recent neurophysiological study (Bonini et al., 2014), we recorded the activity of canonical and mirror neurons from the hand field of macaque ventral premotor cortex while the monkey performed a visuomotor task or observed the same task done by an experimenter, either in the monkey's peripersonal or extrapersonal space (**Figures 1A,B**). One of the main findings of this study was that the previously proposed dichotomy between canonical and mirror neurons appears to be at least too rigid. Indeed, beyond the classical mirror and canonical neurons, grasping neurons have been found showing hallmark features of both categories, that is, they responded both to object presentation and to observation of other's action ("canonical-mirror" neurons—see **Figure 1C**).

A further important result of this study concerns the influence of the space sector in which a target object was presented on the response of these three categories of neurons. Mirror neurons could code others' action both when it was presented in the monkey's peripersonal and extrapersonal space, in line with previous studies (Caggiano et al., 2009). In contrast, object coding by canonical neurons appeared to be markedly constrained to the

peripersonal space, as well as to the visual perspective (subjective view) from which the object was seen by the monkey. This is in line with the classical proposal maintaining that canonical neurons provide a representation of the potential motor act afforded by the observed object, likely participating in the visuomotor transformations of object properties into the appropriate motor act for grasping it (Jeannerod et al., 1995; Fogassi et al., 2001).

Canonical-mirror neurons evidenced different response patterns. Example Neuron 1 (**Figure 1C**) would be classified as a canonical neuron, based on the VMT, but it also responded during the observation of the other's action performed in the extrapersonal space. Example Neuron 2 (**Figure 1C**), in contrast, did not show any response to the presentation of the object during the VMT, while it responded both to objects presented in the monkey's extrapersonal space and the subsequent experimenter's action. This latter finding suggests that the response of part of the canonical-mirror neurons to object presentation should not play a relevant role in visuomotor transformations for grasping. Rather, the object-triggered activation of canonical-mirror neurons may provide a *predictive* representation of the impending action of the observed agent.

In the same study we also showed that space-constrained coding of object, both by canonical and canonical-mirror neurons, relies on a pragmatic rather than metric representation of space. Indeed, most (about 75%) of the recorded canonical and canonical-mirror neurons discharged weakly to object presentation when it occurred behind a transparent plastic barrier, with about half of them showing no significant activation in this condition (see **Figures 2A,B**). This finding clearly demonstrates that neuronal responses to object rely on the actual possibility for the monkey to interact with the observed stimulus. This effect can be explained by the anatomical connections of this sector of area F5 with the adjacent area F4 (Matelli et al., 1986), whose neurons encode monkey's peripersonal space in a pragmatic format (Fogassi et al., 1996).

Space-constrained coding of objects as potential targets for self and others' action appears to rely on different types of neurons located in the same area: some of these neurons, which might enable motor prediction, can play a role for planning actions and for preparing behavioral reactions in the physical and social world.

CONCLUSIONS

Most of the reviewed studies indicate that, besides the purely pictorial description of objects occurring in higher order visual areas, the processing of object features also involves different parallel parieto-frontal circuits constituting the extended motor system (Rizzolatti and Luppino, 2001). In these circuits affordances and contextual elements are crucial for a pragmatic object representation. Among them the peripersonal space appears to play a pivotal role in gating the representation of the potential motor act afforded by the object. When the object is located in the extrapersonal space, its representation as a potential target for the observer's hand action is not activated, while a motor representation of the object appears to be triggered if this latter is a potential target for an observed agent.

ACKNOWLEDGMENT

Supported by the European Commission grant Cogsystems (FP7-250013), Italian PRIN (prot. 2010MEFNF7), and Italian Institute of Technology.

REFERENCES

Ambrosini, E., and Costantini, M. (2013). Handles lost in non-reachable space. *Exp. Brain. Res.* 229, 197–202. doi: 10.1007/s00221-013-3607-0

Anelli, F., Nicoletti, R., Bolzani, R., and Borghi, A. M. (2013). Keep away from danger: dangerous objects in dynamic and static situations. *Front. Hum. Neurosci.* 7:344. doi: 10.3389/fnhum.2013.00344

Arbib, M. A. (1985). Schemas for the temporal organization of behaviour. *Hum. Neurobiol.* 4, 63–72.

Battaglini, P. P., Muzur, A., Galletti, C., Skrap, M., Brovelli, A., and Fattori, P. (2002). Effects of lesions to area V6A in monkeys. *Exp. Brain Res.* 144, 419–422. doi: 10.1007/s00221-002-1099-4

Baumann, M. A., Fluet, M. C., and Scherberger, H. (2009). Context-specific grasp movement representation in the macaque anterior intraparietal area. *J. Neurosci.* 29, 6436–6448. doi: 10.1523/JNEUROSCI.5479-08.2009

Belmalih, A., Borra, E., Contini, M., Gerbella, M., Rozzi, S., and Luppino, G. (2009). Multimodal architectonic subdivision of the rostral part (area F5) of the macaque ventral premotor cortex. *J. Comp. Neurol.* 512, 183–217. doi: 10.1002/cne.21892

Bonini, L., Maranesi, M., Livi, A., Fogassi, L., and Rizzolatti, G. (2014). Space-dependent representation of objects and other's action in monkey ventral premotor grasping neurons. *J. Neurosci.* 34, 4108–4119. doi: 10.1523/JNEUROSCI.4187-13.2014

Borghi, A. M., Flumini, A., Natraj, N., and Wheaton, L. A. (2012). One hand, two objects: emergence of affordance in contexts. *Brain Cogn.* 80, 64–73. doi: 10.1016/j.bandc.2012.04.007

Borra, E., Belmalih, A., Calzavara, R., Gerbella, M., Murata, A., Rozzi, S., et al. (2008). Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb. Cortex* 18, 1094–1111. doi: 10.1093/cercor/bhm146

Borra, E., Belmalih, A., Gerbella, M., Rozzi, S., and Luppino, G. (2010). Projections of the hand field of the macaque ventral premotor area F5 to the brainstem and spinal cord. *J. Comp. Neurol.* 518, 2570–2591. doi: 10.1002/cne.22353

Brozzoli, C., Ehrsson, H. H., and Farne, A. (2014). Multisensory representation of the space near the hand: from perception to action and interindividual interactions. *Neuroscientist* 20, 122–135. doi: 10.1177/1073858413511153

Brozzoli, C., Gentile, G., Bergouignan, L., and Ehrsson, H. H. (2013). A shared representation of the space near oneself and others in the human premotor cortex. *Curr. Biol.* 23, 1764–1768. doi: 10.1016/j.cub.2013.07.004

Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., and Casile, A. (2009). Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science* 324, 403–406. doi: 10.1126/science.1166818

Cardellicchio, P., Sinigaglia, C., and Costantini, M. (2013). Grasping affordances with the other's hand: a TMS study. *Soc. Cogn. Affect. Neurosci.* 8, 455–459. doi: 10.1093/scan/nss017

Costantini, M., Ambrosini, E., Scorolli, C., and Borghi, A. M. (2011a). When objects are close to me: affordances in the peripersonal space. *Psychon. Bull. Rev.* 18, 302–308. doi: 10.3758/s13423-011-0054-4

Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., and Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Exp. Brain Res.* 207, 95–103. doi: 10.1007/s00221-010-2435-8

Costantini, M., Committeri, G., and Sinigaglia, C. (2011b). Ready both to your and to my hands: mapping the action space of others. *PLoS ONE* 6:e17923. doi: 10.1371/journal.pone.0017923

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., and Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: providing a foundation for predicting the future behavior of others. *Front. Hum. Neurosci.* 7:596. doi: 10.3389/fnhum.2013.00596

Culham, J. C., Danckert, S. L., Desouza, J. F., Gati, J. S., Menon, R. S., and Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Exp. Brain Res.* 153, 180–189. doi: 10.1007/s00221-003-1591-5

Davare, M., Rothwell, J. C., and Lemon, R. N. (2010). Causal connectivity between the human anterior intraparietal area and premotor cortex during grasp. *Curr. Biol.* 20, 176–181. doi: 10.1016/j.cub.2009.11.063

Disbrow, E., Litinas, E., Recanzone, G. H., Padberg, J., and Krubitzer, L. (2003). Cortical connections of the second somatosensory area and the parietal ventral area in macaque monkeys. *J. Comp. Neurol.* 462, 382–399. doi: 10.1002/cne.10731

Ellis, R., and Tucker, M. (2000). Micro-affordance: the potentiation of components of action by seen objects. *Br. J. Psychol.* 91(pt 4), 451–471. doi: 10.1348/000712600161934

Fattori, P., Breveglieri, R., Marzocchi, N., Filippini, D., Bosco, A., and Galletti, C. (2009). Hand orientation during reach-to-grasp movements modulates neuronal activity in the medial posterior parietal area V6A. *J. Neurosci.* 29, 1928–1936. doi: 10.1523/JNEUROSCI.4998-08.2009

Fattori, P., Breveglieri, R., Raos, V., Bosco, A., and Galletti, C. (2012). Vision for action in the macaque medial posterior parietal cortex. *J. Neurosci.* 32, 3221–3234. doi: 10.1523/JNEUROSCI.5358-11.2012

Fattori, P., Gamberini, M., Kutz, D. F., and Galletti, C. (2001). 'Arm-reaching' neurons in the parietal area V6A of the macaque monkey. *Eur. J. Neurosci.* 13, 2309–2313. doi: 10.1046/j.0953-816x.2001.01618.x

Fattori, P., Kutz, D. F., Breveglieri, R., Marzocchi, N., and Galletti, C. (2005). Spatial tuning of reaching activity in the medial parieto-occipital cortex (area V6A) of macaque monkey. *Eur. J. Neurosci.* 22, 956–972. doi: 10.1111/j.1460-9568.2005.04288.x

Fattori, P., Raos, V., Breveglieri, R., Bosco, A., Marzocchi, N., and Galletti, C. (2010). The dorsomedial pathway is not just for reaching: grasping neurons in the medial parieto-occipital cortex of the macaque monkey. *J. Neurosci.* 30, 342–349. doi: 10.1523/JNEUROSCI.3800-09.2010

Fitzgerald, P. J., Lane, J. W., Thakur, P. H., and Hsiao, S. S. (2004). Receptive field properties of the macaque second somatosensory cortex: evidence for multiple functional representations. *J. Neurosci.* 24, 11193–11204. doi: 10.1523/JNEUROSCI.3481-04.2004

Fogassi, L., Gallese, V., Buccino, G., Craighero, L., Fadiga, L., and Rizzolatti, G. (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey: a reversible inactivation study. *Brain* 124, 571–586. doi: 10.1093/brain/124.3.571

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., and Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *J. Neurophysiol.* 76, 141–157.

Fogassi, L., Raos, V., Franchi, G., Gallese, V., Luppino, G., and Matelli, M. (1999). Visual responses in the dorsal premotor area F2 of the macaque monkey. *Exp. Brain. Res.* 128, 194–199. doi: 10.1007/s002210050835

Frey, S. H., Vinton, D., Norlund, R., and Grafton, S. T. (2005). Cortical topography of human anterior intraparietal cortex active during visually guided grasping. *Brain Res. Cogn. Brain Res.* 23, 397–405. doi: 10.1016/j.cogbrainres.2004.11.010

Gallese, V., Murata, A., Kaseda, M., Niki, N., and Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport* 5, 1525–1529. doi: 10.1097/00001756-199407000-00029

Galletti, C., Fattori, P., Kutz, D. F., and Gamberini, M. (1999). Brain location and visual topography of cortical area V6A in the macaque monkey. *Eur. J. Neurosci.* 11, 575–582. doi: 10.1046/j.1460-9568.1999.00467.x

Galletti, C., Kutz, D. F., Gamberini, M., Breveglieri, R., and Fattori, P. (2003). Role of the medial parieto-occipital cortex in the control of reaching and grasping movements. *Exp. Brain Res.* 153, 158–170. doi: 10.1007/s00221-003-1589-z

Gallivan, J. P., Cavina-Pratesi, C., and Culham, J. C. (2009). Is that within reach? fMRI reveals that the human superior parieto-occipital cortex encodes objects reachable by the hand. *J. Neurosci.* 29, 4381–4391. doi: 10.1523/JNEUROSCI.0377-09.2009

Gamberini, M., Passarelli, L., Fattori, P., Zucchelli, M., Bakola, S., Luppino, G., et al. (2009). Cortical connections of the visuomotor parietooccipital area V6Ad of the macaque monkey. *J. Comp. Neurol.* 513, 622–642. doi: 10.1002/cne.21980

Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2011). Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. *Brain Struct. Funct.* 216, 43–65. doi: 10.1007/s00429-010-0293-6

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Grefkes, C., Weiss, P. H., Zilles, K., and Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron* 35, 173–184. doi: 10.1016/S0896-6273(02)00741-9

Hadjidimitrakis, K., Bertozzi, F., Breveglieri, R., Fattori, P., and Galletti, C. (2013). Body-centered, mixed, but not hand-centered coding of visual targets in the medial posterior parietal cortex during reaches in 3D space. *Cereb. Cortex* doi: 10.1093/cercor/bht181. [Epub ahead of print].

Hadjidimitrakis, K., Breveglieri, R., Placenti, G., Bosco, A., Sabatini, S. P., and Fattori, P. (2011). Fix your eyes in the space you could reach: neurons in the macaque medial parietal cortex prefer gaze positions in peripersonal space. *PLoS ONE* 6:e23335. doi: 10.1371/journal.pone.0023335

Hurley, S. L. (1998). *Consciousness in Action.* London: Harvard University Press.

Ishida, H., Nakajima, K., Inase, M., and Murata, A. (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *J. Cogn. Neurosci.* 22, 83–96. doi: 10.1162/jocn.2009.21185

Jancke, L., Kleinschmidt, A., Mirzazade, S., Shah, N. J., and Freund, H. J. (2001). The role of the inferior parietal cortex in linking the tactile perception and manual construction of object shapes. *Cereb. Cortex* 11, 114–121. doi: 10.1093/cercor/11.2.114

Janssen, P., Vogels, R., and Orban, G. A. (2000a). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288, 2054–2056. doi: 10.1126/science.288.5473.2054

Janssen, P., Vogels, R., and Orban, G. A. (2000b). Three-dimensional shape coding in inferior temporal cortex. *Neuron* 27, 385–397. doi: 10.1016/S0896-6273(00)00045-3

Jeannerod, M. (1984). The timing of natural prehension movements. *J. Mot. Behav.* 16, 235–254. doi: 10.1080/00222895.1984.10735319

Jeannerod, M., Arbib, M. A., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* 18, 314–320. doi: 10.1016/0166-2236(95)93921-J

Kalenine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., and Buxbaum, L. J. (2013). Visual context modulates potentiation of grasp types during semantic object categorization. *Psychon. Bull. Rev.*. doi: 10.3758/s13423-013-0536-7

Konen, C. S., and Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nat. Neurosci.* 11, 224–231. doi: 10.1038/nn2036

Krubitzer, L., Clarey, J., Tweedale, R., Elston, G., and Calford, M. (1995). A redefinition of somatosensory areas in the lateral sulcus of macaque monkeys. *J. Neurosci.* 15, 3821–3839.

Luppino, G., Murata, A., Govoni, P., and Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Exp. Brain Res.* 128, 181–187. doi: 10.1007/s002210050833

Macfarlane, N. B., and Graziano, M. S. (2009). Diversity of grip in Macaca mulatta. *Exp. Brain Res.* 197, 255–268. doi: 10.1007/s00221-009-1909-z

Matelli, M., Camarda, R., Glickstein, M., and Rizzolatti, G. (1986). Afferent and efferent projections of the inferior area 6 in the macaque monkey. *J. Comp. Neurol.* 251, 281–298. doi: 10.1002/cne.902510302

Matelli, M., Govoni, P., Galletti, C., Kutz, D. F., and Luppino, G. (1998). Superior area 6 afferents from the superior parietal lobule in the macaque monkey. *J. Comp. Neurol.* 402, 327–352.

Milner, A. D., and Goodale, M. A. (1993). Visual pathways to perception and action. *Prog. Brain Res.* 95, 317–337. doi: 10.1016/S0079-6123(08)60379-9

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *J. Neurophysiol.* 78, 2226–2230.

Murata, A., Gallese, V., Kaseda, M., and Sakata, H. (1996). Parietal neurons related to memory-guided hand manipulation. *J. Neurophysiol.* 75, 2180–2186.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *J. Neurophysiol.* 83, 2580–2601.

Noë, A. (2004). *Action in Perception.* Cambrige: MIT Press.

Passarelli, L., Rosa, M. G., Gamberini, M., Bakola, S., Burman, K. J., Fattori, P., et al. (2011). Cortical connections of area V6Av in the macaque: a visual-input node to the eye/hand coordination system. *J. Neurosci.* 31, 1790–1801. doi: 10.1523/JNEUROSCI.4784-10.2011

Poincaré, H. (1908). *Science and Method.* London: Thomas Nelson & sons.

Raos, V., Franchi, G., Gallese, V., and Fogassi, L. (2003). Somatotopic organization of the lateral part of area F2 (dorsal premotor cortex) of the macaque monkey. *J. Neurophysiol.* 89, 1503–1518. doi: 10.1152/jn.00661.2002

Raos, V., Umilta, M. A., Gallese, V., and Fogassi, L. (2004). Functional properties of grasping-related neurons in the dorsal premotor area F2 of the macaque monkey. *J. Neurophysiol.* 92, 1990–2002. doi: 10.1152/jn.00154.2004

Raos, V., Umilta, M. A., Murata, A., Fogassi, L., and Gallese, V. (2006). Functional properties of grasping-related neurons in the ventral premotor area F5 of the macaque monkey. *J. Neurophysiol.* 95, 709–729. doi: 10.1152/jn.00463.2005

Reed, C. L., and Caselli, R. J. (1994). The nature of tactile agnosia: a case study. *Neuropsychologia* 32, 527–539. doi: 10.1016/0028-3932(94)90142-2

Reed, C. L., Shoham, S., and Halgren, E. (2004). Neural substrates of tactile object recognition: an fMRI study. *Hum. Brain Mapp.* 21, 236–246. doi: 10.1002/hbm.10162

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., and Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain Res.* 71, 491–507. doi: 10.1007/BF00248742

Rizzolatti, G., and Fadiga, L. (1998). Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). *Novartis Found Symp.* 218, 81–95; discussion 95–103.

Rizzolatti, G., and Gallese, V. (1997). "From action to meaning," in *Les Neurosciences et la Philosophie de L'action,* ed. J. L. Petit (Paris: Librairie Philosophique J. Vrin), 217–223.

Rizzolatti, G., and Luppino, G. (2001). The cortical motor system. *Neuron* 31, 889–901. doi: 10.1016/S0896-6273(01)00423-8

Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157. doi: 10.1007/s00221-003-1588-0

Sakata, H., and Kusunoki, M. (1992). Organization of space perception: neural representation of three-dimensional space in the posterior parietal cortex. *Curr. Opin. Neurobiol.* 2, 170–174. doi: 10.1016/0959-4388(92)90007-8

Sakata, H., Taira, M., Murata, A., and Mine, S. (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb. Cortex* 5, 429–438. doi: 10.1093/cercor/5.5.429

Sereno, M. E., Trinath, T., Augath, M., and Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. *Neuron* 33, 635–652. doi: 10.1016/S0896-6273(02)00598-6

Silver, M. A., and Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends Cogn. Sci.* 13, 488–495. doi: 10.1016/j.tics.2009.08.005

Srivastava, S., Orban, G. A., De Maziere, P. A., and Janssen, P. (2009). A distinct representation of three-dimensional shape in macaque anterior intraparietal area: fast, metric, and coarse. *J. Neurosci.* 29, 10613–10626. doi: 10.1523/JNEUROSCI.6016-08.2009

Taira, M., Mine, S., Georgopoulos, A. P., Murata, A., and Sakata, H. (1990). Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Exp. Brain Res.* 83, 29–36. doi: 10.1007/BF00232190

Tucker, M., and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Vis. cogn.* 8, 769–800. doi: 10.1080/13506280042000144

Tunik, E., Frey, S. H., and Grafton, S. T. (2005). Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp. *Nat. Neurosci.* 8, 505–511. doi: 10.1038/nn1430

Ungerleider, L. G., and Mishkin, M. (1982). "Two cortical visual systems," in *Analysis of Visual Behavior,* eds D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Cambrige: MIT Press), 549–586.

Valenza, N., Ptak, R., Zimine, I., Badan, M., Lazeyras, F., and Schnider, A. (2001). Dissociated active and passive tactile shape recognition: a case study of pure tactile apraxia. *Brain* 124, 2287–2298. doi: 10.1093/brain/124.11.2287

Van Elk, M., Van Schie, H., and Bekkering, H. (2014). Action semantics: a unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Phys. Life Rev.* 11, 220–250. doi: 10.1016/j.plrev.2013.11.005.

Verhoef, B. E., Vogels, R., and Janssen, P. (2010). Contribution of inferior temporal and posterior parietal activity to three-dimensional shape perception. *Curr. Biol.* 20, 909–913. doi: 10.1016/j.cub.2010.03.058

Xu, Y., and Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends Cogn. Sci.* 13, 167–174. doi: 10.1016/j.tics.2009.01.008

Zachariou, V., Klatzky, R., and Behrmann, M. (2014). Ventral and dorsal visual stream contributions to the perception of object shape and object location. *J. Cogn. Neurosci.* 26, 189–209. doi: 10.1162/jocn_a_00475

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: 10 February 2014; accepted: 14 May 2014; published online: 17 June 2014.
*Citation: Maranesi M, Bonini L and Fogassi L (2014) Cortical processing of object affordances for self and others' action. Front. Psychol. **5**:538. doi: 10.3389/fpsyg.2014.00538*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Maranesi, Bonini and Fogassi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 24 June 2014
doi: 10.3389/fpsyg.2014.00636

Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI

Ricarda I. Schubotz[1,2,3]*, Moritz F. Wurm[4], Marco K. Wittmann[5] and D. Yves von Cramon[2]

[1] *Institute for Psychology, University of Münster, Münster, Germany*
[2] *Max Planck Institute for Neurological Research, Cologne, Germany*
[3] *Department of Neurology, University Hospital Cologne, Köln, Germany*
[4] *Center for Mind/Brain Sciences (CIMeC), University of Trento, Mattarello, Italy*
[5] *Department of Experimental Psychology, University of Oxford, Oxford, UK*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Martin Schürmann, University of Nottingham, UK
Christine E. Watson, Moss Rehabilitation Research Institute, USA

****Correspondence:***
Ricarda I. Schubotz, Institute for Psychology, University of Münster, Fliednerstr. 21, 48149 Münster, Germany
e-mail: rschubotz@uni-muenster.de

Objects are reminiscent of actions often performed with them: knife and apple remind us on peeling the apple or cutting it. Mnemonic representations of object-related actions (action codes) evoked by the sight of an object may constrain and hence facilitate recognition of unrolling actions. The present fMRI study investigated if and how action codes influence brain activation during action observation. The average number of action codes (NAC) of 51 sets of objects was rated by a group of $n = 24$ participants. In an fMRI study, different volunteers were asked to recognize actions performed with the same objects presented in short videos. To disentangle areas reflecting the storage of action codes from those exploiting them, we showed object-compatible and object-incompatible (pantomime) actions. Areas storing action codes were considered to positively co-vary with NAC in both object-compatible and object-incompatible action; due to its role in tool-related tasks, we here hypothesized left anterior inferior parietal cortex (aIPL). In contrast, areas exploiting action codes were expected to show this correlation only in object-compatible but not incompatible action, as only object-compatible actions match one of the active action codes. For this interaction, we hypothesized ventrolateral premotor cortex (PMv) to join aIPL due to its role in biasing competition in IPL. We found left anterior intraparietal sulcus (IPS) and left posterior middle temporal gyrus (pMTG) to co-vary with NAC. In addition to these areas, action codes increased activity in object-compatible action in bilateral PMv, right IPS, and lateral occipital cortex (LO). Findings suggest that during action observation, the brain derives possible actions from perceived objects, and uses this information to shape action recognition. In particular, the number of expectable actions quantifies the activity level at PMv, IPL, and pMTG, but only PMv reflects their biased competition while observed action unfolds.

Keywords: fMRI, object perception, action observation, apraxia, affordance, pantomime

INTRODUCTION

Observed action entails a highly complex stimulus that prompts a multitude of attentional and memory processes. The observer has to be flexible with regard to potential actions that may unroll, but yet quickly discard those which do not pertain to the actual situation. How is this achieved?

When considering object-related action, the observer has access to at least two sources of information that usually help him to quickly recognize the most probable action goal: manipulation movements and objects. These two basic sources of information, rather than being complementary, are intimately interrelated: familiar objects such as mobile phones or knifes are strongly reminiscent of manipulations that we perform with them everyday. Hence, the observer's brain may use these automatically evoked memories of distinct object-related actions (action codes, hereafter) to bias or constrain expectation on upcoming manipulations and hence facilitate recognition of the action, i.e., implemented object function, and thereby the probable actor's goal. For instance, when seeing someone handling a knife and an apple, the object set "knife, apple" evokes two action codes: "cutting apple with knife" and "peeling apple with knife." While tracking the unfolding manipulation, we at a point in time notice that the peeling-action code matches the observed manipulation, and recognize the actor is peeling the apple with the knife (object function), probably to prepare it for eating (goal).

The present fMRI study focused on automatically evoked object-related action codes to find out if, and if so how, they influence the neural basis of action observation. In order to recognize an observed action, it would make no sense to match the observed

action to *all possible* action memories we have. Rather, one could suggest that objects automatically evoke mnemonic codes of the handful of actions we most frequently perform with them (i.e., action codes) (Helbig et al., 2006; Myung et al., 2006; Campanella and Shallice, 2011), information that could greatly constrain the number of expectable actions. Note that we very quickly recognize objects (Bar, 2003), including their pragmatic properties (Liu et al., 2009; Proverbio et al., 2011), while observed manipulation only unfolds and disambiguates in time.

Although the notion of action codes is reminiscent of what is called object *affordance*, they differ in an important respect. According to the classical concept of affordance (Gibson, 1977; see McGrenere and Ho, 2000 for modifications and alternatives), an object affords actions in that seeing the object can automatically prime, and hence facilitate, object-compatible actions in a particular observer (depending also on the observer's body); the object does so by virtue of its physical properties: e.g., size and shape of the object afford appropriate grasping, and its location appropriate pointing (Tucker and Ellis, 1998, 2001, 2004; Craighero et al., 1999; Pavese and Buxbaum, 2002; Phillips and Ward, 2002; Derbyshire et al., 2006; Symes et al., 2007; Cho and Proctor, 2010; Pellicano et al., 2010; Iani et al., 2011; McBride et al., 2012). In contrast, we here were interested in object-evoked representations of actions that do not derive from the object's size, shape or orientation, but from associative memories of how and what for we use these objects in everyday life. Note that our manipulation did not dissociate this "how" and "what for," which can be doubly dissociated in patient groups (Buxbaum and Saffran, 2002).

To show that action codes are effective during action observation, we should find that it makes a difference how many action codes are currently evoked, and whether the observed action matches one of them or not. Accordingly we should find (H1) increased activity in areas that code for currently active action codes; and (H2) increased activity in areas that exploit them for action recognition. In order to disentangle these effects, we presented object-compatible (normal) action and object-incompatible (pantomime) action. As an example for an incompatible action, the actor performed the movements for "cracking an egg" while holding and moving an orange and an orange squeezer. Object-compatible and object-incompatible actions were performed on objects whose NACs, i.e., number of possible action codes related to them, were assessed in a pre-fMRI rating study (see Methods and **Figure 1**).

The NAC effect (H1) should be only driven by the perceived object(s) but be independent of the actually observed action. Thus, we considered areas that positively co-vary with the NAC during object-compatible and object-incompatible action to classify as areas storing action codes. In contrast, currently evoked action codes can only be exploited for action recognition (H2) when observing the former, but not the latter. That means, only if the observed actor executes one of the currently evoked action codes, i.e., in object-compatible actions, can the observer benefit from their automatic pre-activation.

Thus, these action codes put an effective constraint on the to-be-expected possible actions, and identification of the matching action will be enhanced. In terms of neural computation, this results in a continuous, top-down reinforcement of the matching action code, in competition to all currently active action codes, during ongoing action observation (see neuroanatomical hypotheses below). This reinforcement may be achieved by enhancement of the matching action code, or by inhibition of the currently competing but non-matching action codes, or both. Since the present approach could not distinguish between these options, we will refer to this mechanism shortly as "reinforcement" hereafter.

It is particularly essential that, in order to interpret an area's activation as exerting a reinforcement of one particular action code among all currently evoked and hence competing action codes, rather than simply signaling for a successful matching, this activation has to depend on competition strength: to make the particular matching action code to come out on top of three possible actions (NAC 3) is more demanding than on top of two actions or only one (NAC 2 or 1, respectively). Accordingly, regarding (H2), we were not interested in the main effect of object compatibility, but rather in the *interaction* between the NAC and object compatibility: We considered areas that positively co-vary with the NAC during object-compatible *significantly more than* during object-incompatible action to classify as areas exploiting the currently active action codes. These areas should show a significant parametric effect of NAC in object-compatible actions, no significant parametric effect of NAC in object-incompatible actions, and a significant interaction between the NAC and object compatibility.

Regarding the neural correlates of action code storage (H1), we hypothesized that activity in the left anterior inferior parietal lobule (aIPL) increases with the NAC, no matter whether the movie shows an object-compatible or an action-incompatible manipulation. During tool-related tasks, left aIPL is often seen in co-activation with left ventral premotor cortex (PMv) and the posterior middle temporal gyrus (pMTG) (Johnson-Frey, 2004; Culham and Valyear, 2006; Martin, 2007; Creem-Regehr, 2009), i.e., exactly the same network that is reported for action observation (Grèzes and Decety, 2001; Van Overwalle and Baetens, 2009; Caspers et al., 2010), but also for action execution, action imagery, action planning, and action imitation. This network has been referred to as MNS (mirror neuron system) or AON (action observation network), but due to the spectrum of action-related roles of this triad, the label "Action Network" might be more generic. Regarding our hypothesis on areas housing action codes (H1) we focused on aIPL because of converging findings from various studies reporting left aIPL to be engaged in the representation of pragmatic properties of objects, particularly manipulation knowledge (e.g., Chao and Martin, 2000; Kellenbach et al., 2003; Johnson-Frey, 2004; Rumiati et al., 2004; Boronat et al., 2005; Ishibashi et al., 2011). In her thorough review, Creem-Regehr (2009) proposed to conceive of the inferior IPL/IPS as a region for motor cognition, including the generation of internal representations for action and knowledge about actions. Patient studies indicate that the ability to retrieve the correct manipulation for a given tool can be selectively impaired, while in the same patient, the ability to correctly name the tool or point to the tool when named by the experimenter are preserved (Ochipa et al., 1989). This defect in tool utilization has

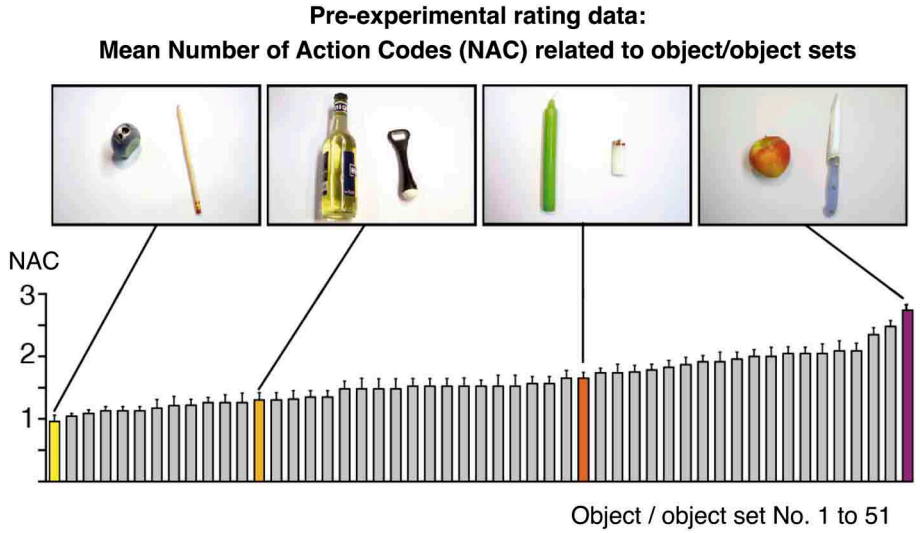


FIGURE 1 | Rating values for the number of action codes (NAC) related to 51 objects/object sets as assessed pre-experimentally in an independent group of $n = 24$ volunteers. Values were subsequently used to model the BOLD fMRI amplitude during object-compatible and object-incompatible action. Some values are highlighted and correspond to the objects shown in the photos, based on which participants delivered their ratings. For instance, the object set "pencil and sharpener" were mostly reported to be related to "sharpening a pencil" (resulting in NAC=1), whereas "candle and lighter" participants often associated "lighting a candle," and sometimes also "melting a candle's foot" (in order to firmly fixate it in a candle stand) (NAC=1.65).

been coined limb apraxia (Rothi and Heilman, 1997). In spite of considerable variance between findings, evidence converges that patients with impaired object use and pantomiming to visually presented objects mostly suffer from lesions that include the left IPL (Rumiati et al., 2004). This region is considered crucial for gestural praxis, tool knowledge, body part knowledge, and manipulation knowledge, together coined as the ability to generate internal models of object-interaction (Buxbaum et al., 2005).

The frontal component of the Action Network, the left PMv, was expected to respond quite differently than aIPL. In relation to our second question, whether there would be areas reflecting the selection among competing action codes, we hypothesized (H2) the left PMv to be enhanced by the number of action codes, but in contrast to aIPL only for object-compatible, not object-incompatible action videos. That is, we should see left PMv only for the interaction between NAC and object compatibility of the observed action.

This hypothesis was motivated by the notion that premotor regions serve the top-down selection among alternative manipulation options provided in parietal areas (Fagg and Arbib, 1998; Rushworth et al., 2003). The lateral premotor cortex is made of a variety of functionally highly specialized sub-areas which in turn are connected in multiple parallel, largely segregated loops to a mosaic of sub-areas making of the parietal cortex (Luppino and Rizzolatti, 2000). Among these premotor-parietal loops, the ventral premotor—inferior parietal loop was reported to code for grasping and manipulation (Rizzolatti et al., 1987), but also for the sight of graspable objects (via so-called canonical neurons in PMv, Murata et al., 1997; Rizzolatti and Fadiga, 1998). As for fronto-parietal loops in general, the functional role of PMv with regard to IPL is providing inhibitory and reinforcing input to focus and elevate currently relevant codes in IPL to modulate adaptive perception, attention and behavior.

Addressing the interplay between lateral premotor and parietal areas in object-directed action, Fagg and Arbib (1998) put forward that anterior parietal cortex provides ventral premotor cortex with a multiple description of how the object can be grasped and used. In PMv, then, all corresponding motor acts are first activated, and then the currently required one is selected (or reinforced, to keep with the more process-dynamic notion adopted above). For instance, neurophysiological studies in macaques indicate that potential plans for movements to multiple targets are simultaneously represented in parietal and frontal areas (Andersen and Cui, 2009) and, as information accumulates, eliminated in a competition for overt execution (Cisek and Kalaska, 2005) (for an application of the notion of selection as frontoparietal reinforcement signal in humans, see e.g., Ramsey et al., 2013). Fagg and Arbib (1998) proposed that in action execution, this selection needs prefrontal input (via pre-supplementary motor area) that signals the current goals of the individual. However recent imaging findings indicate that action selection that emerges from the race between competitive decision-units is reflected in premotor, not prefrontal, areas (Rowe et al., 2010), suggesting that action selection in premotor sites does not necessarily need prefrontal bias.

In the present experimental approach, competition between the action codes evoked by the perceived object was to be resolved by the actually observed manipulation. We expected that in case of a successful match (which was only possible for object-compatible actions), the PMv would reinforce the matching action codes in aIPL, just as it does during action execution. Load on this reinforcement would be a function of action codes only in object-compatible action, as outlined above, manifesting in an

interaction of the NAC and object compatibility of the observed action. Of course, if PMv does exert an action code dependent reinforcing signal on aIPL during action observation, this effect should be reflected in both of these areas.

Finally, in object-incompatible action, reinforcement load should be generally higher than in object-compatible action, as action recognition is unrestricted by the currently evoked action codes: there are objects that evoke action codes, but none of them matches the observed action. That does also mean, reinforcement load should *not* depend on the number of action codes in case of object-incompatible actions. Since objects employed in object-incompatible actions evoke action codes that are not effective to constrain the matching process, the number of possible actions is the number of all possible actions that humans do perform with objects. Accordingly, our third hypothesis (H3) was that object-incompatible actions lead to an overall higher response than object-compatible actions in left PMv and IPL (replicating Schubotz and von Cramon, 2009), but show no positive co-variance with the number of currently activated action codes.

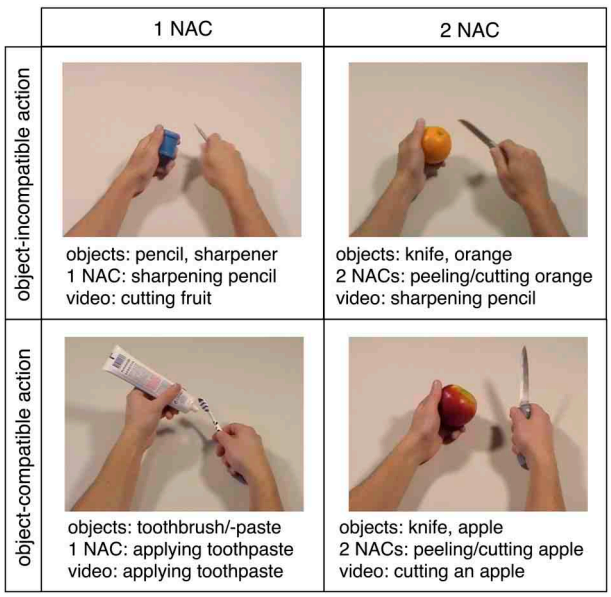


FIGURE 2 | Snapshots of videos showing actions with objects mostly related to either one (left panel) or two (right panel) actions. Actions could be exploited to constrain action recognition only in object-compatible actions (lower panel; examples show "applying toothpaste" on the left and "cutting an apple" on the right), but not in object-incompatible actions (upper panel; examples show "cutting a fruit" on the left and "sharpening a pencil" on the right). The corresponding videos can be found in the supplementary material.

MATERIALS AND METHODS

PARTICIPANTS

Seventeen right-handed, healthy volunteers (13 women, 20–31 years, mean age 25.6 years) participated in the study. After being informed about potential risks and screened by a physician of the institution, participants gave informed consent before participating. The local ethics committee of the University of Cologne approved the experimental standards. Data were handled anonymously.

STIMULI AND TASKS

Subjects were presented with two kinds of trials, videos showing actions (snapshots in **Figure 2**; for examples of videos, see supplementary material) and short verbal action descriptions (without video) referring to these actions (e.g., "cutting bread," "peeling an apple," "cleaning a cell phone"). Each trial lasted 6 s and started with a movie (2 s) followed by a fixation phase (**Figure 3**). The length of the fixation phase (2.5–4 s) depended on the variable jitter times (0, 500, 1000, or 1500 ms) that were inserted before the movie to enhance the temporal resolution of the BOLD response. Actions were either performed on appropriate objects (object-compatible actions, e.g., peeling an apple with a knife) or on inappropriate objects (object-incompatible actions, e.g., making the same movements with a pencil and a sharpener).

Subjects were instructed to attend to the presented movies. They were informed that some of the movies were followed by a trial providing a verbal action description that either matched or did not match the content of the preceding movie. Subjects then performed a verification task, i.e., they were asked to indicate by button press whether the verbal description was consistent with the action movie previously presented or not. It was emphasized that it did not play any role whether actions, to which the action description referred, were object-compatible or not. Thus, when subjects saw the action "peeling an apple" performed with an apple and a knife, or with a pencil and a sharpener, and the subsequent trial delivered the verbal action description "peeling an apple," the correct answer was "yes." In the case of an action description trial, participants immediately delivered their responses on a two-button response box using their index finger for affirmative responses (description pertains to the movie in the preceding trial) and their middle finger for rejections (description did not pertain to the movie in the preceding trial). Half of the action descriptions were to be affirmed and half to be rejected.

Action movies varied with regard to the number of action codes (**Figure 1**; see also "Pre-experimental assessment on objects' average number of action codes"). Importantly, in case that two or three objects were involved in an action, they always made up object sets that were indicative of possible actions (which were, of course, not actually performed in the case of object-incompatible actions); e.g., participants were presented the object-incompatible action "cracking an egg" performed in an as-if manner on the objects "orange" and "orange squeezer," i.e., a pair of objects that could be used to prepare orange juice (**Figure 2**). Thus, videos showing object-incompatible actions never involved meaningless object sets such as e.g., an orange and a sharpener.

Twenty percent of the movies (i.e., 21 of 105 object-compatible actions and 21 of 105 object-incompatible actions) were followed by an action description that had the length of a regular trial (2 s description, including response phase, plus 4 s fixation phase), resulting in 42 additional trials. Finally, 20 empty trials

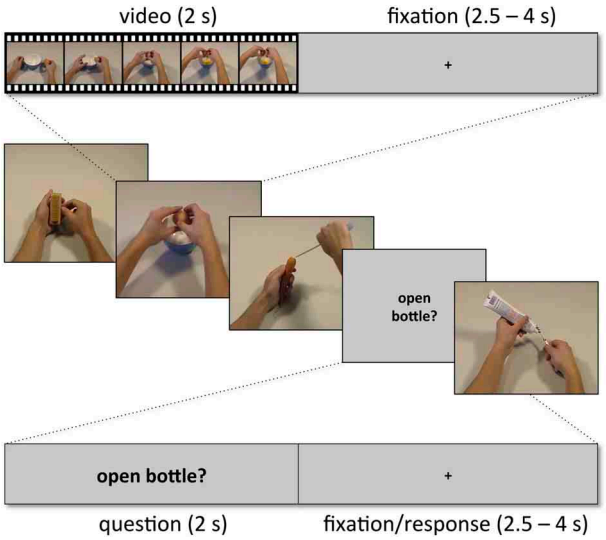


FIGURE 3 | Example sequence of five trials with within-trial time course for one video trial (top) and for one question trial (bottom).

(resting state) of 6 s duration were presented intermixed with the experimental trials. Thus, 272 trials were presented altogether.

For each subject, each action was presented four times during the course of the experimental session, two times object-compatible and two times object-incompatible, with different objects each time. Importantly, we balanced the order of appearance of object-compatible and incompatible actions in the time course of the appearance. Hence, all combinations (1: compatible, compatible, incompatible, incompatible; 2: compatible, incompatible, compatible, incompatible; 3: incompatible, compatible, incompatible, compatible; 4: incompatible, incompatible, compatible, compatible) occurred equally often in the experiment.

PRE-EXPERIMENTAL ASSESSMENT ON AVERAGE NUMBER OF ACTION CODES

In order to determine the NACs of the objects later used in the action movies, we assessed the spontaneous assignment of actions to these objects in a group of $n = 24$ volunteers. To avoid mnemonic confounds, this group was not identical to the group tested in the fMRI session, i.e., none of the participants of the pre-experimental assessment was included in the fMRI study. Participants were given photographs of 51 objects (e.g., cell phone) or object sets (e.g., apple and knife). There were 27% single objects, 63% two-object sets and 10% 3-object sets. Participants were asked to write down all potential actions that the presented objects were typically reminiscent of in their eyes. For instance, participants rated an apple and a knife to be most suggestive of "cutting an apple into halves," "peeling an apple," and "coring an intact apple" (3 actions), whereas an orange and an orange squeezer were rated suggestive of "squeezing an orange" (1 action). To assess NACs rather than object affordances, participants were explicitly asked to provide object-specific goal-directed actions, not object grasping or transport.

We did not impose a temporal restriction onto this rating process and participants had time to thoroughly ponder on potential actions. The collection of actions typically took less than 1 min per object or object set; moreover, no participant came up with invalid or odd actions. On the basis of this rating, the average NAC score was calculated for each object or set of objects (**Figure 1**). NAC scores, ranging from 0.95 to 2.73, were subsequently used in the parametric analysis of fMRI data (see below). Importantly, there was no systematic relation between NAC score and number of objects in a set. Thus, single objects yielded a mean NAC of 1.63 ± 0.33, two-object-sets 1.66 ± 0.4, and three-object sets 1.18 ± 0.1. To statistically rule out the potential confound that NAC co-vary with the number of objects displayed in an action, we calculated a correlation of NAC with the number of objects. There was no correlation [$r_{(49)} = -0.223$, $p = 0.12$].

MRI DATA ACQUISITION

Twenty-two axial slices (192 mm field of view; 64×64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3×3 mm) parallel to bi-commissural line (AC–PC) covering the whole brain were acquired using a single-shot gradient EPI sequence (2 s repetition time; 30 ms echo time; 90° flip angle; 116 kHz acquisition bandwidth) sensitive to BOLD contrast. Prior to the functional imaging, 26 anatomical T1-weighted MDEFT images (Ugurbil et al., 1993; Norris, 2000) with the same spatial orientation as the functional data were acquired. In a separate session, high-resolution whole-brain images (160 slices of 1 mm thickness) were acquired from each participant to improve the localization of activation foci using a T1-weighted 3-D-segmented MDEFT sequence covering the whole brain.

fMRI DATA ANALYSIS

After offline motion-correction using the Siemens motion protocol PACE (Siemens, Erlangen, Germany), fMRI data were processed using the software package LIPSIA (Lohmann et al., 2001). To correct for the temporal offset between the slices acquired in one image, a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter with a cutoff frequency of 1/90 Hz. Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM ($SD = 0.8$ voxel). To align the functional data slices with a 3-D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed.

The rotational and translational parameters were acquired on the basis of the MDEFT slices to achieve an optimal match between these slices and the individual 3-D reference dataset. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to the same standard size. The resulting parameters were then used to transform the functional slices employing a trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system. Resulting data had a spatial resolution of $3 \times 3 \times 3$ mm (27 mm3).

The statistical evaluation was based on a least-squares estimation using the general linear model for serially auto-correlated observations (Friston et al., 1995; Worsley and Friston, 1995). The design matrix was generated with a delta function, convolved with the hemodynamic response function (gamma function) (Glover, 1999). The design matrix comprised the following events: object-compatible action videos, object-incompatible action videos, object-compatible action videos with an amplitude modeled by the corresponding objects' NAC, object-incompatible action videos with an amplitude modeled by the corresponding objects' NAC, question trials, and empty trials (null events).

Brain activations were analyzed time-locked to onset of the videos. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM to account for the temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images, that is, beta value estimates of the raw-score differences between specified conditions were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts.

One-sample *t*-tests were employed for the group analyses across the contrast images of all participants that indicated whether observed differences between conditions were significantly distinct from zero. The *t*-values were subsequently transformed into *z*-scores. To correct for false-positive results, an initial *z*-threshold was set to 2.33 ($p < 0.01$, one-tailed). In a second step, the results were corrected for multiple comparisons at the cluster level, using cluster size and cluster value thresholds that were obtained by Monte-Carlo simulations at a significance level of $p = 0.05$, i.e., the reported activations were significantly activated at $p \leq 0.05$, corrected for multiple comparison at cluster level.

RESULTS

BEHAVIORAL RESULTS

Performance was assessed by error rates and reaction times. We calculated paired-*t*-tests (one-tailed, in expectation of lower performance in object-incompatible actions) for each of these measures between question trials addressing object-incompatible and object-compatible actions. While error rates became not significant ($t_{16} = -0.66$, $p = 0.259$), reaction times showed a small effect ($t_{16} = 1.75$, $p = 0.049$). Thus, object-compatible and object-incompatible actions were responded to equally correct (mean ± SE: object-compatible action 5.8 ± 1.2% errors and object-incompatible 4.8 ± 1.4% errors), but recognition of object-incompatible actions took 40 ms longer (object-compatible action 1192 ± 61 ms and object-incompatible 1232 ± 69 ms).

Moreover, we calculated bivariate correlations between NACs and reaction times or error rates, respectively. As a result, there was neither an effect on error rates [$r_{(16)} = -0.012$, $p = 0.96$], nor on reaction times [$r_{(16)} = -0.026$, $p = 0.92$]. Together, behavioral statistics suggested that recognition times were slightly but significantly reduced by object information, but both object-incompatible and object-compatible actions could be successfully identified.

As a caveat, we employed a retrospective judgment in order to control for the participants' performance in action recognition. This task was implemented by extra question trials that followed an action observation trial in order to avoid response-related confounds: motor execution, and, even worse, trial-specific interactions between executed button press, implied and observed manipulations. That is, reaction times and error rates refer to a response delivered one trial after action observation. So, our paradigm was optimized for fMRI rather than for specific behavioral effects.

fMRI RESULTS

Our two main hypotheses H1 and H2 addressed the parametric effect of number of action codes. In left aIPL, this effect was expected to be independent of the object-compatibility of the observed action (Hypothesis H1), but depend on object-compatibility in PMv (Hypothesis H2). H1 was tested by calculating a conjunction of the thresholded parametric contrast in object-compatible actions and the thresholded parametric contrast in object-incompatible actions. H2 was tested by an interaction contrast, i.e., by contrasting the parametric effect of number of action codes in object-compatible actions with the parametric effect of number of action codes in object-incompatible actions. While H2 addressed compatible > incompatible actions, we also report the reverse contrast for exploratory reasons. We follow the view that in order to show that differences between the two parametric effects (number or action codes in object-compatible actions, number of actions in object-incompatible actions) are statistically significant, it is not enough to show each of them, but to calculate a contrast between both, i.e., an interaction (Nieuwenhuis et al., 2011). In order to make their respective contributions to these two effects descriptively transparent, however, we also report the parametric effect in object-compatible actions and the parametric effect in object-incompatible actions separately. Finally, we report findings on the main effect of object compatibility of observed action (Hypothesis H3).

Parametric effect of number of action codes (NAC) common to object-compatible and object-incompatible actions (Hypothesis H1)

The conjunction between the thresholded parametric effect of action codes in *object-compatible* actions and the thresholded parametric effect of action codes in *object-incompatible* actions revealed activation in the left aIPS and in the left posterior middle temporal gyrus (pMTG). Both areas showed increasing activation with the number of possible actions associated with the objects shown in the movies (**Figure 4A**, **Table 1**).

Interaction effect of NAC and object-compatibility (Hypothesis H2)

Parametric effect of NAC in object-incompatible actions. For object-incompatible action, activity increased with number of action codes in the left aIPS, the left pMTG, encroaching into adjacent pSTS and TPJ, and in the left Cuneus (**Figure 4D**, **Table 2**).

Parametric effect of NAC in object-compatible actions. For object-compatible action, activity increased with number of

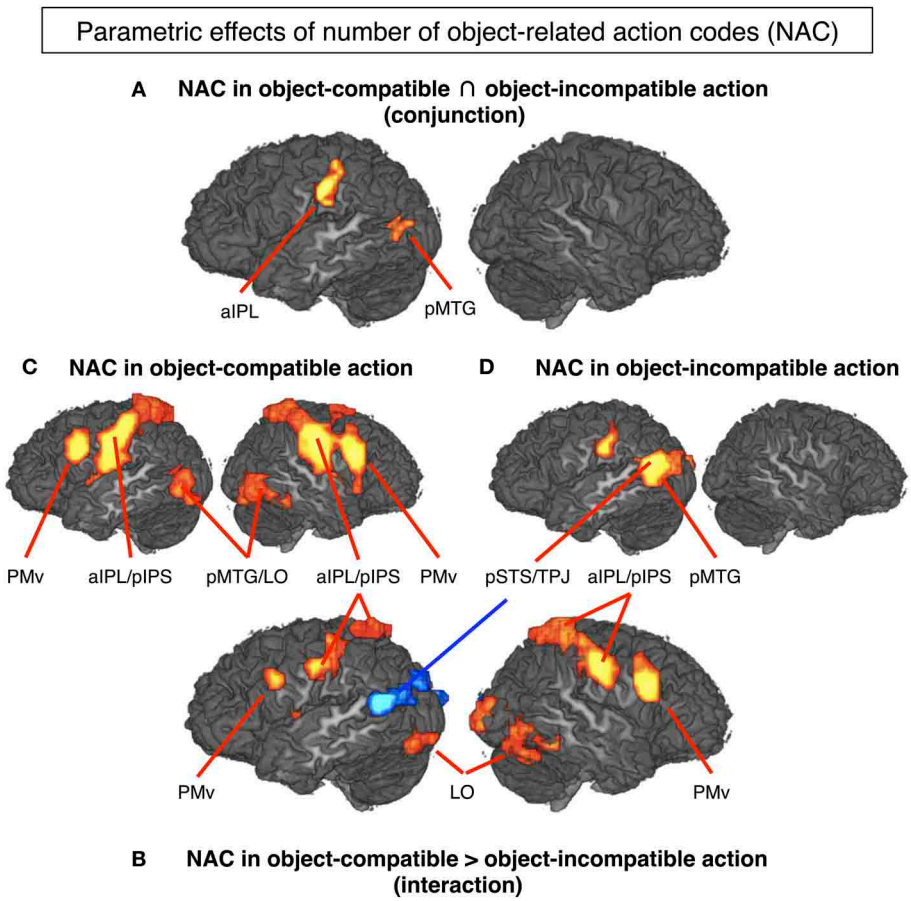


FIGURE 4 | The mere sight of objects triggered the representation of a number of possible object-related actions that quantified the activation in several cortical areas ($P < 0.05$, corrected for multiple comparisons). **(A)** Left aIPL and pMTG can be considered to code for the number of object-related action codes (NAC), as their activity increased with the NAC regardless of whether the observed action was object-compatible (and hence matched one of the evoked NACs) or object-incompatible (cf. Hypothesis H1). **(B)** In contrast, PMv, right IPS, left pIPS, bilateral LO and mid-insula increased with NAC only in object-compatible action, presumably reflecting a top-down competition bias between the observed and the remaining object-related but unobserved actions (cf. Hypothesis H2). Interestingly, left pSTS/TPJ increased with NAC in object-incompatible actions blue spot in **(B)** suggesting that in case of a non-match, lower constraints on expectable actions (i.e., higher NACs) increased efforts to read out the actor's hand postures and movements. For descriptive purposes, **(C)** and **(D)** show the NAC effect separately for videos on object-compatible and object-incompatible (pantomime-with-incompatible-objects) actions. **Table 1** lists Talairach coordinates for *z*-maps shown in **(A)** and **(B)**, **Table 2** for **(C)** and **(D)**.

action codes in the Action Network (PM, aIPL, pMTG) as well as the fusiform gyrus/lateral occipital cortex and mid-insula. All activation spots were found in both hemispheres (**Figure 4C**, **Table 2**).

Interaction effect of NAC in object-compatible > incompatible actions. For object-compatible action, additional activation was found to increase with the NAC in right ventral and dorsal premotor cortex (PMv, PMd) as well as in bilateral mid-insula. Moreover, activation in IPS was recorded bilaterally and extended from anterior into its horizontal segments. Finally, activation in the pMTG extended inferiorly and posteriorly into the lateral occipital cortex (LO) and emerged particularly in the right hemisphere (**Figure 4B**, **Table 1**).

Interaction effect of NAC in object-incompatible > object-compatible actions. Some areas responded to increasing number of evoked action codes exclusively during object-incompatible actions. These were located in the left pSTS, extending posteriorly and dorsally into the temporo-parietal junction (TPJ), and in left cuneus (**Figure 4B**, **Table 1**).

Main effects of observation of object-compatible and object-incompatible action (Hypothesis H3)

The present study employed object-compatible and object-incompatible action to investigate the effects of action codes and their impact on action observation. However, it is important to consider that all effects so far reported supervened on the typical network found for action observation (cf. Introduction), including the lateral premotor-parietal loops as well as temporo-occipital areas related to attention to motion, movements and objects (**Figure 5A**). Recorded activation patterns were almost identical for object-compatible and object-incompatible actions when compared to rest, but direct contrasts revealed significant

Table 1 | Anatomical area (for abbreviations, see main text), Talairach coordinates (*x, y, z*) and maximal *Z*-score (max) of activated clusters ($p = 0.05$, corrected for multiple comparisons) for parametric effects of the number of automatically evoked object-related action codes (NAC) that were common to both for object-compatible and object-incompatible actions (conjunction; Hypothesis H1) or interacted with object compatibility of observed action (Hypothesis H2) (cf. Figures 4A,B).

Area	*x*	*y*	*z*	Max
H1: NAC IN BOTH OBJECT COMPATIBLE ∩ INCOMPATIBLE ACTION				
aIPS	−62	−24	36	3.08
pMTG	−47	−66	9	2.83
H2: INTERACTION NAC *x* OBJECT COMPATIBLE > INCOMPATIBLE ACTIONS				
PMv	−56	9	27	3.08
	52	9	24	4.43
aIPS	−50	−27	36	4.18
	40	−27	42	5.22
pIPS	−35	−54	63	3.66
	31	−45	66	4.61
Mid-Insula	−38	−6	21	3.24
	34	−3	34	3.42
LO / pMTG	37	−69	−3	4.07
Cuneus	22	−90	6	5.10
pSTS / TPJ	−59	−57	15	−3.65
Cuneus	−8	−81	0	−4.81

Table 2 | Anatomical area (for abbreviations, see main text), Talairach coordinates (*x, y, z*) and maximal Z-score (*max*) of activated clusters ($p = 0.05$, corrected for multiple comparisons) for parametric effects of the number of automatically evoked object-related action codes (NAC), separately analyzed for object-compatible actions and for object-incompatible actions (cf. Figures 4C,D).

Area	*x*	*y*	*z*	Max
NAC IN OBJECT-COMPATIBLE ACTIONS				
aIPS/SMG	−47	−24	33	5.38
	49	−21	33	5.67
pMTG, LO	−47	−63	3	3.51
	43	−63	0	4.79
Fusiform gyrus / LO	−47	−75	−6	3.06
Fusiform gyrus	46	−45	−12	3.18
pIPS	−32	−51	60	5.74
	28	−48	60	5.39
	37	−30	39	5.91
PMv	−56	6	27	4.23
	52	9	27	5.29
PMd	22	−12	51	3.64
Mid-Insula	−38	3	12	4.55
	31	−6	18	4.77
NAC IN OBJECT-INCOMPATIBLE ACTIONS				
aIPS	−62	−24	33	3.15
pMTG/pSTS/TPJ	−53	−69	21	4.13
Cuneus	−14	−78	−3	4.59

differences in the modulations of this network as well, particularly enhanced activity in left PMv and IPL for object-incompatible actions (**Figure 5B**). For object-compatible actions, we found enhanced activity solely in fusiform areas (**Figure 5C**). These findings fully replicate those of a previous study (Schubotz and von Cramon, 2009; **Figure 3**).

In contrast to the findings in Schubotz and von Cramon (2009), object-incompatible actions in the current study additionally activated mesial Brodmann Area (BA) 8, the ventral tegmental area and the bilateral dorsal anterior insula (not shown in **Figure 5**). These activations probably reflect dopaminergic enhancement during decision uncertainty (Volz et al., 2005). We consider this difference to be due to the use of object sets that always implied valid action options for object-compatible as well as for object-incompatible actions, in contrast to our previous study. Thus, uncertainty was somewhat higher for object-incompatible, as objects did not indicate whether the to-be-expected action would be an object-incompatible or an object-compatible action, and action codes could not become effective to constrain the process of identifying the observed action.

DISCUSSION

Objects are reminiscent of actions that we typically perform with them. These object-related actions (action codes) may influence action observation by providing a constraint on the number of expectable actions, and hence facilitate action recognition. We used fMRI in an action observation paradigm to test whether left aIPL codes for action codes, i.e., whether its activation level varies as a function of the currently evoked number of action codes (main effect action codes; Hypothesis H1). Moreover, we employed object-compatible and object-incompatible action videos to test whether left PMv reflects the exploitation of evoked action codes. Here we reasoned that an area that exploits action codes in action observation should positively co-vary with the NAC in case of object-compatible, but not object-incompatible action, since action codes can act as a constraint only in the former (interaction effect action codes x object compatibility; Hypothesis H2).

In expectation to replicate findings from a previous study (Schubotz and von Cramon, 2009), we hypothesized that object-compatible and object-incompatible action differ in highly similar way from the resting level, but when directly contrasted with one another show enhanced activity for object-incompatible actions in the entire Action Network, including left PMv and IPL (Hypothesis H3).

RESPONSES TO AUTOMATICALLY EVOKED CODES OF OBJECT-RELATED ACTIONS

Object-compatible and object-incompatible actions differed with respect to the usability of object information, but objects implied possible actions in both. To tap this object-based action-preactivation, we computed the parametric effect of the number of action codes separately for object-compatible actions and object-incompatible actions, and subsequently built the conjunction of both. As a result, activity was recorded in only two areas, the left aIPL and the left pMTG (**Figure 4A**). Finding aIPL confirmed our

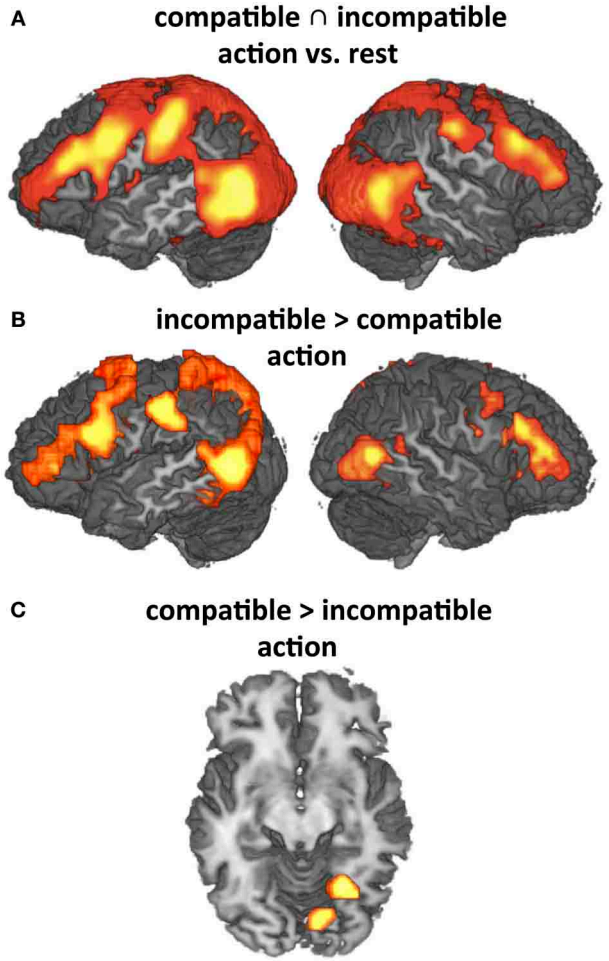


FIGURE 5 | Significant activation differences addressed by Hypothesis (H3) ($P < 0.05$, corrected for multiple comparisons) that replicate previous findings (Schubotz and von Cramon, 2009). (A) Main effects of observation of object-compatible and object-incompatible action as compared to rest (conjunction). **(B)** and **(C)**: Differential effects as revealed by direct contrasts between observation of object-compatible and object-incompatible action. Both object-compatible and object-incompatible actions induce strong activation in the Action Network, including premotor-parietal and temporo-occipital areas. This activation pattern is even intensified by object-incompatible action **(B)** where participants have to rely on manipulation information to recognize the action. Correspondingly, activity is significantly enhanced in fusiform gyrus when object information is valid **(C)**.

hypothesis, which was based on the role of inferior parietal lobe in the appraisal of pragmatic implications provided by objects. Left pMTG was not hypothesized and will be discussed as a *post-hoc* finding.

The left IPL activation was restricted to the anterior bank of the intraparietal sulcus (aIPS) and did not encroach into supramarginal gyrus (SMG). This is an important observation, since these two areas have distinct functions, as implicated by research in their putative homologues in the macaque, AIP and PF, respectively (Committeri et al., 2007; McGeoch et al., 2007). The latter mediates between PMv and pSTS in a network coined "mirror neuron system" or MNS for both action observation and action execution (as lucidly outlined in Keysers and Perrett, 2004), whereas the former provides PMv with a pragmatic description of objects (Fagg and Arbib, 1998). The core difference here is that neurons in AIP already respond to objects even when not manipulated, whereas PF neurons are particularly tuned to the sight of the experimenter grasping and manipulating objects (Gallese et al., 2002).

This difference seems particularly relevant in the context of the present findings, as the conjunction contrast aimed to tap only the parametric effects of object-evoked action knowledge, *independent of the object-compatibility of the observed manipulation*. It makes perfect sense that the parametric action codes contrast did not identify SMG (as putative human PF-homolog), because activation that was caused by observation was accounted for by the main effect action vs. rest (**Figure 5**), i.e., it was canceled out in the parametric action codes contrast. Notably, exploitation of action codes was reflected by extension of activation into SMG that we found only for object-compatible actions, as will be discussed later. Thus, our findings perfectly corroborate the assumption of a functional dissociation or relative weighting of AIP/aIPS reflecting object-related action information and PF/SMG reflecting the observation of object manipulation.

Human and macaque data converge with regard to the manipulation-related role of anterior intraparietal cortex. The role of macaque AIP in providing pragmatic object descriptions has been related to "hand manipulation neurons" (Gardner et al., 2007a) in this region and to the encoding of context-specific hand grasping movements to perceived objects (Gallese et al., 1994; Murata et al., 1996; Baumann et al., 2009). Human left aIPL is selectively activated during the explicit retrieval of specific ways of grasping tools (Chao and Martin, 2000) and manipulating objects (Kellenbach et al., 2003). Using an interaction design implementing two cue types (naming and pantomiming) and two response triggers (objects and actions), Rumiati et al. (2004) showed that the left aIPL is particularly active for the transforming objects into skilled object manipulation. A recent fMRI study showed that activity in human aIPS reflects the relationship between object features and grasp type, as in macaques (Begliomini et al., 2007). Also paralleling macaque data, aIPS is particularly enhanced when object information is to be transferred between the visual and the tactile modality (Grefkes et al., 2002). Our results crucially extend these findings, showing that activity in aIPS increases with the mere implication of more possible actions, i.e., the more visual properties of the objects are mentally transferred to different, merely imagined tactile properties.

The present study did not distinguish between semantic/conceptual ("what") and procedural/motor ("how to") representations triggered by the sight of objects, and its perfectly possible that both are automatically evoked. However, there is some evidence that aIPL is more related to the "how to" knowledge related to objects. For instance, Boronat et al. (2005) asked participants to determine whether two given objects are manipulated similarly (e.g., a piano and a laptop keyboard) or serve the same function (e.g., a box of matches and a lighter). Only the left IPL was more engaged during judgments on manipulation than

during judgments on object function (cf. Kellenbach et al., 2003 for parallel findings).

Patient studies support this interpretation, showing that damage to the left IPL can result in an inability to recognize and produce precise hand postures associated with familiar objects while functional knowledge of objects seems spared (Buxbaum and Saffran, 2002; Buxbaum et al., 2003). Binkofski and Buxbaum (2013) proposed that two action systems have to be distinguished in the dorsal stream: a bilateral dorso-dorsal "grasp" stream linking superior parietal to dorsal premotor sites for reaching and grasping objects based on their size, shape or orientation; and a ventro-dorsal "use" system linking inferior parietal to ventral premotor sites for skilled functional object use.

In the present study, objects varied with regard to the number of implied actions, and thereby ways to use the objects, but of course, also in the way to grasp them. Although our parametric approach—object-evoked action options—tapped a very subtle source of variance in our stimulus material (videos), this approach did not allow distinguishing between automatically evoked representations of object-related ways of manipulating, and object-related ways of grasping (i.e., affordances). However, participants were required to recognize the observed actions, and hence could not solely rely on the observed kind of grasping; rather, they had to exactly analyze the way of subsequent usage to determine the observed action with confidence. Moreover, finding PMv and aIPL to increase with the number of active action codes points to the ventro-dorsal "use" system rather than to the dorso-dorsal "grasp" stream.

Left pMTG showed up in the action vs. rest contrast, as expected, as left pMTG is mostly seen in action observation, and also for tool perception (cf. Introduction). However, just as left aIPL, pMTG was also found to positively co-vary with the number of object-implied actions (**Figure 4A**). Fusiform gyrus, pMTG, and aIPL are considered sensitive to the three types of information required for identification of tools: their visual form, the typical motion with which they move when we use them, and the way they are manipulated, respectively (Beauchamp et al., 2002; Mahon et al., 2007). Following this view, we suggest pMTG and aIPL both co-varied in activation with the number of active action codes, because action codes differed not only with regard to the way we use objects, but also in the way the object moves while used. For instance, when participants saw the actor handling a knife and an apple, the automatically evoked action codes included two sorts of knife manipulation, but also the corresponding two sorts of knife motion. Of course, the visual form of the knife was invariant, fitting to the fact that action codes showed no effect in fusiform gyrus. Note that other authors have put forward that pMTG rather than being a motion-coding area, represents conceptual object knowledge (Johnson-Frey, 2004; Fairhall and Caramazza, 2013). There might be also subtle regional differences in functions, as the posterior temporal region contains a variety of functionally specialized areas.

In fact, pMTG refers to an only vague macroanatomical definition of a cortical region that lies in direct vicinity of functional related areas. The peak coordinates of the left MTG in our study were at Talairach $x = -47, y = -67, z = 9$, which is nearly identical to peak coordinates of the extrastriate body area (EBA; $x = \pm 47.2, y = -66.7, z = 4.7$) when averaging across 13 recent fMRI studies (Downing et al., 2006a,b; Taylor et al., 2007; Myers and Sowden, 2008). Moreover, human motion selective area hMT (Greenlee, 2000; Peuskens et al., 2005) overlaps with EBA (e.g., Downing et al., 2007; Taylor et al., 2007). Although the parametric increase of activity in EBA or hMT in the current experimental design cannot be due to demands on body part and motion perception, it could reflect the range of movements and body postures associated with a given object. On the one hand, EBA's contribution in the processing of body posture (Downing et al., 2006b) could be required here as referring to typical hand postures and configurations indicative of the manipulations applicable to an object. On the other hand, hMT is engaged in the processing of complex motion patterns (Peuskens et al., 2005), but also in motion as merely implied or announced by hand postures or objects (Kourtzi and Kanwisher, 2000). Interestingly, both hMT and EBA, together with pSTS sensitive to the perception of biological motion, were found to adapt to the repetition of observed actions even when novel exemplars of object manipulation were shown, suggesting a role of these areas in the representation of the type of manipulation rather than its particular instantiation (Kable and Chatterjee, 2006). Our findings fit to this notion, as the parametric effect of action codes in pMTG and aIPL was independent of the actual observation of one of these actions (revealed by the conjunction of both).

EXPLOITING OBJECT-RELATED KNOWLEDGE TO RECOGNIZE ACTIONS

Only object-compatible, not object-incompatible actions matched one of the action codes supposedly evoked by the sight of the involved objects. Thus, the NAC quantified the constraint imposed onto recognizing object-compatible, not object-incompatible actions: the actually observed action was one out of about one, two or three expectable actions. As hypothesized, PMv activity positively co-varied with the NAC in object-compatible but not incompatible actions. We found this activation in both hemispheres, together with corresponding activation foci in anterior IPL, bilateral lateral occipital cortex extending into pMTG in the right hemisphere, and bilateral mid-insula (**Figure 4B**). Activity in pIPS was bilateral, but pronounced in the right hemisphere spanning from a ventral postcentral region and anterior SMG up to the horizontal segment of the IPS. The fact that left aIPL, the area that we had found for the parametric effect of the NAC, surfaced in this interaction contrast as well indicate that it was dominant, though not specific, for object-compatible actions.

As outlined in the Introduction, we take this network to reflect the fronto-parietal reinforcement of object-implied action options while tracking the unfolding action. Here, the observed action matched one of actions the observer was expecting due to the observed object or set of objects. In this case, and only then, PMv reinforced the matching action manipulation in IPL, and the matching tool motion in pMTG. Further extrastriate visual activation located in the right cuneus may point to modulations going even further downstream.

Importantly, this interaction contrast tapped only into areas whose activation increased with the competition load between object-evoked action options. This effect was observed not only

at the frontal component of the Action Network, but also at the corresponding parietal and posterior temporal sites. Thus, reinforcement increases activation also at the targets in the posterior brain. It is well-known that frontal and parietal/temporal areas interact for selective purposes in attention, with the latter providing "bottom-up" externally driven, perceptual input on which the frontal areas exert a "top-down" selective modulation for goal-directed cognition and behavior (Frith, 2001; Bar, 2003; Pessoa et al., 2003). For sure, the relevant parietal and temporal areas themselves provide highly integrated information of the stimulus. That is, they build rather a "mid" level between frontal and lower visual areas, exerting "top-down" biasing signals on the latter as well (Kastner and Ungerleider, 2001).

Premotor-parietal-temporal activation patterns during action observation have been suggested to reflect a re-activation of actions stored in memory (Decety and Ingvar, 1990; Jeannerod, 1999). Our findings specify this formula by showing that during action observation, the premotor, parietal, and temporal components of this network differ with regard to their sensitivity of object-implied actions: Unlike IPL and pMTG, PMv was only sensitive for the exploitation of competing implied actions, but not for the mere number of implied actions. While IPL and pMTG reflected the action options both as evoked by the sight of the objects (bottom-up) and as competition resolved by frontal biasing signals (top-down), PMv was indifferent with regard to the former: it showed for the number of action codes effect for the interaction between, but not for the conjunction of, object-compatible and object-incompatible actions.

In order to understand and interpret this finding, it helps to consider three converging results in the present study: PMv was present in both object-compatible and object-incompatible action (conjunction contrast, **Figure 5**), more pronounced in object-incompatible as in object-compatible actions (masked direct contrast, **Figure 5**) (H3), and driven by object-evoked action options in object-compatible but not in object-incompatible actions (interaction of action codes and object compatibility, **Figure 4B**). This data pattern suggests that PMv not only registers object-evoked action representations, as aIPL and pMTG do, but also dynamically applies these internal action representations, either in order to adapt to, or to predict, the ongoing action (cf. Schubotz, 2007).

As to the parietal activation revealed in the interaction contrast, it was found to extend from anterior to posterior IPS in the right hemisphere. Why was pIPS activity so pronounced for the right hemisphere? Mruczek et al. (2013) recently reported that tools evoke stronger responses than non-tools in an anterior intraparietal region. Authors suggest that posterior IPS encode features common to any graspable object (including tools), whereas anterior IPS integrate this grip-relevant information with "experience-dependent knowledge of action associations, affordances, and goals, which are uniquely linked to tools" (Mruczek et al., 2013, p. 2892). Coordinates of activation maxima in posterior IPS were most closely located to those related to macaque area MIP (Grefkes and Fink, 2005). MIP is suggested to be involved in coordination of hand movements and visual targets (Eskandar and Assad, 2002), particularly in transforming the spatial coordinates of a target into a representation that is exploited by the motor system for computing the appropriate movement vector (Cohen and Andersen, 2002). Interestingly, these computations take place even in advance of the motion execution itself (Johnson et al., 1996) and hence point to a role of MIP in the detection of movement errors and their correction already on the basis of internal models (Kalaska et al., 2003). More recent studies specify this region as providing tactile information to circuits linking anterior intraparietal to ventral premotor regions, giving on-line feedback needed for goal-directed hand movements (Gardner et al., 2007a,b).

This computational profile was perfectly reflected in the increase of activity in this region reported here, when pragmatic object-implied constraints on expectable manipulations could be integrated with the currently unfolding action. Also the notion of detection of movement errors and their correction on the basis of internal models (Kalaska et al., 2003) fits very well to the present finding, as our parametric contrast pinpointed the competition load between action options. Thus, when multiple action options were implied by the perceived object or object set, and hence represented as multiple internal models of potentially observable manipulations, pIPS may contribute to the detection of discrepancies between expected and observed manipulations.

To finally address activation in the mid-insula, this region relays tactile information from the somatosensory cortex to the frontal cortex (Burton and Sinclair, 2000). Activity was located at the posterior short insula gyrus, which is delimited by the precentral and central insular sulci. This dysgranular region has connections to SI and SII (cf. Guenot et al., 2004 for review). Together with SII and SMG, the mid-insula is suggested to play a crucial role in tactile object recognition, and to integrate somatosensory information to provide a coherent image of an object appropriate for cognitive action (Reed et al., 2004; Milner et al., 2007). Since we found this region to positively co-vary in activity with the number of object-implied expectable manipulations, but only when the observed action matched one of them, we speculate that the enhancement of the matching action comprised also a tactile representation of the observed object manipulation.

OBSERVING OBJECT-INCOMPATIBLE ACTIONS: OBJECTS EVOKE ACTION OPTIONS THAT DO NOT FOSTER ACTION RECOGNITION

In the present study, videos showing object-incompatible actions (i.e., pantomimes with incompatible objects) were employed as a control condition that served to tell apart effects that could be only due to the sight of objects (common to object-compatible and incompatible actions) from effects that could be only due their manipulation (different between object-compatible and incompatible actions). Objects and object sets were always reminiscent of valid action options, both in object-compatible and object-incompatible actions. Moreover, object-incompatible and object-compatible actions were presented randomly intermixed and each occurred with equal probability of 0.5. Together, these design features provoked, as intended, an initial analysis of object information and an attempt to match the observed actions on one of the automatically evoked action code.

To be sure, object-incompatible action is certainly more than just some kind of "incomplete" action, and there are positive effects of object-incompatible action, i.e., activations that come in addition to what we see during observation of object-compatible action. We have investigated these effects elsewhere

(Schubotz and von Cramon, 2009). Replicating these findings, object-compatible and object-incompatible action observation yielded highly similar activation patterns in the resting contrast, with significant differences only in the emphasis of different parts of the Action Network (cf. conjunction in **Figure 5**). Thus, differences in our parametric analyses in object-compatible and object-incompatible action could not be due to principally absent activations in the object-incompatible condition, i.e., "positive effects" in the object-compatible action condition cannot be just due to "negative effects" in object-incompatible actions.

It is important to note that humans are perfectly able to decode actions from object-incompatible manipulations, even from early childhood on (Fein, 1981). Also in the present study, participants performed as well in object-incompatible trials (4.8% errors, 1232 ± 69 ms reaction time) as in object-compatible action trials (5.8% errors, 1192 ± 61 ms reaction time). Saying that participants failed to exploit object-evoked action representations in the case of object-incompatible actions thus does not mean that they failed in the task, but rather, that their strategy had to be adapted according to the stimulus. Object-incompatibility was revealed by a mismatch between the currently active action codes and the actually observed manipulation. As object information was invalid here, observers had to entirely rely on the analysis of hand movements when trying to decode the currently pursued action. In accordance with this suggestion, and replicating a previous study (Schubotz and von Cramon, 2009), we found that the entire Action Network engaged in manipulation recognition (PMv, aIPS, pMTG) enhanced for object-incompatible as compared to object-compatible actions (cf. Hypothesis H3).

However, we also reasoned that action codes should not result in any significant effect on object-incompatible actions since they cannot help to constrain the recognition process. However, the novel and striking finding here was that activity in an area comprising left pSTS and TPJ did increase with the NAC. More specifically, this activation was located in the horizontal posterior segment of the pSTS, extending toward the ascending posterior segment, and hence comprised a temporal-parietal-occipital junction (BA 37, 39 and 19) (**Figure 4D** and blue spot in **Figure 4B**). Activation was left-lateralized, corresponding to the processing of information from the right visual field, and, in the present study, the dominant (right) hand of the actor. This fits well with the experimental setting as, to an observer, motion and posture of the dominant hand is more informative than that of the non-dominant hand; the latter typically holds and stabilizes the object while the former performs the relevant manipulations. Focus on the right visual field was also indicated by increased activation in left cuneus (cf. Machner et al., 2009).

These effects of an increasing NAC were only found for object-incompatible actions and indicated that, although object information was in fact not usable here, the expectations of particular hand movements announced by the objects still affected further stimulus analysis. Note that the more action options were evoked by the object, the lower were the constraints on the to-be-expected manipulations. At the same time, the probability increased that the actually observed manipulation eventually matches one of the pre-activated actions: When you expect one out of three potential actions to occur, a fourth and unexpected action is more difficult to detect than in a case when you expect exactly one specific action to occur.

With this in mind, we take left pSTS activation to reflect the intensified focus on the hand's movements in an attempt to decode the displayed action. Interestingly, activation extended posteriorly and dorsally into TPJ. Due to its particular functional profile in attentional orienting as well as in mentalizing paradigms (Decety and Lamm, 2007; Mitchell, 2008), TPJ has been discussed to have a "where-to" functionality in analogy with the spatial "where" functionality of the dorsal stream (Van Overwalle, 2009). That is, it responds to externally generated behaviors with the aim of identifying the possible end-state of these behaviors (cf. discussion in Van Overwalle and Baetens, 2009). Note that the "end-state" of behavior can be read as being actually related to the physical body, as TPJ is related to the sensation of the position and the movement of one's own body (Blanke et al., 2004). Our findings corroborate this interpretation as they showed TPJ activation to proportionally increase with the number of expectable end-states of the unfolding action. Importantly, TPJ activation was not specific to or indicative of object-incompatible (pantomime) perception in general, as the TPJ effect was only found for the action code parameter in object-incompatible actions, whereas it was absent in the direct contrast object-incompatible vs. object-compatible action.

CONCLUDING REMARKS

The faculty of understanding what other persons are doing is based, among other factors, on the analysis of object and manipulation information. The present study shows that the action-observing brain accurately extrapolates the expectable actions from the objects that the actor is handling, and, when detecting a match between these expectable actions and the actually observed one, subsequently reinforces the matching action against the competition of the remaining but unobserved actions. These findings impressively reflect that object-evoked actions constrain the recognition process in action observation

ACKNOWLEDGMENTS

We cordially thank Matthis Drolet and Christiane Ahlheim for helpful comments on the manuscript, and Florian Riegg for their experimental assistance. Marco Wittmann's contribution was supported by the Wellcome Trust.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00636/abstract

REFERENCES

Andersen, R. A., and Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* 63, 568–583. doi: 10.1016/j.neuron.2009.08.028

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Baumann, M. A., Fluet, M. C., and Scherberger, H. (2009). Context-specific grasp movement representation in the macaque anterior intraparietal area. *J. Neurosci.* 29, 6436–6448. doi: 10.1523/JNEUROSCI.5479-08.2009

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159. doi: 10.1016/S0896-6273(02)00642-6

Begliomini, C., Wall, M. B., Smith, A. T., and Castiello, U. (2007). Differential cortical activity for precision and whole-hand visually guided grasping in humans. *Eur. J. Neurosci.* 25, 1245–1252. doi: 10.1111/j.1460-9568.2007.05365.x

Binkofski, F., and Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain Lang.* 127, 222–229. doi: 10.1016/j.bandl.2012.07.007

Blanke, O., Landis, T., Spinelli, L., and Seeck, M. (2004). Out-of-body experience and autoscopy of neurological origin. *Brain* 127, 243–258. doi: 10.1093/brain/awh040

Boronat, C. B., Buxbaum, L. J., Coslett, H. B., Tang, K., Saffran, E. M., Kimberg, D. Y., et al. (2005). Distinctions between manipulation and function knowledge of objects: evidence from functional magnetic resonance imaging. *Cogn. Brain Res.* 23, 361–373. doi: 10.1016/j.cogbrainres.2004.11.001

Burton, H., and Sinclair, R. J. (2000). Attending to and remembering tactile stimuli: a review of brain imaging data and single-neuron responses. *J. Clin. Neurophysiol.* 17, 575–591. doi: 10.1097/00004691-200011000-00004

Buxbaum, L. J., and Saffran, E. M. (2002). Knowledge of object manipulation and object function: dissociations in apraxic and nonapraxic subjects. *Brain Lang.* 82, 179–199. doi: 10.1016/S0093-934X(02)00014-7

Buxbaum, L. J., Sirigu, A., Schwartz, M. F., and Klatzky, R. (2003). Cognitive representations of hand posture in ideomotor apraxia. *Neuropsychologia* 41, 1091–1113. doi: 10.1016/S0028-3932(02)00314-7

Buxbaum, L. J., Johnson-Frey, S. H., and Bartlett-Williams, M. (2005). Deficient internal models for planning hand–object interactions in apraxia. *Neuropsychologia* 43, 917–929. doi: 10.1016/j.neuropsychologia.2004.09.006

Campanella, F., and Shallice, T. (2011). Manipulability and object recognition: is manipulability a semantic feature? *Exp. Brain Res.* 208, 369–383. doi: 10.1007/s00221-010-2489-7

Caspers, S., Zilles, K., Laird, A. R., and Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50, 1148–1167. doi: 10.1016/j.neuroimage.2009.12.112

Chao, L., and Martin, A. (2000). Representation of manipulable manmade objects in the dorsal stream. *Neuroimage* 12, 478–484. doi: 10.1006/nimg.2000.0635

Cho, D., and Proctor, R. W. (2010). The object−based Simon effect: Grasping affordance or relative location of the graspable part? *J. Exp. Psychol.* 36, 853–861. doi: 10.1037/a0019328

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. doi: 10.1016/j.neuron.2005.01.027

Cohen, Y. E., and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nat. Rev. Neurosci.* 3, 553–562. doi: 10.1038/nrn873

Committeri, G., Pitzalis, S., Galati, G., Patria, F., Pelle, G., Sabatini, U., et al. (2007). Neural bases of personal and extrapersonal neglect in humans. *Brain* 130, 431–441. doi: 10.1093/brain/awl265

Craighero, L., Fadiga, L., Rizzolatti, G., and Umiltà, C. J. (1999). Action for perception: a motor-visual attentional effect. *Exp. Psychol. Hum. Percept. Perform* 25, 1673–1692. doi: 10.1037/0096-1523.25.6.1673

Creem-Regehr, S. H. (2009). Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. *Neurobiol. Learn. Mem.* 91, 166–171. doi: 10.1016/j.nlm.2008.10.004

Culham, J. C., and Valyear, K. F. (2006). Human parietal cortex in action. *Curr. Opin. Neurobiol.* 16, 205–212. doi: 10.1016/j.conb.2006.03.005

Decety, J., and Ingvar, D. H. (1990). Brain structures participating in mental simulation of motor behavior: a neuropsychological interpretation. *Acta Psychol.* 73, 13–24. doi: 10.1016/0001-6918(90)90056-L

Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* 13, 580–593. doi: 10.1177/1073858407304654

Derbyshire, N., Ellis, R., and Tucker, M. (2006). The potentiation of two components of the reach-to-grasp action during object categorisation in visual memory. *Acta Psychol.* 122, 74–98. doi: 10.1016/j.actpsy.2005.10.004

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., and Kanwisher, N. (2006a). Domain specificity in visual cortex. *Cereb. Cortex* 16, 1453–1461. doi: 10.1093/cercor/bhj086

Downing, P. E., Peelen, M. V., Wiggett, A. J., and Tew, B. D. (2006b). The role of the extrastriate body area in action perception. *Soc. Neurosci.* 1, 52–62. doi: 10.1080/17470910600668854

Downing, P. E., Wiggett, A., and Peelen, M. (2007). fMRI investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *J. Neurosci.* 27, 226–233. doi: 10.1523/JNEUROSCI.3619-06.2007

Eskandar, E. N., and Assad, J. A. (2002). Distinct nature of directional signals among parietal cortical areas during visual guidance. *J. Neurophysiol.* 88, 1777–1790.

Fagg, A., and Arbib, M. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Netw.* 11, 1277–1303. doi: 10.1016/S0893-6080(98)00047-1

Fairhall, S. L., and Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *J. Neurosci.* 33, 10552–10558. doi: 10.1523/JNEUROSCI.0051-13.2013

Fein, G. G. (1981). Pretend play in childhood: an integrative review. *Child Develop.* 52, 1095–1118. doi: 10.2307/1129497

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Frith, C. (2001). A framework for studying the neural basis of attention. *Neuropsychologia* 39, 1367–1371. doi: 10.1016/S0028-3932(01)00124-5

Gallese, V., Murata, A., Kaseda, M., Niki, N., and Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport* 5, 1525–1529. doi: 10.1097/00001756-199407000-00029

Gallese, V., Fogassi, L., Fadiga, L., and Rizzolatti, G. (2002). "Action representation and the inferior parietal lobule," in *Common Mechanisms in Perception and Action: Attention and Performance, Vol. XIX,* eds. W. Prinz and B. Hommel (Oxford: Oxford Univ Press), 334–355.

Gardner, E. P., Babu, K. S., Ghosh, S., Sherwood, A., and Chen, J. (2007b). Neurophysiology of prehension. III. Representation of object features in posterior parietal cortex of the macaque monkey. *J. Neurophysiol.* 98, 3708–3730. doi: 10.1152/jn.00609.2007

Gardner, E. P., Babu, K. S., Reitzen, S. D., Ghosh, S., Brown, A. S., Chen, J., et al. (2007a). Neurophysiology of prehension. I. Posterior parietal cortex and object-oriented hand behaviors. *J. Neurophysiol.* 97, 387–406. doi: 10.1152/jn.00558.2006

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.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage* 9, 416–429. doi: 10.1006/nimg.1998.0419

Greenlee, M. W. (2000). Human cortical areas underlying the perception of optic flow: brain imaging studies. *Int. Rev. Neurobiol.* 44, 269–292. doi: 10.1016/S0074-7742(08)60746-1

Grefkes, C., and Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *J. Anat.* 207, 3–17. doi: 10.1111/j.1469-7580.2005.00426.x

Grefkes, C., Weiss, P. H., Zilles, K., and Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron* 35, 173–184. doi: 10.1016/S0896-6273(02)00741-9

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.

Guenot, M., Isnard, J., and Sindou, M. (2004). Surgical anatomy of the insula. *Adv. Tech. Stand. Neurosurg.* 29, 265–288. doi: 10.1007/978-3-7091-0558-0_7

Helbig, H. B., Graf, M., and Kiefer, M. (2006). The role of action representations in visual object recognition. *Exp. Brain Res.* 174, 221–228. doi: 10.1007/s00221-006-0443-5

Iani, C., Baroni, G., Pellicano, A., and Nicoletti, R. (2011). On the relationship between affordance and Simon effects: are the effects really independent? *Eur. J. Cogn. Psychol.* 23, 121–131. doi: 10.1080/20445911.2011.467251

Ishibashi, R., Lambon Ralph, M. A., Saito, S., and Pobric, G. (2011). Different roles of lateral anterior temporal lobe and inferior parietal lobule in coding function and manipulation tool knowledge: evidence from an rTMS study. *Neuropsychologia* 49, 1128–1135. doi: 10.1016/j.neuropsychologia.2011.01.004

Jeannerod, M. (1999). To act or not to act: perspectives on the representation of actions. *Q. J. Exp. Psychol.* 52, 1–29. doi: 10.1080/713755803

Johnson, P. B., Ferraina, S., Bianchi, L., and Caminiti, R. (1996). Cortical networks for visual reaching: physiological and anatomical organization of frontal and parietal lobe arm regions. *Cereb. Cortex* 6, 102–119. doi: 10.1093/cercor/6.2.102

Johnson-Frey, S. H. (2004). The neural basis of complex tool use in humans. *Trends Cogn. Sci.* 8, 71–78. doi: 10.1016/j.tics.2003.12.002

Kable, J. W., and Chatterjee, A. (2006). Specificity of action representations in the lateral occipitotemporal cortex. *J. Cogn. Neurosci.* 18, 1498–1517. doi: 10.1162/jocn.2006.18.9.1498

Kalaska, J. F., Cisek, P., and Gosselin-Kessiby, N. (2003). Mechanisms of selection and guidance of reaching movements in the parietal lobe. *Adv. Neurol.* 93, 97–119.

Kastner, S., and Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39, 1263–1276. doi: 10.1016/S0028-3932(01)00116-6

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, 20–46. doi: 10.1162/089892903321107800

Keysers, C., and Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends Cogn Sci.* 8, 501–507. doi: 10.1016/j.tics.2004.09.005

Kourtzi, Z., and Kanwisher, N. (2000). Implied motion activates extrastriate motion-processing areas: response to David and Senior (2000). *Trends Cogn Sci.* 4, 295–296. doi: 10.1016/S1364-6613(00)01512-6

Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281–290. doi: 10.1016/j.neuron.2009.02.025

Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al. (2001). LIPSIA - a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comp. Med. Imag. Graph.* 25, 449–457. doi: 10.1016/S0895-6111(01)00008-8

Luppino, G., and Rizzolatti, G. (2000). The organization of the frontal motor cortex. *News Physiol. Sci.* 15, 219–224.

Machner, B., Sprenger, A., Sander, T., Heide, W., Kimmig, H., Helmchen, C., et al. (2009). Visual search disorders in acute and chronic homonymous hemianopia: lesion effects and adaptive strategies. *Ann. N. Y. Acad. Sci.* 1164, 419–426. doi: 10.1111/j.1749-6632.2009.03769.x

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., and Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron* 55, 507–520. doi: 10.1016/j.neuron.2007.07.011

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143

McBride, J., Sumner, P., and Husain, M. (2012). Conflict in object affordance revealed by grip force. *Q. J. Exp. Psychol.* 65, 13–24. doi: 10.1080/17470218.2011.588336

McGeoch, P. D., Brang, D., and Ramachandran, V. S. (2007). Apraxia, metaphor and mirror neurons. *Med. Hyp.* 69, 1165–1168. doi: 10.1016/j.mehy.2007.05.017

McGrenere, J., and Ho, W. (2000). "Affordances: clarifying and evolving a concept," in *Proceedings of the Graphics Interface 2000 Conference* (Montreal, QC), 179–186.

Milner, T. E., Franklin, D. W., Imamizu, H., and Kawato, M. (2007). Central control of grasp: manipulation of objects with complex and simple dynamics. *Neuroimage* 36, 388–395. doi: 10.1016/j.neuroimage.2007.01.057

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* 18, 262–271. doi: 10.1093/cercor/bhm051

Mruczek, R. E., von Loga, I. S., and Kastner, S. (2013). The representation of tool and non-tool object information in the human intraparietal sulcus. *J. Neurophysiol.* 109, 2883–2896. doi: 10.1152/jn.00658.2012

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V. and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *J. Neurophysiol.* 78, 2226–2230.

Murata, A., Gallese, V., Kaseda, M., and Sakata, H. (1996). Parietal neurons related to memory-guided hand manipulation. *J. Neurophysiol.* 75, 2180–2186.

Myers, A., and Sowden, P. T. (2008). Your hand or mine? The extrastriate body area. *Neuroimage* 42, 1669–1677. doi: 10.1016/j.neuroimage.2008.05.045

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. doi: 10.1016/j.cognition.2004.11.010

Nieuwenhuis, S., Forstmann, B. U., and Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat. Neurosci.* 14, 1105–1107. doi: 10.1038/nn.2886

Norris, D. G. (2000). Reduced power multiscale MDEFT imaging. *J. Magn. Res. Imaging* 11, 445–451.

Ochipa, C., Rothi, L. J. G., and Heilman, K. M. (1989). Ideational apraxia: a deficit in tool selection and use. *Ann. Neurol.* 25, 190–193. doi: 10.1002/ana.410250214

Pavese, A., and Buxbaum, L. J. (2002). Action matters: the role of action plans and object affordances in selection for action. *Vis. Cogn.* 9, 559–590. doi: 10.1080/13506280143000584

Pellicano, A., Iani, C., Borghi, A. M., Rubichi, S., and Nicoletti, R. (2010). Simon-like and functional affordance effects with tools: the effects of object perceptual discrimination and object action state. *Q. J. Exp. Psychol.* 63, 2190–2201 doi: 10.1080/17470218.2010.486903

Pessoa, L., Kastner, S., and Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *J. Neurosci.* 23, 3990–3998.

Peuskens, H., Vanrie, J., Verfaillie, K., and Orban, G. A. (2005). Specificity of regions processing biological motion. *Eur. J. Neurosci.* 21, 2864–2875. doi: 10.1111/j.1460-9568.2005.04106.x

Phillips, J. C., and Ward, R. (2002). S–R correspondence effects of irrelevant visual affordance: time course and specificity of response activation. *Vis. cogn.* 9, 540–558. doi: 10.1080/13506280143000575

Proverbio, A. M., Adorni, R., and D'Aniello, G. E. (2011). 250 ms to code for action affordance during observation of manipulable objects. *Neuropsychologia* 2711–2717. doi: 10.1016/j.neuropsychologia.2011.05.019

Ramsey, R., Hansen, P., Apperly, I., and Samson, D. (2013). Seeing it my way or your way: frontoparietal brain areas sustain viewpoint-independent perspective selection processes. *J. Cogn. Neurosci.* 25, 670–684. doi: 10.1162/jocn_a_00345

Reed, C. L., Shoham, S., and Halgren, E. (2004). Neural substrates of tactile object recognition: an fMRI study. *Hum. Brain Mapp.* 21, 236–246. doi: 10.1002/hbm.10162

Rizzolatti, G., and Fadiga, L. (1998). Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). *Novartis Found. Symp.* 218, 81–103.

Rizzolatti, G., Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., and Ponzoni-Maggi, S. (1987). Neurons related to goal-directed motor acts in inferior area 6 of the macaque monkey. *Exp. Brain Res.* 67, 220–224. doi: 10.1007/BF00269468

Rothi, L. J., and Heilman, K. M. (1997). *Apraxia: The Neuropsychology of Action*. Hove: Psychology Press.

Rowe, J. B., Hughes, L., and Nimmo-Smith, I. (2010). Action selection: a race model for selected and non-selected actions distinguishes the contribution of premotor and prefrontal areas. *Neuroimage* 51, 888–896. doi: 10.1016/j.neuroimage.2010.02.045

Rumiati, R. I., Weiss, P. H., Shallice, T., Ottoboni, G., Noth, J., Zilles, K., et al. (2004). Neural basis of pantomiming the use of visually presented objects. *Neuroimage* 21, 1224–1231. doi: 10.1016/j.neuroimage.2003.11.017

Rushworth, M. F., Johansen-Berg, H., Göbel, S. M., and Devlin, J. T. (2003). The left parietal and premotor cortices: motor attention and selection. *Neuroimage* 1, 89–100. doi: 10.1016/j.neuroimage.2003.09.011

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* 11, 211–218. doi: 10.1016/j.tics.2007.02.006

Schubotz, R. I., and von Cramon, D. Y. (2009). The case of pretense: observing actions and inferring goals. *J. Cogn. Neurosci.* 21, 642–653. doi: 10.1162/jocn.2009.21049

Symes, E., Ellis, R., and Tucker, M. (2007). Visual object affordances: object orientation. *Acta Psychol.* 124, 238–255. doi: 10.1016/j.actpsy.2006.03.005

Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme.

Taylor, J. C., Wiggett, A. J., and Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J. Neurophysiol.* 98, 1626–1633. doi: 10.1152/jn.00012.2007

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. doi: 10.1037/0096-1523.24.3.830

Tucker, M., and Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Vis. cogn.* 8, 769–800. doi: 10.1080/13506280042000144

Tucker, M., and Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychol.* 116, 185–203. doi: 10.1016/j.actpsy.2004.01.004

Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., et al. (1993). Imaging at high magnetic fields: Initial experiences at 4 T. *Magn. Res. Q.* 9, 259–277.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858. doi: 10.1002/hbm.20547

Van Overwalle, F., and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584. doi: 10.1016/j.neuroimage.2009.06.009

Volz, K. G., Schubotz, R. I., and von Cramon, D. Y. (2005). Variants of uncertainty in decision-making and their neural correlates. *Brain Res. Bull.* 67, 403–412. doi: 10.1016/j.brainresbull.2005.06.011

Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited - again. *Neuroimage* 2, 173–181. doi: 10.1006/nimg.1995.1023

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: 31 January 2014; accepted: 04 June 2014; published online: 24 June 2014.
*Citation: Schubotz RI, Wurm MF, Wittmann MK and von Cramon DY (2014) Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. Front. Psychol. **5**:636. doi: 10.3389/fpsyg.2014.00636*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Schubotz, Wurm, Wittmann and von Cramon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 20 May 2014
doi: 10.3389/fpsyg.2014.00441

Prediction-learning in infants as a mechanism for gaze control during object exploration

***Matthew Schlesinger*[1]*, *Scott P. Johnson*[2] *and Dima Amso*[3]**

[1] *Department of Psychology, Southern Illinois University Carbondale, Carbondale, IL, USA*
[2] *Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA*
[3] *Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
George L. Malcolm, The George Washington University, USA
Anne Sanda Warlaumont, University of California, Merced, USA

***Correspondence:**
Matthew Schlesinger, Department of Psychology, Southern Illinois University Carbondale, Life Science II, Room 281, Carbondale, IL 62901, USA
e-mail: matthews@siu.edu

We are pursuing the hypothesis that visual exploration and learning in young infants is achieved by producing gaze-sample sequences that are sequentially predictable. Our recent analysis of infants' gaze patterns during image free-viewing (Schlesinger and Amso, 2013) provides support for this idea. In particular, this work demonstrates that infants' gaze samples are more easily learnable than those produced by adults, as well as those produced by three artificial-observer models. In the current study, we extend these findings to a well-studied object-perception task, by investigating 3-month-olds' gaze patterns as they view a moving, partially occluded object. We first use infants' gaze data from this task to produce a set of corresponding center-of-gaze (COG) sequences. Next, we generate two simulated sets of COG samples, from image-saliency and random-gaze models, respectively. Finally, we generate learnability estimates for the three sets of COG samples by presenting each as a training set to an SRN. There are two key findings. First, as predicted, infants' COG samples from the occluded-object task are learned by a pool of simple recurrent networks faster than the samples produced by the yoked, artificial-observer models. Second, we also find that resetting activity in the recurrent layer increases the network's prediction errors, which further implicates the presence of temporal structure in infants' COG sequences. We conclude by relating our findings to the role of image-saliency and prediction-learning during the development of object perception.

Keywords: object perception, prediction-learning, infant development, eye movements, visual saliency

INTRODUCTION

The capacity to perceive and recognize objects begins to develop shortly after birth (e.g., Fantz, 1956; Slater, 2002). A critical skill that emerges during this time and supports object perception is gaze control, that is, the ability to direct gaze toward informative or distinctive regions of an object, such as edges and contours, as well as to shift gaze from one part of the object to another (e.g., Haith, 1980; Bronson, 1982, 1991). There are a number of relatively well-studied mechanisms that help drive the development of gaze control – in particular, during infants' visual object exploration – including improvements in acuity and contrast perception, inhibition-of-return, and selective attention (e.g., Banks and Salapatek, 1978; Clohessy et al., 1991; Dannemiller, 2000). While these mechanisms help to explain when, why, and in which direction infants shift their gaze, they may offer limited explanatory power in accounting for gaze-shift patterns at a more fine-grained level (e.g., the particular visual features sampled by the fovea at the next fixation point).

In the current paper, we present and evaluate a microanalytic approach for analyzing infants' gaze shift sequences during visual exploration. Specifically, we convert the sequence of fixations produced by each infant into a stream of "center-of-gaze" (or COG) image samples, where each sample approximates the portion of the image visible to the fovea of a human observer while fixating the given location on the image (for a related approach, see Dragoi and Sur, 2006; Kienzle et al., 2009; Mohammed et al., 2012). We then use a simple recurrent network (SRN) as a computational tool for estimating the presence of temporal or sequential structure within infants' COG gaze patterns.

The rationale for our analytical strategy is guided by two key ideas: first, that a core learning mechanism in infancy is driven by the detection of statistical regularities in the environment (e.g., Saffran et al., 1996), and second, that a wide range of infants' exploratory actions, such as visual scanning and object manipulation, are future-oriented (e.g., Haith, 1994; Johnson et al., 2003; von Hofsten, 2010). Together, these ideas suggest that infants' ongoing gaze patterns are predictive or prospective. Thus, our primary hypothesis is that if infants' gaze patterns are sequentially structured, we should then find that the stream of recent fixations toward an object or scene will provide sufficient information to predict the content of upcoming fixations. A related hypothesis is, given that sequential structure is observed in infants' gaze patterns, these sequences should be more predictable (i.e., more easily learned by an SRN) than those generated by other types of observers (e.g., human adults, ideal, or artificial observers, etc.).

Our recent work has provided preliminary support for both of these hypotheses. In particular, we compared the gaze sequences produced by 3-month-old infants and adults during an image

free-viewing task with those from three sets of artificial observers (i.e., image-saliency, image-entropy, and random-gaze models) that were presented with the same natural images (Schlesinger and Amso, 2013; Amso et al., 2014). The real and artificial observers' fixation data were first transformed into corresponding sequences of COG samples. We then measured the learnability of the five sets of COG image sequences by presenting each set to an SRN, which was trained to reproduce the corresponding sequences. A key finding from this work, over two simulation studies, was that the COG sequences produced by the human infants resulted in both more accurate and rapid learning than the adult COG sequences, or any of the three artificial-observer sequences.

In the current paper, we extended our model in a number of important ways to investigate the development of object perception in 3-month-olds. First, our dataset derives from a paradigm called the *perceptual-completion task*, which is specifically designed to assess infants' perception of a moving, partially occluded object (Kellman and Spelke, 1983; Johnson and Aslin, 1995). **Figure 1A** illustrates this occluded-rod display, which is presented first to infants, and then repeated until they habituate to the display. Two subsequent displays are then presented to infants and used to probe their perception and memory of the occluded-rod display (see **Figures 1B,C**). Because our focus here is on infants' initial gaze patterns at the beginning of the task, before they have accumulated extensive experience with the display, we therefore restrict our analyses to gaze data from the first trial of the occluded-rod display. Although this display is somewhat simplified relative to the natural images from our previous study, it also has the benefit that infants will likely devote much of their attention to either of the two primary objects in the scene (i.e., the moving rod and/or the occluder), thereby producing a rich source of object-directed gaze data to analyze.

A second important advance in the current paper concerns how the artificial-observer gaze patterns are produced. Specifically, in our previous model, several parameters of the artificial observers were left to vary freely, which resulted in systematic differences between the kinematics of the gaze patterns produced by the human-infant and artificial observers. For example, the artificial observers generated significantly longer gaze shifts than the infants. We address this issue in the current model by carefully yoking the gaze patterns of each artificial observer to a corresponding individual infant, so that the average kinematic measures were the same for each observer group.

A third advance is that we also simplified the architecture of the model used to learn the COG sequences. In particular, our previous model focused specifically on the process of visual exploration, including a component in the model that simulated an *intrinsically motivated learner* (i.e., an agent that is motivated to improve its own behavior, rather than to reach an externally defined goal). However, because the issue of intrinsic motivation is not central to the current paper, we have stripped this component from the model, resulting in a more direct and straightforward method for assessing the relative learnability of the COG sequences produced by each of the observer groups.

In the next section, we provide a detailed description of (1) the procedure used to transform infants' gaze data into COG sequences, (2) the comparable steps used to generate the artificial observers' gaze data and COG sequences, and (3) the training regime employed to measure COG sequence learnability. In the meantime, we briefly sketch the procedure here, followed by our primary hypotheses and analytical strategy.

The infant gaze data were obtained from a sample of 3-month-old infants who viewed the occluded-rod display illustrated in **Figure 1A**. Fixation locations for each infant were acquired by an automated eye-tracker. These locations were then mapped to the corresponding spatial position and frame number from the occluded-rod display, and a small (41 × 41 pixel) image sample, centered at the fixation location, was obtained for each gaze point. Next, two sets of artificial gaze sequences were generated. First, an image-saliency model was used to produce a sequence of gaze points in which gaze direction is determined by bottom-up visual features, such as motion or regions with strong light/dark contrast (e.g., Itti and Koch, 2000). Second, in the random-gaze model, locations were selected at random from the occluded-rod display. Each of the artificial-observer models was used to generate a set of COG sequences, with each sequence in the set yoked to the timing and gaze-shift distance of a corresponding infant.

Given our previous findings with the image free-viewing paradigm, our primary hypothesis was that the COG sequences produced by infants during the occluded-rod display would be more easily learned by a set of SRNs than either of the two artificial-observer sequences. We evaluated this hypothesis by

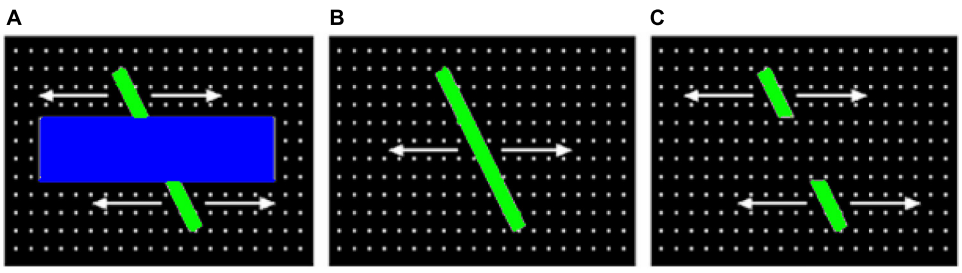


FIGURE 1 | Displays used to assess perceptual completion in infants: (A) occluded-rod (habituation) display, and (B) complete-rod and (C) broken-rod test displays.

assigning an SRN to each of the infants, and then training each network simultaneously on the three corresponding COG sequences (i.e., the infant's sequence, plus the yoked image-saliency and random-gaze sequences). Learning was implemented in each SRN by presenting it with the three corresponding COG sequences, one image sample at a time as input, and then using a supervised learning algorithm to train the SRN to produce as output the next image sample from the sequence. We then assessed *learnability* by ranking the three observers assigned to each SRN by mean prediction error after each training epoch. Given this measure, we predicted that infants would not only have the highest average rank at the start of training (i.e., their COG sequences would be learned first by the SRNs), but also that this difference would persist throughout training.

In addition, we also probed the training process further by exploring the effect of manipulating the context units on the performance of the SRN. In particular, we implemented a "forgetting function" in which the context units were reset at one of three intervals (every 1, 2, or 5 COG training samples; for a related discussion, see Elman, 1993). In the most extreme condition, resetting the context units after each COG sample enabled us to determine if the network was learning exclusively on the basis of each current COG sample – in which case, the 1-sample reset would have no impact on performance – or alternatively, if the memory trace of recent COG samples encoded within the recurrent pathway was also being used as a predictive cue. Accordingly, we predicted that resetting the context layer units would not only impair performance of the SRN, but also that this interference effect would be greatest for the infants' COG sequences.

It is important to stress in the 2- and 5-sample reset conditions, though, that this trace accumulates in a fashion that weights the memory toward COG samples that are more distal in time (i.e., past COG samples are not weighted equally). For example, in the 5-sample case, the first COG sample in a wave of five is effectively presented to the network as input (directly or indirectly) four times: once as the first COG sample, and then four more times as the trace of the sample cycles through the context units. By this logic, the fourth COG sample in the same wave of five is presented twice. Thus, the forgetting function provides a somewhat qualitative method for revealing whether or not sequential or temporal structure is present in infants' COG image samples, but may not directly specify how those regularities are distributed over time. We return to this issue in the discussion and raise a potential strategy for addressing it.

STIMULI

OCCLUDED-ROD DISPLAY

During the collection of eye-tracking data (see below), the occluded-rod display was rendered in real-time. In order to convert this display into a sequence of still frames for the current simulation study, it was first captured as a video file (AVI format, 1280 × 1024 pixels, 30 fps), and then parsed by Matlab into still frames. A complete cycle of the rod's movement, from the starting position on the far right, to the far left, and then back to the starting location, was extracted from the video and resulted in 117 frames (∼3.5 s in real-time). Note that during video presentation, the dimensions of the occluded-rod display were 480 × 360 pixels, which was presented at the center of the monitor, surrounded by a black border. This border was subsequently cropped from the still-frame images, so that the occluded-rod display filled the frame. The gaze data obtained from infants were adjusted to reflect this cropping process; meanwhile, as we describe below, the simulated gaze data from the image-saliency and random-gaze models were obtained by presenting the cropped (480 × 360) occluded-rod displays to each model.

OBSERVER GROUPS

Infants

Twelve 3-month-old infants (age, $M = 87.7$ days, SD = 12 days; 5 females) participated in the study. Infants sat on their parents' laps approximately 60 cm away from a 76 cm monitor in a darkened room. Eye movements were recorded using the Tobii 1750 remote eye tracker. Before the beginning of each trial, an attention-getter (an expanding and contracting children's toy) was used to attract infants' gaze to the center of the screen. As soon as infants fixated the screen, the attention-getter was replaced with the experimental stimulus and timing of trials began. Each trial ended when the infant looked away for 2 s or when 60 s had elapsed. Note that all analyses described below were based on the eye-tracking data acquired during each infant's first habituation trial (i.e., the occluded-rod display).

Image-saliency model

The saliency model was designed to simulate the gaze patterns of an artificial observer whose fixations and gaze shifts are determined by image salience, that is, by bottom-up visual features such as motion and light/dark contrast. In particular, the 117 still frames extracted from the occluded-rod display were transformed into a set of corresponding saliency maps by first creating four feature maps (tuned to motion, oriented edges, luminance, and color contrast, respectively) from each still-frame image, and then summing the feature maps into a saliency map. The sequence of 117 saliency maps was then used to generate a series of simulated fixations. We describe each of these processing steps in detail below.

Feature maps. Each of the still-frame images was passed through a bank of image filters, resulting in four sets of feature maps: one motion map (i.e., using frame-differencing between consecutive frames), four oriented edge maps (i.e., tuned to 0°, 45°, 90°, and 135°), one luminance map, and two color-contrast maps (i.e., red–green and blue–yellow color-opponency maps). In addition, this process was performed over three spatial scales (i.e., to capture the presence of the corresponding features at high, medium, and low spatial frequencies), by successively blurring the original image and then repeating the filtering process [for detailed descriptions of the algorithms used for each filter type, refer to Itti et al. (1998) and Itti and Koch (2000)]. As a result, 24 total feature maps were computed for each still-frame image.

Saliency maps. Each saliency map was produced by first normalizing the corresponding feature maps (i.e., by scaling the values on each map between 0 and 1), and summing the 24 maps together. For the next step (simulating gaze data), each

saliency map was then downscaled to 40 × 30. These resulting saliency maps were then normalized, by dividing each map by the average of the highest 100 saliency values from that map. **Figure 2** illustrates a still-frame image from the occluded-rod display on the left, and the corresponding saliency map on the right.

Simulated gaze data. Next, 12 sets of simulated gaze sequences were produced with the image-saliency model. Each set was yoked to the gaze data from a specific infant, and in particular, four dimensions of the infant and artificial-observer gaze sequences were equated: (1) the location (i.e., gaze point) of the first fixation, (2) the total number of fixations, (3) the duration of each fixation (i.e., dwell-time), and (4) the distance traveled between each successive fixation (i.e., gaze-shift distance).

At the start of the simulated trial, the image-saliency model's initial gaze point was set equal to the location of the infant's first fixation. The model's gaze point was then held at this location for the same duration as the infant's. For example, if the infant's initial fixation was 375 ms, the model's gaze point remained at the same location for 11 frames (i.e., 375 ms ÷ 33 ms/frame = 11 frames). In a comparable manner, each gaze shift produced by the image-saliency model was therefore synchronized with the timing of the corresponding infant's gaze shift.

Subsequent fixation locations were selected by the image-saliency model by iteratively updating a fixation map for the duration of the fixation. The fixation map represents the difference between the *cumulative* saliency map (i.e., the sum of the saliency maps that span the current fixation) and a decaying inhibition map (see below). Note that the inhibition map served as an analog for an inhibition-of-return (IOR) mechanism, which allowed the saliency model to release its gaze from the current location and shift it to other locations on the fixation map.

Each trial began by selecting the initial fixation as described above. Next, the inhibition map was initialized to 0, and a 2D Gaussian surface was added to the map, centered at the current fixation point, with an activation peak equal to the value at the corresponding location on the saliency map. The Gaussian surface spanned a 92 × 92 pixel region, slightly larger than twice the size of a single COG sample (see COG Image Sequences, below). Over the subsequent fixation duration, activity on the inhibition map decayed at a rate of 10% per 33 ms. At the end of the fixation, the next fixation point was selected: (a) the fixation map was updated by subtracting the inhibition map from the saliency map (negative values were set to 0), (b) the top 500 values on the saliency map were chosen as potential target locations, and (c) the gaze-shift distance between the current fixation and each target location was computed. Finally, the target location with the gaze-shift distance closest to that produced by the infant (on the corresponding gaze shift) was selected as the next fixation location (any ties were resolved with a simulated coin-toss). The process continued until the model produced the same number of fixations as the corresponding infant (note that the sequence of 117 saliency maps were repeated as necessary).

Random-gaze model

The random-gaze model was designed as a control condition, to simulate the gaze pattern of an observer who scanned the occluded-rod display by following a policy in which all locations (at a given distance from the current gaze point) are equally likely to be selected. Thus, the gaze sequences were produced by the random-gaze model following the same four constraints as those for the image-saliency model (i.e., number and duration of fixations, gaze-shift distance, etc.), with the one key difference that upcoming fixation locations were selected at random (rather than based on image salience).

To help provide a qualitative comparison between typical gaze patterns produced by the three types of observers, **Figure 3** presents the cumulative scanplot from one of the infants (**Figure 3A**), as well as the corresponding scanplots from the image-saliency and random-gaze models that were yoked to the same infant (**Figures 3B,C**, respectively).

SUMMARY STATISTICS

Prior to the training phase, we computed summary statistics for the three models, in order to verify that the yoking procedure resulted in comparable performance patterns for each yoked dimension. **Table 1** presents the mean summary statistics for the three observer groups (with standard deviations presented in parentheses). Note that the values presented in italics represent two of the four

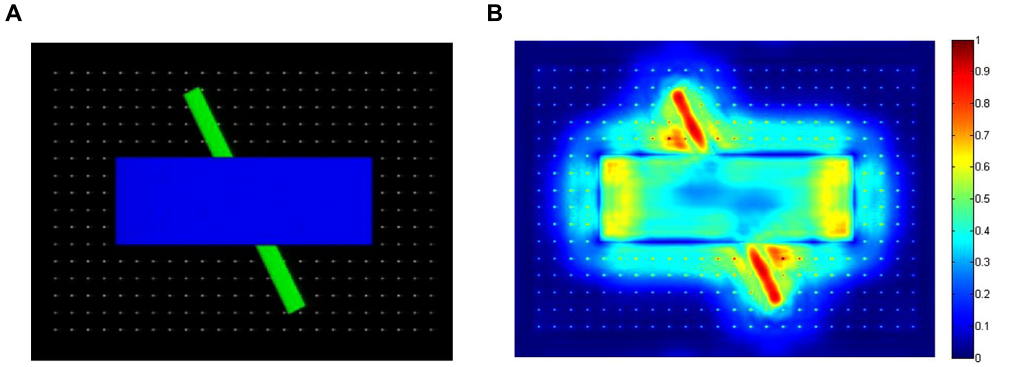


FIGURE 2 | Illustration of one of the still-frame images from the occluded-rod display (A), and the corresponding saliency map (B).

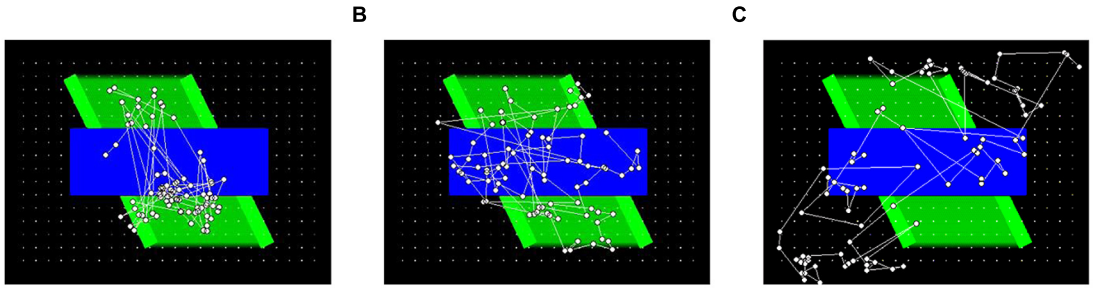


FIGURE 3 | Scanplot (sequence of fixation points) produced by one of the infants (A), together with the corresponding scanplots from the yoked image-saliency (B) and random-gaze models (C).

Table 1 | Summary statistics as a function of observer group.

	Fixation duration	Saliency captured	Revisit rate	Fixation dispersion	Gaze-shift distance
Infant	*339.38*	0.66	0.23	78.55	*59.20*
	(96.03)	(0.07)	(0.07)	(15.08)	*(18.82)*
Saliency	*356.19*	0.65	0.19	82.46	*60.36*
	(95.47)	(0.03)	(0.11)	(18.68)	*(18.44)*
Random	*356.19*	0.47*	0.16	110.60*	*59.21*
	(95.47)	(0.05)	(0.08)	(28.75)	*(18.82)*

*$p < 0.01$ (paired comparison vs. infant observer group). Standard deviation presented in parentheses; values in italics correspond to the two measures that were yoked across the three observer models.

dimensions (i.e., fixation duration and gaze-shift distance) that were systematically equated between observer groups. In general, except where noted below, *post hoc* comparisons across the three observer groups revealed no significant differences. The first column presents the mean fixation duration (in milliseconds) for the infant, image-saliency, and random-gaze groups. The net difference between real and artificial observers was approximately 17 ms, and was presumably due to the fact while the infant data were measured continuously, the artificial observers were simulated in discrete time steps of 33.3 ms.

The second column presents the mean saliency "captured" by each model, that is, the degree to which each group's fixations were oriented toward regions of maximal saliency in the display. This was computed by projecting the gaze points produced by each of the observer groups on to the corresponding saliency maps, and then calculating the average saliency for those locations. Recall that values on the saliency maps were scaled between 0 and 1; the average saliency values from each group therefore reflected the proportion of optimal or maximal saliency captured by that group. There are two key results. First, the saliency model achieved an average of 0.65 saliency, indicating that – due to the constraint imposed on allowable gaze-shift distance – the model did not consistently fixate the most salient locations in the display. The second noteworthy finding is that infants' gaze patterns captured a comparable level of saliency, that is, 0.66. As **Table 1** notes, the average saliency captured by the random observer group was significantly lower than the infant and image-saliency groups [both $ts(22) > 8.46$, $ps < 0.001$].

The third column presents the mean revisit rate for each observer group. Revisit rate was estimated by first creating a null frequency map (a 480 × 360 matrix with all locations initialized to 0). Next, for each fixation, the values within a 41 × 41 square (centered at the fixation location) on the frequency map were incremented by 1. This process was repeated for all of the fixations generated by an observer, and the frequency map was then divided by the number of fixations. For each observer, the maximum value from this map was recorded, reflecting the location in the occluded-rod display that was *most frequently* visited (as estimated by the 41 × 41 fixation window). The maximum value was then averaged across observers within each group, providing a metric for the peak proportion of fixations that a particular location in the occluded-rod display was visited, on average. As **Table 1** illustrates, a key finding from this analysis is that infants had the highest revisit rate (23%), while the two artificial observer groups produced lower rates.

The last two columns present kinematic measures of the gaze patterns. First, dispersion was computed by calculating the centroid of the fixations (i.e., the mean fixation location), then calculating the mean distance of the fixations (in pixels) from the centroid for each observer, and then averaging the resulting dispersion values for each group. As Figure, **Table 1** indicates, infants tended to have the least-disperse gaze patterns. Fixation dispersion in the image-saliency observer group did not differ significantly from the infant group, although it was significantly higher in the random-observer group [$t(22) = 3.63$, $p < 0.01$]. Finally, the fifth column presents the mean gaze shift distance (measured in pixels) for each group. Because this measure was yoked across groups, as expected, the artificial-observer groups produced mean gaze-shift distances that were comparable to the infants' mean distance.

COG IMAGE SEQUENCES

The final step, prior to training the model, was the process of mapping each set of gaze patterns into a sequence of COG image samples. This was accomplished by determining the frame number that corresponded to the start of each fixation, projecting the gaze point on to the resulting still-frame image, and then sampling a 41 × 41 pixel image, centered at that location. The dimensions of the COG sample were derived from the display size and infants'

viewing distance, and correspond to a visual angle of 1.8°, which falls within the estimated range of the angle subtended by the human fovea (Goldstein, 2010). In order to facilitate the training process, note that each of the COG samples was converted from color (RGB) to grayscale.

MATERIALS AND METHODS

MODEL ARCHITECTURE AND LEARNING ALGORITHM

Recall that our primary hypothesis was that infants' COG sequences would be more easily learned by an SRN than the sequences from the two artificial-observer models. To evaluate this hypothesis, we trained a set of 3-layer Elman networks, with recurrent connections from the hidden layer back to the input layer (context units; Elman, 1990). In particular, this architecture implements a forward model, in which the current sensory input (plus a planned action) is used to generate a prediction of the next expected input (e.g., Jordan and Rumelhart, 1992). The complete model (including the training stimuli, network architecture, and learning algorithm) was written and tested by the first author (Schlesinger) in the Matlab programming environment.

The input layer of the SRN was composed of 2083 units, including 1681 units that encoded the grayscale pixel values of the current COG sample, 400 context units (which copied back the activity of the hidden layer from the previous time step), and two input units that encoded the x- and y-coordinates of the upcoming COG sample (normalized between 0 and 1). The input layer was fully connected to the hidden layer (400 hidden units, i.e., approximately 75% compression of the COG sample), which in turn was fully connected to the output layer (1681 units). The standard logistic function was used at the hidden and output layers to maintain activation values between 0 and 1; in addition, the bias terms were fixed to 0 for the hidden and output units.

An individual training trial proceeded as follows: given the selection of a COG sequence, the first COG sample in the sequence was presented to the SRN. For this first sample, the activation of the context units was set to 0.5. Activity in the network was propagated forward, resulting in the predicted next COG sample. This output was compared to the second COG sample in the sequence, and the root mean-squared error (RMSE) was calculated. Next, the standard backpropogation-of-error (i.e., backprop) learning algorithm was used to adjust the SRN's connection weights (i.e., training was pattern-wise). The activation values from the hidden layer were then copied back to the input layer, and the second COG sample was presented to the SRN. This process continued until the second-to-last COG sample in the sequence was presented.

TRAINING REGIME

A total of 10 training runs were conducted. At the start of each run, a single SRN was initialized with random connection weights between 0 and 1, which were then divided by the number of incoming units to the given layer (i.e., fan-in). This network was cloned 12 times, once for each of the infants. This duplication process ensured that any subsequent performance differences between SRNs during a run were due to the training samples unique to each infant, rather than to the initialization procedure.

Accordingly, each of the 12 SRNs was paired with one of the infants, and subsequently trained on the three COG sequences associated with that infant: the selected infant's sequence, as well as the image-saliency and random-gaze sequences that were yoked to the same infant. A single training epoch was defined as a sweep through the three COG sequences. Order of observer type (i.e., infant, saliency, random) was randomized for each epoch. Pilot data collection indicated that the SRNs reached asymptotic performance, with a learning rate of 0.1, between 200 and 300 training epochs. As a result, each training run continued for 300 epochs.

In order to evaluate our second hypothesis – that resetting the activation of the context layer would have the largest interference effect on the infants' COG sequences – we "paused" training every 10 epochs to test each of the SRNs. During the testing phase, learning was turned off and all connections were frozen in the SRN. Next, the SRN was tested by presenting the three COG sequences, four times each: (1) with recurrence functioning normally, and (2–4) with the activity of the context units reset to 0.5 every 1, 2, or 5 input steps, respectively.

RESULTS

Two sets of planned analyses were conducted. First, we converted RMSE values into rank scores, and then compared the performance of the 12 SRNs as a function of mean rank of each observer group. In particular, this analysis focused on our predictions that the COG sequences from the infant group would have the highest mean ranking at the start of training, and that this difference would persist throughout the training period. The second analysis examined the influence of resetting the context-layer units on the SRNs' performance, which allowed us to indirectly measure the presence of temporal dependencies in the COG sequences, between both adjacent samples as well as those as many as five samples apart.

Figure 4 presents the RMSE produced by the 12 SRNs during the 300 training epochs, as a function of the observer group (i.e., infant, image-saliency, and random-observer models, respectively). Note that these data are pooled over the 12 SRNs and the 10 training runs. In addition, the RMSE values presented in **Figure 4** were those generated by the SRNs during the test phase, that is, in which learning was turned off every 10 epochs. As a result, these data reflect the performance of the SRNs while removing the transient effect of testing order (i.e., recall that the order of the observer groups during training was randomized across epochs).

There are two important trends suggested by **Figure 4**. First, the RMSE values produced by the image-saliency group remain consistently highest during training. Second, there is an early "trade-off" between the infant and random-gaze groups, which eventually results in a stable difference, favoring the infant group. In order to determine whether these trends were statistically reliable, we first converted the RMSE values into ranks. In particular, for each epoch, the RMSE for the three observer groups were sorted in ascending order, and assigned the corresponding rank (i.e., 1, 2, or 3). As before, ranks were then averaged over the 12 SRNs and 10 training runs.

Figure 5 presents the rank-transformed performance data. (Note that in describing these data, we adopt the convention that the rank of 1 is treated as "highest" while the rank of 3 is the

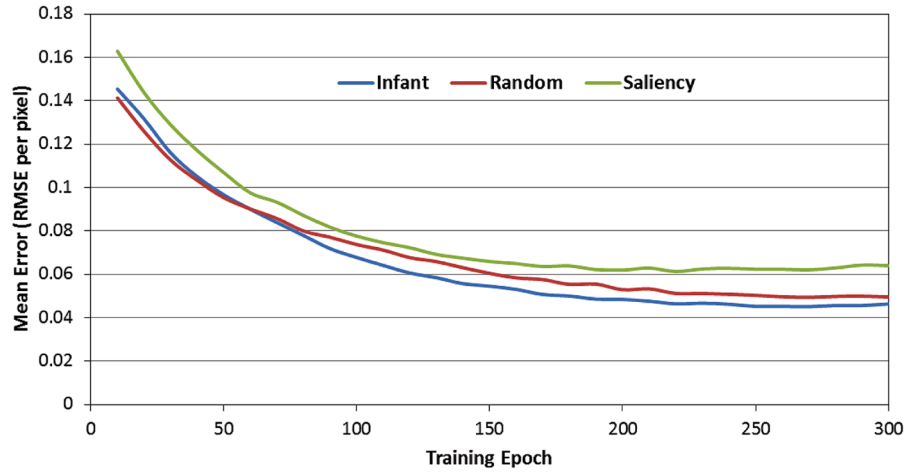


FIGURE 4 | Mean prediction error (MRSE per pixel) over the 300 training epochs, as a function of the three observer groups.

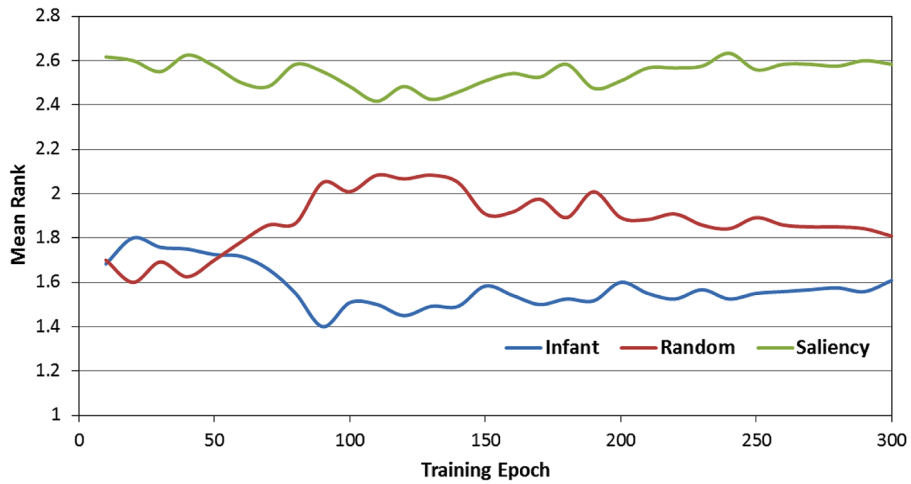


FIGURE 5 | Mean rank scores over the 300 training epochs, as a function of the three observer groups.

"lowest." In other words, a higher average rank corresponds to a lower RMSE). In order to compare the three observer groups, a 2-way ANOVA was conducted with epoch and observer group as the two factors. As expected, there was a main effect of observer group [$F(2,357) = 124.24$, $p < 0.001$]. We examined this effect with planned paired comparisons between the three groups (using Bonferroni corrections), which also confirmed our prediction: specifically, the infant observer group had significantly higher overall mean rank than the image-saliency and random-gaze groups. However, these findings were qualified by a significant epoch × observer group interaction [$F(58,10353) = 6.48$, $p < 0.001$]. As **Figure 5** indicates, near the start of training, the infant and random-gaze groups had similar ranks; in contrast, a large, stable difference between the two groups emerged after approximately 50 epochs.

In order to examine this interaction, we conducted a *post hoc* analysis by first dividing training time into two phases (0 to 50 and 60 to 300 epochs). We then repeated the previous 2-way ANOVA for each phase (i.e., epoch × observer group), including comparisons between the three observer groups. This analysis revealed that while there was no significant difference between the infant and random-gaze groups during the first 50 epochs ($p = 0.64$), the infant group averaged a significantly higher rank than the random-gaze group during the remaining 250 epochs ($p < 0.005$). In particular, these results confirm our prediction that the infant observer group would be ranked highest at the start of training, albeit after an initial period of equivalent performance in two of the three groups. In addition, the stability of this pattern for the remainder of the training phase also provides support for our prediction that the infant observer group would maintain the highest rank throughout training.

The second set of analyses focused on the role of the context layer in the SRN architecture, and more specifically, on the question of whether periodically resetting the activity of this layer

during training would disrupt performance. In order to address this question, recall that during each test phase, each of the SRNs was not only tested under canonical conditions (e.g., full recurrence; see **Figure 4**), but also under three conditions in which the context layer was reset (i.e., all values were set to 0.5) after every 1, 2, or 5 training samples. Because it was anticipated that resetting the context layer would produce an increase in prediction errors, RMSE difference scores were therefore computed between each of the reset conditions and the canonical condition. These difference scores were then transformed into percent-change scores, relative to the canonical condition (that is, percent increase in the RMSE due to resetting the context layer). **Figure 6** presents the resulting percent-change values for each of the observer groups, within the three reset conditions (i.e., 6A = every sample, 6B = every two samples, and 6C = every five samples, respectively).

There are three primary findings from this analysis. First, a consistent pattern observed across the three observer groups and reset conditions is that the percent change of the RMSE starts near 0 at the beginning of training. However, for all groups and conditions, this value quickly increases, reflecting a progressively greater impact of resetting the context layer over training time. For example, **Figure 6A** illustrates that by the end of training, resetting the context layer after each COG sample results in approximately a 200% increase in the RMSE, on average for the three observer groups. Second, there is a positive association between the reset frequency and the percent increase in RMSE. In other words, resetting the context layer after every sample produced a larger interference effect than resetting every two samples, and likewise for resetting every five samples.

Third, we conducted a 2-way ANOVA for each of the reset conditions, again with epoch and observer groups as the two factors. This comparison revealed a significant epoch × observer group interaction for all three reset conditions [all Fs(58, 10353) > 3.87, ps < 0.001]. In general, as **Figure 6** illustrates, this interaction reflects the tendency for percent-change scores to begin near 0 for each of the observer groups, and then subsequently increase at different rates over training time. We pursued this interaction by dividing training time into three blocks of epochs (i.e., 0–100, 100–200, and 200–300 epochs), and then conducting a simple-effects test of observer group for each of the three blocks. Two consistent findings emerged from this test. First, across each of the three training blocks and two of the three reset conditions, the percent increase of the RMSE in the infant group was significantly higher than in the random-gaze group [all ts(238) > 2.79, ps < 0.02]. The only exception to this result was in the condition where the context layer was reset every five samples, during the final block of epochs; in this case, the infant and random-gaze groups did not significantly differ. Second, a significant difference between the infant and saliency groups was not present during the first two blocks of epochs (i.e., through epoch 200). However, by the third block of epochs, the percent increase in RMSE in the infant group was significantly higher than in the saliency group, for all three reset conditions [all ts(238) > 2.38, ps < 0.05]. Taken together, these findings collectively support our prediction that resetting the context-layer activation values would have the largest interference effect on the infants' COG sequences.

DISCUSSION

The current simulation study focused on two goals. First, we sought to demonstrate that our previous gaze-sequence learnability findings, from an infant free-viewing task (Schlesinger and Amso, 2013), would generalize and extend to a task that was specifically designed to study object perception in young infants. Second, we not only implemented several key improvements in our model, but also modified the training and testing procedure to allow us to assess whether learnability of the infants' COG samples was due, at least in part, to the presence of sequential dependencies between both adjacent and non-adjacent training samples.

The results were consistent with each of our four hypotheses. First, we predicted that infants' COG sequences would be learned first by the 12 SRNs. We assessed this prediction by converting each observer group's error scores into ranks and then analyzing the respective ranks over 300 epochs of training time. As we predicted, the infant group eventually established a significant advantage over the other two observer groups. Unexpectedly, however, this advantage did not appear at the onset of training. Instead, the average ranks of the infant and random-gaze groups were comparable for the first 50 epochs of training. One potential explanation for this early similarity of performance in the two observer groups is that there was a higher initial "learning cost" associated with the infant group, due to the (presumed) presence of temporal dependencies in their COG sequences, which ostensibly required additional time for the SRNs to detect and exploit (through the context layer). Second, we also predicted that this advantage would persist and remain stable across the remaining time. Again, the results supported our prediction.

Our third and fourth predictions focused on whether the success of the SRN architecture in learning the infants' COG sequences benefited from the (presumed) presence of temporal or sequential dependencies embedded within the infants' COG training samples. Luckily, the use of the random-gaze model provides a critical role in addressing this question, as the gaze sequences from this model were specifically produced with a stochastic procedure (although it should be noted that the selection of each gaze point was constrained by a fixed gaze-shift distance rule). As a result, we can thus assume that there were no *a priori* regularities or dependencies within the random-gaze model's COG sequences, other than those broadly present in the display itself (e.g., the baseline probability of fixating the background, or the occluding screen, at random).

We therefore predicted that disrupting information flow within the recurrent pathway of the network by periodically resetting the context layer would increase the overall errors produced by the SRNs. Indeed, across all three observer groups we observed significant increases in the SRN prediction errors when the recurrent layer was reset. Our last prediction was that the interference effect would be greatest for the infants' COG sequences, and as **Figure 6** illustrates, this prediction was confirmed as well.

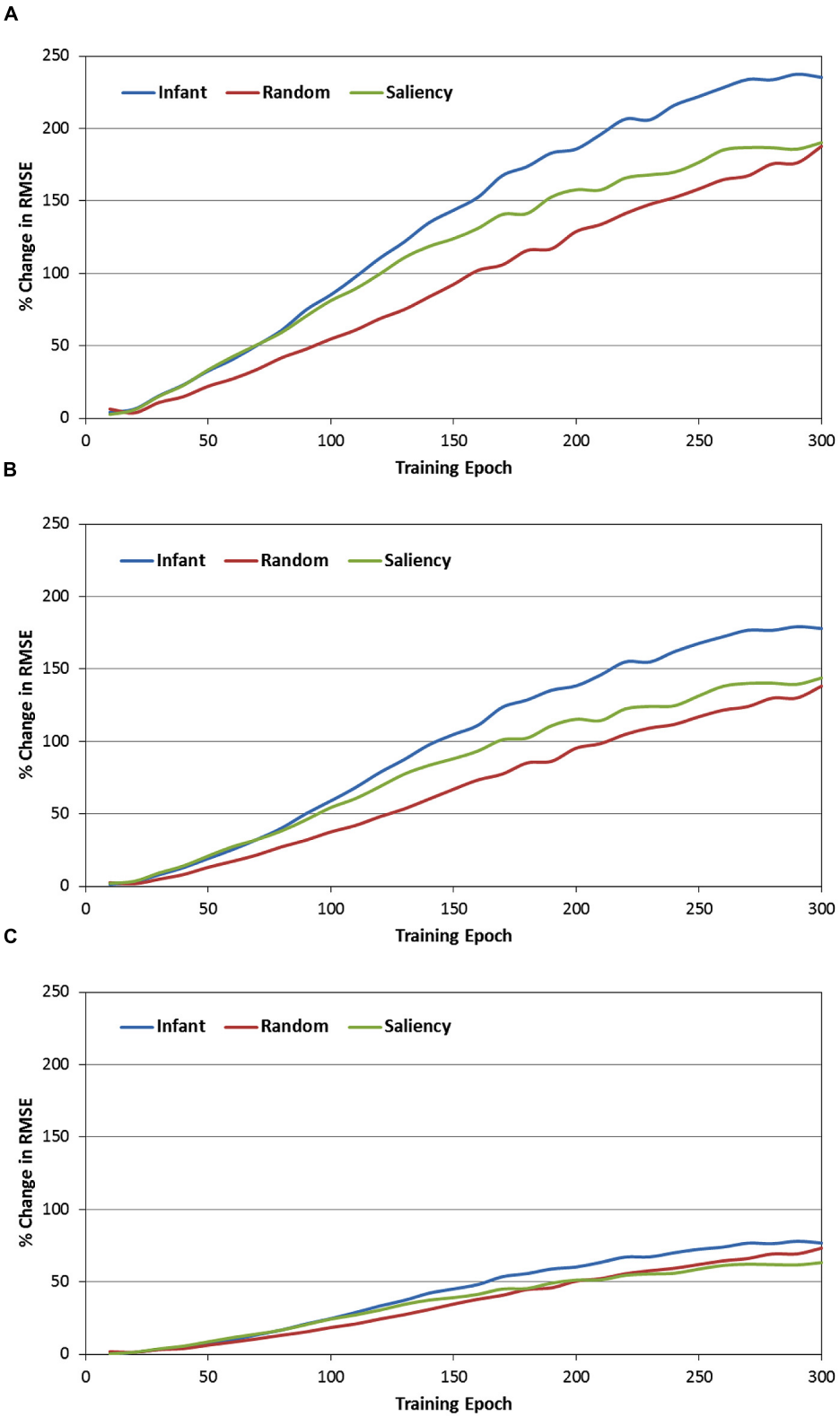


FIGURE 6 | Percent change in the MRSE during testing of the three observer groups, while resetting the recurrent layer units after every sample (A), every other sample (B), and every five samples (C).

Further inspection of **Figure 6** may offer three additional insights. First, as we suggested above, the gaze sequences produced by the random-gaze model should include minimal (if any) sequential structure. Nevertheless, note that – like the other two observer groups – the interference effect increased with training time in the random-gaze group. This trend provides a statistical baseline for estimating the contribution of the context layer for prediction learning on the current task, as the training sequences from the random-gaze model were ostensibly sequentially independent. We can therefore estimate the presence of any additional structure embedded within infants' COG sequences by subtracting the RMSE change values produced by the random-gaze model. For example, in the first reset condition (i.e., reset after every sample) and pooling over training time, the overall difference in RMSE change between the infant and random-gaze groups is 42%. This value provides an important clue toward understanding the function of infants' object-directed gaze behavior, as it demonstrates that infants' gaze sequences are significantly more structured than sequences produced by chance, and that this embedded sequential structure also provides a measurable advantage to an active observer that is learning to forecast or predict the content of upcoming fixations.

An additional insight offered by manipulating the context layer is reflected by the regular order of performance observed across the three observer groups. In particular, note that the interference effect was consistently lowest in the random-gaze group, highest in the infant group, and midway between the two in the image-saliency group. This finding suggests that the simple strategy of orienting toward relatively high-saliency regions in the occluded-rod display is sufficient to generate statistically reliable temporal structure in the COG sequences.

Finally, a third insight suggested by these findings is that image-saliency may provide, at best, a partial account for how infants' gaze patterns are structured over time and space. In particular, our previous work has demonstrated that a saliency-based model captures several global-level features of infants' gaze patterns, such as the frequency of fixations toward the rod segments, as well as individual differences in the rate of rod fixations between infants (Amso and Johnson, 2006; Schlesinger et al., 2007, 2012). In addition, our current model provides two additional pieces of evidence that also implicate the role of image saliency. First, as **Table 1** indicates, the infant and image-saliency groups fixated regions of the occluded-rod display that were on average nearly equal in salience. Second, as **Figure 6** illustrates, resetting the context layer had a comparable effect on the infant and image-saliency groups during the first 75–80 epochs of training (the same pattern was also consistent across the three reset conditions).

However, after approximately 80 epochs, the interference effect continued to increase at a faster rate in the infant group. One potential interpretation for this pattern is that, due to similar levels of saliency in the infants' and image-saliency models' COG samples, the SRNs "focused" during early learning on saliency-related features in the input (e.g., luminance contrast) as a predictive cue. In contrast, the random-gaze model fixated salient locations less frequently (i.e., 42% of maximal salience, vs. 66 and 65% in the infant and image-saliency models, respectively), and as a result, recurrent feedback in the SRN had less impact on prediction learning for the sequences from this observer model. If this reasoning is correct, it suggests that the subsequent performance split between the infant and image-saliency models was presumably due to additional temporal structure – beyond that provided by saliency – in the infants' sequences, which the SRNs continued to learn to detect and exploit. To put the point concisely: while infants and the image-saliency model fixated (on average) equally-salient regions in the occluded-rod display, we are proposing that it was *the particular temporal order in which infants scanned salient regions of the display* that provided an additional predictive cue to the SRNs. We are currently exploring computational strategies for teasing apart these spatial and temporal cues, and isolating their influence on the prediction-learning process.

Two key issues remain unaddressed by our work thus far. First, it is important to note that our use of the SRN architecture, as well as our manipulation of the context layer, provide a somewhat indirect method for identifying sequential structure in infants' COG samples. In general, this strategy tells us that temporal structure is present and it also provides a method for quantifying the interference caused by periodically resetting the context units, but it does not directly identify the visual features detected by the SRN, not does it specify how variation in these cues over time (i.e., correlations between successive COG samples) improves the outcome of sequence learning. An additional limitation of the reset method, which we noted in the introduction, is that the samples that are processed before a reset occurs do not contribute equally to the memory trace that accumulates in the recurrent pathway (i.e., distal samples are weighted more than recent samples).

There are several strategies available to address these issues. For example, alternative analytical methods (e.g., principal-component or clustering analysis of the hidden layer activations) as well as alternative modeling architectures and learning algorithms (e.g., Kohonen networks, Kalman filters, etc.) may provide additional insights. We are also currently exploring the strategy of constructing artificial gaze sequences in which we strictly control the statistical dependencies over time (e.g., alternating gaze between 2, or 3, or 4 narrowly defined regions in an image). Ideally, this will allow us to examine the influence of resetting the context layer versus learning/detecting temporal dependencies that vary in their duration over time. A related limitation of the modeling strategy we have employed here is that the SRNs were trained over multiple repetitions of the same COG sequences. In particular, this repetition provides an important learning cue to the SRNs, independent of the temporal structure embedded within the COG sequences. One way to address this issue is to employ a "leave-out" training regime, in which a subset of training patterns are set aside and reserved for testing the model.

Second, we should also note that our current simulation study focused exclusively on infants' first trial during the perceptual-completion task. An open question is whether infants' scanning patterns change systematically over subsequent trials (e.g., do rod fixations increase?), and if so, what effect if any will such changes have on the predictability of the COG sequences that are

produced during later trials? Our intuition is that if infants' gaze patterns during later trials are less variable (e.g., as estimated by our dispersion measure), *their COG sequences will become more predictable* (due to greater similarity between sequences). In addition, recall that after habituating to the occluded-rod display, infants then view the solid-rod and broken-rod test displays (**Figure 1**). Therefore, a related question is whether predictability of the COG sequences will increase or decrease during the test trials, and in particular, whether it will vary across the two display types. Answering these questions is essential to understanding the role of visual prediction-learning during the development of object perception.

We now return to the issue of early object-perception development in young infants. Our work has not only implicated the role of active visual scanning as an essential skill for object perception (Johnson et al., 2004; Amso and Johnson, 2006), but also demonstrated how this skill can emerge developmentally through interactions between the parietal and occipital cortex (Schlesinger et al., 2007). Recent work has also implicated visual prediction-learning as a complementary mechanism that may also support object perception (Schlesinger et al., 2011; Schlesinger and Amso, 2013). Our current findings help to integrate these ideas into a coherent developmental mechanism, by not only demonstrating that sequential structure is present within infants' time-ordered gaze patterns, but also that this structure is manifest across both complex, naturalistic displays as well as the relatively simplified ones that are used to investigate object perception in the laboratory. An additional important insight from both our recent behavioral and modeling work is that perceptual salience is likely a necessary, though not sufficient cue for driving visual scanning and object exploration in young infants (Schlesinger and Amso, 2013; Amso et al., 2014). We are optimistic that future work on this question will help to identify the other cues and sources of temporal structure that infants are learning to detect and exploit.

Finally, we conclude by noting that our modeling approach has the potential to offer two important innovations for the study of perceptual development in infants. First, our current strategy is to analyze infants' COG sequences offline, that is, *after they have been produced*. Thus, one of our long-term goals is to design an architecture that can accurately forecast infants' upcoming fixations *before they are produced*. One application of this forecasting technique would then be to manipulate the features or properties of the gaze destination before the infant gazed at that location, as a way of gauging their sensitivity to those features (i.e., a kind of gaze-contingent change-blindness paradigm). Second, we have previously observed variation across infants at the same age with visual displays such as the perceptual-completion task (e.g., Amso and Johnson, 2006). We are now excited to see if infants' performance on the perceptual-completion task will correlate with the relative learnability of the COG sequences they produce during the occluded-rod display, which would provide further support for the idea that individual differences in information pick-up have a fundamental effect on the development of object perception.

REFERENCES

Amso, D., Haas, S., and Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS ONE* 9:e85701. doi: 10.1371/journal.pone.0085701

Amso, D., and Johnson, S. P. (2006). Learning by selection: visual search and object perception in young infants. *Dev. Psychol.* 42, 1236–1245. doi: 10.1037/0012-1649.42.6.1236

Banks, M. S., and Salapatek, P. (1978). Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Investigat. Ophthalmol. Vis. Sci.* 17, 361–365.

Bronson, G. (1982). *The Scanning Patterns of Human Infants: Implications for Visual Learning*. Norwood, NJ: Ablex.

Bronson, G. (1991). Infant differences in rate of visual encoding. *Child Dev.* 62, 44–54. doi: 10.2307/1130703

Clohessy, A. B., Posner, M. I., Rothbart, M. K., and Vecera, S. P. (1991). The development of inhibition of return in early infancy. *J. Cogn. Neurosci.* 3, 345–350. doi: 10.1162/jocn.1991.3.4.345

Dannemiller, J. L. (2000). Competition in early exogenous orienting between 7 and 21 weeks. *J. Exp. Child Psychol.* 76, 253–274. doi: 10.1006/jecp.1999.2551

Dragoi, V., and Sur, M. (2006). Image structure at the center of gaze during free viewing. *J. Cogn. Neurosci.* 185, 737–748. doi: 10.1162/jocn.2006.18.5.737

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99. doi: 10.1016/0010-0277(93)90058-4

Fantz, R. L. (1956). A method for studying early visual development. *Percept. Mot. Skills* 6, 13–15. doi: 10.2466/pms.1956.6.g.13

Goldstein, B. (2010). *Sensation and Perception*. Belmont, CA: Wadsworth.

Haith, M. M. (1980). *Rules that Babies Look by: The Organization of Newborn Visual Activity*. Hillsdale, NJ: Erlbaum.

Haith, M. M. (1994). *The Development of Future-Oriented Processes*. Chicago: University of Chicago Press.

Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual-attention for rapid scene analysis. *IEEE Trans. Pattern Analysis Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558

Johnson, S. P., Amso, D., and Slemmer, J. A. (2003). Development of object concepts in infancy: evidence for early learning in an eye-tracking paradigm. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10568–10573. doi: 10.1073/pnas.1630655100

Johnson, S. P., and Aslin, R. N. (1995). Perception of object unity in 2-month-old infants. *Dev. Psychol.* 31, 739–745. doi: 10.1037/0012-1649.31.5.739

Johnson, S. P., Slemmer, J. A., and Amso, D. (2004). Where infants look determines how they see: eye movements and object perception performance in 3-month-olds. *Infancy* 6, 185–201. doi: 10.1207/s15327078in0602_3

Jordan, M. J., and Rumelhart, D. E. (1992). Forward models: supervised learning with a distal teacher. *Cogn. Sci.* 16, 307–354. doi: 10.1207/s15516709cog1603_1

Kellman, P. J., and Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognit. Psychol.* 15, 483–524. doi: 10.1016/0010-0285(83)90017-8

Kienzle, W., Franz, M. O., Schölkopf, B., and Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *J. Vis.* 9, 1–15. doi: 10.1167/9.5.7

Mohammed, R. A. A., Mohammed, S. A., and Schwabe, L. (2012). BatGaze: a new tool to measure depth features at the center of gaze during free viewing. *Brain Inform.* 7670, 85–96. doi: 10.1007/978-3-642-35139-6_9

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Schlesinger, M., and Amso, D. (2013). Image free-viewing as intrinsically-motivated exploration: estimating the learnability of center-of-gaze image samples in infants and adults. *Front. Psychol.* 4:802. doi: 10.3389/fpsyg.2013.00802

Schlesinger, M., Amso, D., and Johnson, S. P. (2007). The neural basis for visual selective attention in young infants: a computational account. *Adapt. Behav.* 15, 135–148. doi: 10.1177/1059712307078661

Schlesinger, M., Amso, D., and Johnson, S. P. (2011). "Increasing spatial competition enhances visual prediction learning," in *Proceedings of the First Joint IEEE Conference on Development and Learning and on Epigenetic Robotics*, eds A. Cangelosi, J.

Triesch, I. Fasel, K. Rohlfing, F. Nori, P.-Y. Oudeyer, M. Schlesinger, and Y. Nagai (New York: IEEE), 1–6.

Schlesinger, M., Amso, D., and Johnson, S. P. (2012). Simulating the role of visual selective attention during the development of perceptual completion. *Dev. Sci.* 15, 739–752. doi: 10.1111/j.1467-7687.2012.01177.x

Slater, A. (2002). Visual perception in the newborn infant: issues and debates. *Intellectica* 34, 57–76.

von Hofsten, C. (2010). Prospective control: a basic aspect of action development. *Hum. Dev.* 36, 253–270. doi: 10.1159/000278212

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: 04 February 2014; accepted: 25 April 2014; published online: 20 May 2014.
*Citation: Schlesinger M, Johnson SP and Amso D (2014) Prediction-learning in infants as a mechanism for gaze control during object exploration. Front. Psychol. **5**:441. doi: 10.3389/fpsyg.2014.00441*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Schlesinger, Johnson and Amso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 14 January 2015
doi: 10.3389/fpsyg.2014.01457

Binocular fusion and invariant category learning due to predictive remapping during scanning of a depthful scene with eye movements

Stephen Grossberg*, Karthik Srinivasan and Arash Yazdanbakhsh

Center for Adaptive Systems, Graduate Program in Cognitive and Neural Systems, Center of Excellence for Learning in Education, Science and Technology, Center for Computational Neuroscience and Neural Technology, and Department of Mathematics, Boston University, Boston, MA, USA

Edited by:
Chris Fields, Independent Scientist, USA

Reviewed by:
Greg Francis, Purdue University, USA
Christopher W. Tyler, Smith-Kettlewell Eye Research Institute, USA

****Correspondence:***
Stephen Grossberg, Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA
e-mail: steve@bu.edu

How does the brain maintain stable fusion of 3D scenes when the eyes move? Every eye movement causes each retinal position to process a different set of scenic features, and thus the brain needs to binocularly fuse new combinations of features at each position after an eye movement. Despite these breaks in retinotopic fusion due to each movement, previously fused representations of a scene in depth often appear stable. The 3D ARTSCAN neural model proposes how the brain does this by unifying concepts about how multiple cortical areas in the What and Where cortical streams interact to coordinate processes of 3D boundary and surface perception, spatial attention, invariant object category learning, predictive remapping, eye movement control, and learned coordinate transformations. The model explains data from single neuron and psychophysical studies of covert visual attention shifts prior to eye movements. The model further clarifies how perceptual, attentional, and cognitive interactions among multiple brain regions (LGN, V1, V2, V3A, V4, MT, MST, PPC, LIP, ITp, ITa, SC) may accomplish predictive remapping as part of the process whereby view-invariant object categories are learned. These results build upon earlier neural models of 3D vision and figure-ground separation and the learning of invariant object categories as the eyes freely scan a scene. A key process concerns how an object's surface representation generates a form-fitting distribution of spatial attention, or attentional shroud, in parietal cortex that helps maintain the stability of multiple perceptual and cognitive processes. Predictive eye movement signals maintain the stability of the shroud, as well as of binocularly fused perceptual boundaries and surface representations.

Keywords: depth perception, perceptual stability, predictive remapping, saccadic eye movements, object recognition, spatial attention, gain fields, category learning

1. INTRODUCTION

1.1. STABILITY OF 3D PERCEPTS ACROSS EYE MOVEMENTS

Our eyes continually move from place to place as they scan a scene to fixate different objects with their high resolution foveal representations. Despite the evanescent nature of each fixation, we perceive the world continuously in depth. Such percepts require explanation, if only because each eye movement causes the fovea to process a different set of scenic features, and thus there are breaks in retinotopic fusion due to each movement. Within a considerable range of distances and directions of movement, the fused scene appears stable in depth, despite the fact that new retinotopic matches occur after each movement. How does the brain convert such intermittent fusions into a stable 3D percept that persists across eye movements?

This article develops the 3D ARTSCAN model to explain and simulate how the brain does this, and makes several predictions to further test model properties. The model builds upon and integrates concepts and mechanisms from earlier models:

FACADE (Form-And-Color-And-DEpth) is a theory of 3D vision and figure-ground separation that proposes how 3D boundaries and surfaces are formed from 3D scenes and 2D pictures that may include partially occluding objects (Grossberg, 1994, 1997; Grossberg and McLoughlin, 1997; Grossberg and Kelly, 1999; Kelly and Grossberg, 2000; Grossberg et al., 2002, 2007, 2008; Grossberg and Swaminathan, 2004; Cao and Grossberg, 2005, 2012; Grossberg and Yazdanbakhsh, 2005; Fang and Grossberg, 2009). The articles that develop FACADE also summarize and simulate perceptual and neurobiological data supporting the model's prediction that 3D boundary and surface representations are, indeed, the perceptual units of 3D vision.

aFILM (Anchored Filling-In Lightness Model) simulates psychophysical data about how the brain generates representations of anchored lightness and color in response to psychophysical displays and natural scenes (Hong and Grossberg, 2004; Grossberg and Hong, 2006).

ARTSCAN (Grossberg, 2007, 2009; Fazl et al., 2009) models and simulates perceptual, attentional, and neurobiological data about how the brain can coordinate spatial and object attention across the Where and What cortical streams to learn and recognize view-invariant object category representations as it scans a

2D scene with eye movements. These category representations form in the inferotemporal cortex in response to 2D boundary and surface representations that form across several parts of the visual cortex. In order to learn view-invariant object categories, the model showed how spatial attention maintains its stability in head-centered coordinates during eye movements as a result of the action of eye-position-sensitive gain fields.

These earlier models did not, however, consider how 3D boundary and surface representations that are formed from binocularly fused information from the two eyes is maintained as the eyes move to fixate different sets of object features. The current article shows how the stability of 3D boundary and surface representations *and* of spatial attention are ensured using gain fields. With this new competence incorporated, the 3D ARTSCAN model can learn view-invariant object representations as the eyes scan a depthful scene.

3D ARTSCAN is also consistent with the pARTSCAN (positional ARTSCAN) model (Cao et al., 2011), which clarifies how an observer can learn both positionally-invariant and view-invariant object categories in a 2D scene; the dARTSCAN (distributed ARTSCAN) model (Foley et al., 2012), which clarifies how visual backgrounds do not become dark when spatial attention is focused on a particular object, how Where stream transient attentional components and What stream sustained attentional components interact, and how prefrontal priming interacts with parietal attention mechanisms to influence search efficiency; and the ARTSCAN Search model (Chang et al., 2014), which, in addition to supporting view- and positionally-invariant object category learning and recognition using Where-to-What stream interactions, can also search a scene for a valued goal object using reinforcement learning, cognitive-emotional interactions, and What-to-Where stream interactions. It thereby proposes a neurobiologically-grounded solution of the Where's Waldo problem. With the capacity of searching objects in depth added, which the results hereby about 3D perceptual stability permit, a 3D ARTSCAN Search model could learn and recognize both positionally-invariant and view-invariant object categories in a depthful scene, and use eye movements to search for a Where's Waldo target in such a scene, without disrupting perceptual stability during the search.

Section 1 summarizes conceptual issues and processes that are needed to understand and model the maintenance of 3D perceptual stability across saccadic eye movements. Section 2 heuristically reviews the ARTSCAN model upon which the 3D ARTSCAN model builds. Section 3 provides a heuristic description of 3D ARTSCAN concepts and mechanisms. Section 4 summarizes simulation results using the 3D ARTSCAN model that demonstrate 3D perceptual stability across saccadic eye movements. Section 5 summarizes the mathematical equations and parameters that define the 3D ARTSCAN model. Sections 3 and 5 are written with a parallel structure, and with cross-references to model equation numbers and model system diagrams, in order to facilitate model understanding. Section 6 provides a comparative discussion of key concepts and their relationships to other data and models. A reader can skip from Section 4 to 6 if the mathematical structure of the model is not of primary interest.

The main theoretical goal of the current article is to demonstrate the property of perceptual stability of 3D visual boundaries and surfaces across saccadic eye movements, which has been clarified using a variety of experimental paradigms (Irwin, 1991; Carlson-Radvansky, 1999; Cavanagh et al., 2001; Fecteau and Munoz, 2003; Henderson and Hollingworth, 2003; Beauvillain et al., 2005). The article also predicts how this process interacts with processes of spatial and object attention, invariant object category learning, predictive remapping, and eye movement control, notably how they all regulate and/or respond to adaptive coordinate transformations. As explained more fully below, the brain can prevent a break in binocular fusion after an eye movement occurs by using predictive gain fields to maintain 3D boundary and surface representations in head-centered coordinates, even though these representations are not maintained in retinotopic coordinates. This property is demonstrated by simulations using 2D geometrical shapes and natural objects that are viewed in 3D. In particular, the simulations show that the 3D boundary and surface representations of these objects are maintained in head-centered coordinates as the eyes move.

These simulation results generalize immediately to 3D objects that have multiple 2D planar surfaces, since the simulations due not depend upon a particular binocular disparity. Indeed, other modeling studies have demonstrated how the same retinotopic binocular mechanisms can process object features at multiple disparities (Grossberg and McLoughlin, 1997; Grossberg and Howe, 2003; Cao and Grossberg, 2005, 2012), including objects perceived from viewing stereograms (Fang and Grossberg, 2009) and natural 3D scenes (Cao and Grossberg, submitted), as well as objects that are slanted in depth (Grossberg and Swaminathan, 2004). All these results should be preserved under the action of predictive gain fields to convert their retinotopic boundary and surface representations into head-centered ones, since the gain fields merely predictively shift the representations that are created by the retinotopic mechanisms. The key point is thus that the gain field mechanism does not disrupt the retinotopically computed 3D boundary and surface representations. It just changes their coordinates from retinotopic to head-centered to create invariance under eye movements.

The current model computes target positions to which the eyes are commanded to move, but does not model the neural machinery that is needed to accomplish the yoked saccadic movements themselves. Earlier models of the saccadic and smooth pursuit eye movement brain systems that are commanded by such positional representations can be used to augment the current model in future studies (e.g., Grossberg and Kuperstein, 1986; Grossberg et al., 1997, 2012; Gancarz and Grossberg, 1998, 1999; Srihasam et al., 2009; Silver et al., 2011).

1.2. PREDICTIVE REMAPPING AND GAIN FIELDS: MAINTAINING FUSION ACROSS SACCADES

The brain compensates for the changes in retinal coordinates of fused object features fast enough to prevent fusion from being broken. This compensatory property is called *predictive remapping*. Predictive remapping has been used to interpret neurophysiological data about the updating of the representation of visual space by intended eye movements, particularly in cortical areas

such as the parietal cortex, prestriate cortical area V4, and frontal eye fields (Duhamel et al., 1992; Umeno and Goldberg, 1997; Gottlieb et al., 1998; Tolias et al., 2001; Sommer and Wurtz, 2006; Melcher, 2007, 2008, 2009; Saygin and Sereno, 2008; Mathot and Theeuwes, 2010a). Predictive remapping is often explained as being achieved by *gain fields* (Andersen and Mountcastle, 1983; Andersen et al., 1985; Grossberg and Kuperstein, 1986; Gancarz and Grossberg, 1999; Deneve and Pouget, 2003; Pouget et al., 2003), which enable featural representations to incorporate information about the current or predicted gaze position. Gain fields are populations of cells that enable movement-sensitive transformations to occur between one coordinate frame (say, retinotopic), whose representations change due to eye movements, and another (say, head-centered), whose representations are invariant under eye movements.

In both the ARTSCAN model and the 3D ARTSCAN model, gain fields are proposed to be updated by corollary discharges of outflow movement signals that act before the eyes stabilize on their next movement target. In the ARTSCAN model, these predictive gain field signals maintain the stability of spatial attention to an object as eye movements scan the object; see Section 2. In the 3D ARTSCAN model, gain field signals also prevent binocularly-fused object boundary and surface representations of the object from being reset by such eye movements. The 3D ARTSCAN model hereby proposes how the process of predictive remapping of 3D boundary and surface representations is linked to the processes of figure-ground separation of multiple objects in a scene, and of learning to categorize and attentively recognize these objects during active scanning of the scene with saccadic eye movements. The following sections summarize how these processes are predicted to be coordinated.

2. REVIEW OF ARTSCAN MODEL

2.1. SOLVING THE VIEW-TO-OBJECT BINDING PROBLEM WHILE SCANNING A SCENE

The ARTSCAN model and its variants propose answers to the following basic questions: What is an object? How does the brain learn what an object is under both unsupervised and supervised learning conditions? ARTSCAN predicts how spatial and object attention are coordinated to achieve rapid object learning and recognition during eye movement search. In particular, ARTSCAN proposes how the brain learns to recognize an object when it is seen from multiple views, or perspectives. How does such view-invariant object category learning occur?

As the eyes scan a scene, two successive eye movements may focus on different parts of the same object or on different objects. ARTSCAN proposes how the brain avoids erroneously classifying views of different objects together, even before the brain knows what the object is. ARTSCAN also proposes how the brain controls eye movements that enable it to learn multiple view-specific categories and to associately link them with view-invariant object category representations.

The ARTSCAN model (**Figure 1**) predicts how spatial attention may play a crucial role in controlling view-invariant object category learning, using attentionally-regulated signals from the Where cortical stream to the What cortical stream to modulate category learning. Several studies have reported that the distribution of spatial attention can configure itself to fit an object's form. Form-fitting spatial attention is sometimes called an *attentional shroud* (Tyler and Kontsevich, 1995). ARTSCAN explained how an object's pre-attentively formed surface representation in prestriate cortical area V4 may induce such a form-fitting attentional shroud in parietal cortex. In particular, feedback between the surface representation and the shroud are predicted to form a *surface-shroud resonance* that locks spatial attention on the object's surface. While this surface-shroud resonance remains active, it is predicted to accomplish the following: First, it ensures that eye movements tend to end at locations on the object's surface, thereby enabling different views of the same object to be sequentially explored (Theeuwes et al., 2010). Second, it keeps the emerging view-invariant object category active while different views of the object are learned by view-specific categories and associated with it.

The ARTSCAN model thus addressed what would otherwise appear to be an intractable infinite regress: If the brain does not already know what the object is, then how can it, without external guidance, prevent views from several objects from being associated and thus distort the learning of object categories? How does such unsupervised learning until naturalistic viewing conditions get started? The ARTSCAN model shows that an object's pre-attentively and automatically formed surface representation (**Figure 1**) provides the object-sensitive substrate that enables view-invariant object category learning to occur, and thereby circumvents this infinite regress.

The fact that a surface representation can form pre-attentively is consistent with the burgeoning psychophysical literature showing that 3D boundaries and surfaces are the units of pre-attentive visual perception (Grossberg and Mingolla, 1987; Grossberg, 1987a,b, 1994; Paradiso and Nakayama, 1991; Elder and Zucker, 1993; He and Nakayama, 1995; Rogers-Ramachandran and Ramachandran, 1998; Raizada and Grossberg, 2003) and that attention selects these units for recognition (Kahneman and Henik, 1981; He and Nakayama, 1995; LaBerge, 1995).

The ARTSCAN model used the simplest possible front end from the FACADE model of 3D vision and figure-ground perception (Grossberg, 1994, 1997; Grossberg and McLoughlin, 1997) in order to process letters of variable sizes and fonts in simple 2D images. The 3D ARTSCAN Search model elaborates this front end to enable binocular fusion of objects in a 3D scene (see **Figures 2–4** and Section 3 for details).

2.2. ATTENTIONAL SHROUD INHIBITS RESET OF AN INVARIANT OBJECT CATEGORY DURING OBJECT LEARNING

ARTSCAN processes can be described as a temporally coordinated interaction between multiple brain regions within and between the What and Where cortical processing streams, including the Lateral Geniculate Nucleus (LGN), cortical areas V1, V2, V3A, V4, MT, MST, PPC, LIP, ITp, and ITa, and the superior colliculus (SC): The Where stream maintains an attentional shroud whose spatial coordinates mark the surface locations of a current "object of interest," whose identity has yet to be

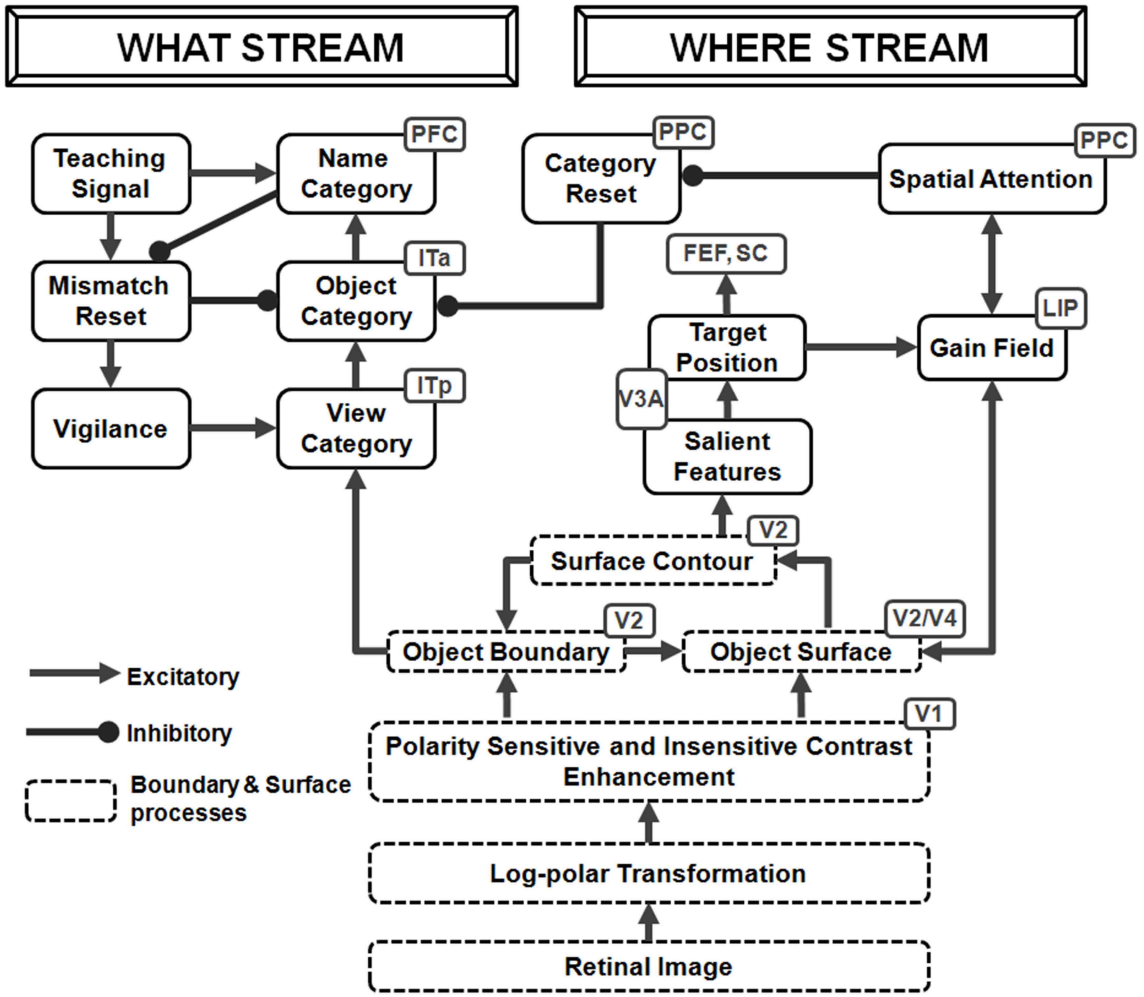


FIGURE 1 | Model diagram of the ARTSCAN model (reprinted with permission from Chang et al., 2014). A few simplified stages from the FACADE model (Grossberg and Todorović, 1988; Grossberg, 1994, 1997; Grossberg and McLoughlin, 1997) preprocess 2D images. The 3D ARTSCAN model is a synthesis and further development of the ARTSCAN model, the aFILM model of anchored lightness and color perception (Hong and Grossberg, 2004; Grossberg and Hong, 2006), and the FACADE model to enable 3D surface percepts to remain stable as saccadic eye movements scan a scene (as elaborated in **Figures 2–5**).

determined in the What stream. As each view-specific category is learned by the What stream, say in posterior inferotemporal cortex (ITp), it focuses object attention via a learned top-down expectation on the critical features in the visual cortex (e.g., in prestriate cortical area V4) that will be used to recognize that view and its variations in the future. When the first such view-specific category is learned, it also activates a cell population at a higher cortical level, say anterior inferotemporal cortex (ITa), that will become the view-invariant object category.

Suppose that the eyes or the object move sufficiently to expose a new view whose critical features are significantly different from the critical features that are used to recognize the first view. Then the first view category is reset, or inhibited. This happens due to the mismatch of its learned top-down expectation, or prototype of attended critical features, with the newly incoming view information. This top-down prototype focuses object attention on the incoming visual information. Object attention hereby helps to control which view-specific categories are learned by determining when the currently active view-specific category should be reset, and a new view-specific category should be activated.

However, the view-invariant object category should *not* be reset every time a view-specific category is reset, or else it can never become view-invariant. This is what the attentional shroud accomplishes: It inhibits a tonically-active reset signal that would otherwise shut off the view-invariant category when each view-based category is reset. As the eyes foveate a sequence of views on a single object's surface through time, they trigger learning of a sequence of view-specific categories, and each of them is associatively linked through learning with the still-active view-invariant category.

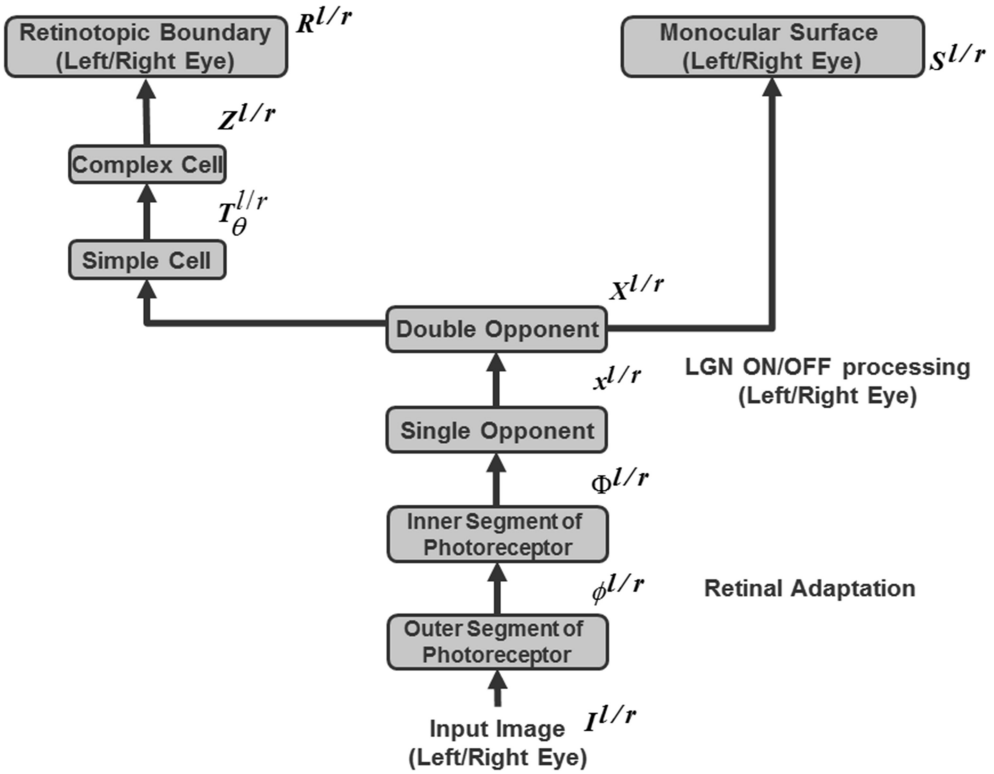


FIGURE 2 | Retinal adaptation of input scene followed by pre-attentive boundary and surface processing in the 3D ARTSCAN model. Light adaptation at the model's outer segment of the photoreceptors and spatial contrast adaptation at the inner segments of photoreceptors are implemented as in the aFILM model (Grossberg and Hong, 2006) (Equations 1–8). The outputs from the inner segment of the photoreceptors input to the model LGN. These inputs are contrast-normalized by single opponent networks of ON and OFF cells via on-center off-surround and off-center on-surround interactions, respectively, among cells that obey membrane equation, or shunting, dynamics (Equations 9–14) and then by double-opponent networks (Equations 15, 16). LGN double-opponent outputs are used to compute orientationally- and contrast-selective simple cells that are selective to four different orientations (Equations 17–20). Simple cell outputs are pooled across all four orientations to yield complex cells (Equation 21).Complex cells, in turn, input to monocular left (L) and right (R) eye retinotopic boundaries.

When the eyes move off an object, its attentional shroud collapses in the Where stream, thereby transiently disinhibiting the reset mechanism that shuts off the view-invariant category in the What stream. When the eyes look at a different object, its shroud can form in the Where stream and a new view-specific category can be learned that can, in turn, activate the cells that will become a new view-invariant category in the What stream. Chiu and Yantis (2009) have described rapid event-related fMRI experiments in humans showing that a spatial attention shift causes a domain-independent transient parietal burst that correlates with a change of categorization rules. This transient parietal signal is a marker against which further experimental tests of model mechanisms can be based; e.g., a test of the predicted sequence of V4-parietal surface-shroud collapse (shift of spatial attention), transient parietal burst (reset signal), and collapse of currently active invariant object category in cortical area ITa (shift of categorization rules). These and related results (e.g., Corbetta et al., 2000; Yantis et al., 2002; Cabeza et al., 2008) are consistent with the model prediction of how different regions of the parietal cortex maintain sustained attention to a currently attended object (shroud) and control transient attention switching (reset burst) to a different object.

2.3. BOUNDARY AND SURFACE REPRESENTATIONS FORM PRE-ATTENTIVELY

Convergent psychophysical and neurobiological data (e.g., He and Nakayama, 1992; Elder and Zucker, 1998; Rogers-Ramachandran and Ramachandran, 1998; Lamme et al., 1999) support the 1984 prediction of Grossberg and colleagues that the units of pre-attentive visual perception are boundaries and surfaces (Cohen and Grossberg, 1984; Grossberg, 1984; Grossberg and Mingolla, 1985a,b; Grossberg and Todorović, 1988). The model that embodies this prediction is often called the BCS/FCS model, for Boundary Contour System and Feature Contour System. This hypothesis was generalized by Grossberg in 1987 to the prediction that 3D boundaries and surfaces are the units of 3D vision and figure-ground perception. This prediction is part of the FACADE (Form-And-Color-And-DEpth) theory of 3D vision and figure-ground separation, which has been used to explain and predict a wide range of perceptual and neurobiological data; see Grossberg (1994, 2003) and Raizada and Grossberg (2003) for reviews. Perceptual boundaries are predicted to form in the (LGN Parvo)-(V1 Interblob)-(V2 Interstripe)-V4 cortical stream, while perceptual surfaces are predicted to form in the (LGN Parvo)-(V1 Blob)-(V2 Thin Stripe)-V4 stream. Various psychophysical

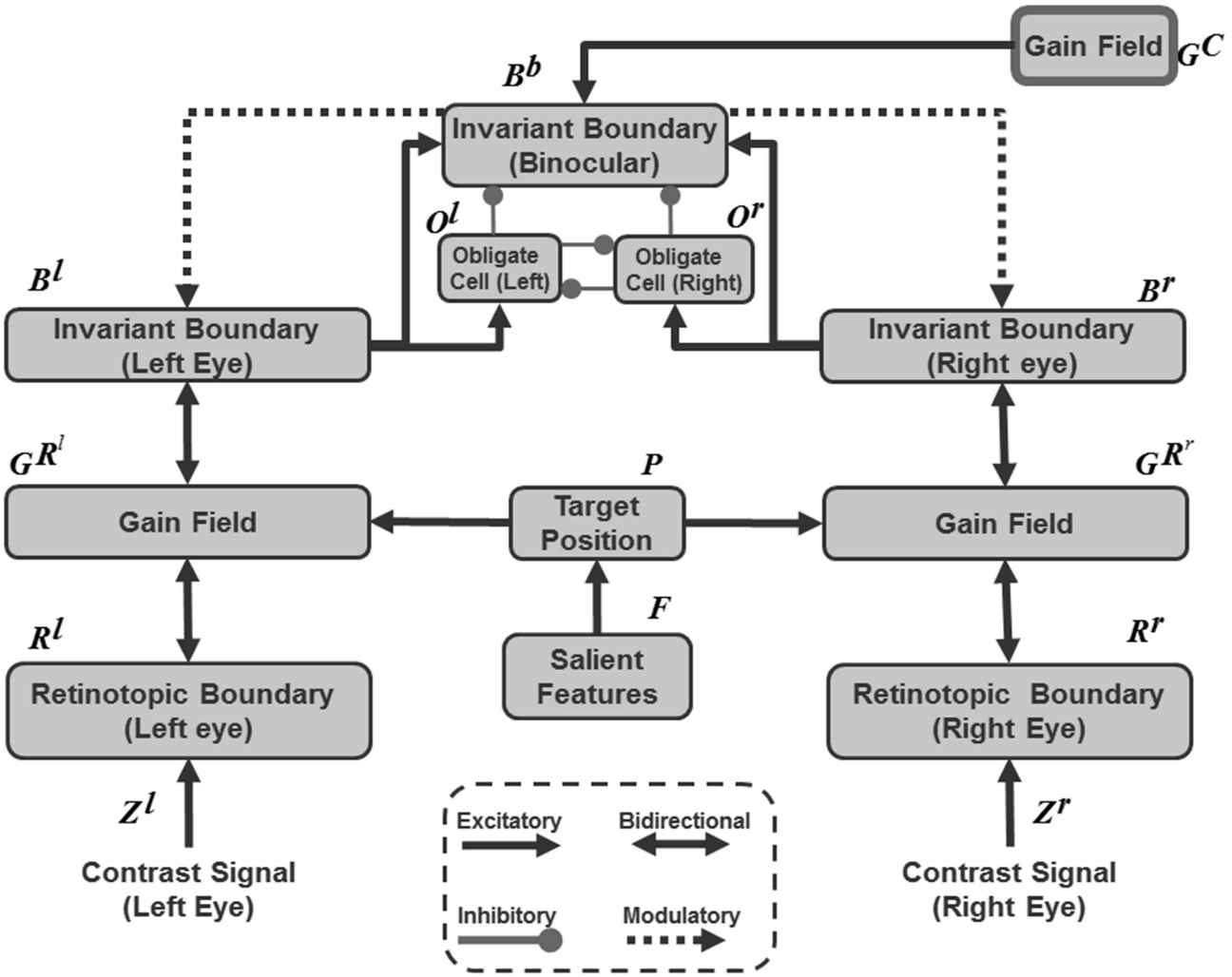


FIGURE 3 | 3D ARTSCAN model macrocircuit for maintaining the stability of fused binocular boundaries during eye movements. Retinotopic monocular boundaries (Equation 22) are computed from complex cell inputs (Equation 21). These boundaries are reset whenever the eyes move. The retinotopic monocular boundaries input to invariant monocular boundaries via gain fields. The invariant boundaries are not reset by eye movements because they are predictively remapped by eye position-selective gain fields before the eyes move to a new fixation position. The invariant monocular boundaries, in turn, feed back to modulate the retinotopic monocular boundaries. The gain fields receive their inputs from target positions that are computed from salient features on surface contours (see Sections 3.4, 3.6, and Equations 45, 64–66). The invariant monocular boundaries (Equation 26) are binocularly fused to form the invariant binocular boundaries (Equation 33). Both excitatory and inhibitory (obligate) inputs to the invariant binocular boundaries are needed to ensure their disparity selectivity. The maintained fusion of binocular boundaries is a primary goal of predictive remapping, since these boundaries support the persistence of object percepts during saccadic eye movements. These fused binocular boundaries modulate the activities of the invariant monocular boundaries and thus the activity of the retinotopic boundary layer via top-down feedback. This top-down feedback ensures that any changes or collapse in the invariant boundary activity is propagated all the way back to the retinotopic boundaries (see Section 3.3 and Equations 22–35).

(Rubin, 1921; Beardslee and Wertheimer, 1958; Driver and Baylis, 1996), fMRI (Kourtzi and Kanwisher, 2001), and electrophysiological data (Baylis and Driver, 2001) support the hypothesis that boundaries and surfaces can form pre-attentively as they help to separate figures from their backgrounds in depth. These experiments show that whether an edge is assigned to a figure or to a background serves as an important factor for attracting attention, activating object recognition areas, and remembering it later. It has also been argued that, prior to attentive selection of an object, figure-ground segregation occurs (Baylis and Driver, 2001), and that it is yoked to bottom-up processes that do not need a top-down attentive influence to be initiated. The boundaries and surfaces that are implemented in the 3D ARTSCAN Search model are generalized in two ways beyond their implementation in the ARTSCAN model:

2.3.1. 3D boundaries and surfaces

As noted above, the monocular boundaries and surfaces in the ARTSCAN model are generalized using FACADE theory mechanisms to form disparity-selective boundaries and surfaces that

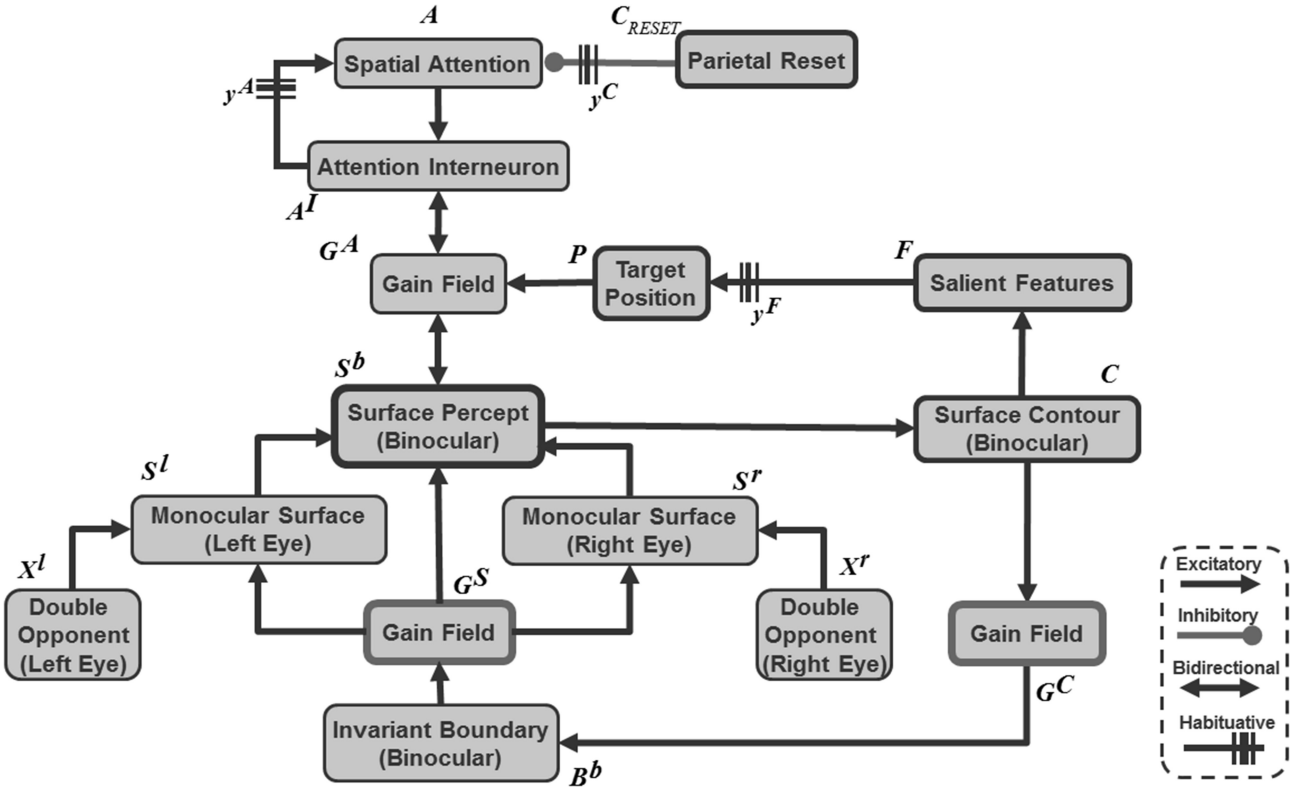


FIGURE 4 | Maintenance and perceptual stability of fused binocular surfaces during eye movements. The monocular surfaces $S^{l/r}$ fill-in (Equation 36) (**Figure 2**) brightness signals from the double-opponent ON and OFF cells (Equations 15, 16). The diffusion that governs filling-in is gated by invariant binocular boundaries (Equation 33) after they are converted into retinotopic binocular boundaries (Equation 40) via gain fields (Equations 42–44). The monocular surfaces are fused to form a binocular surface S^b (Equation 39). The rectified sum of the ON and OFF filling-in domains is the final binocular surface percept (Equation 41) and is assumed to be the consciously seen retinotopic surface percept in depth. Gain fields operating at different levels guarantee the stability of the binocular percept (Section 3.4 and Equations 36–50). Binocular surface representations give rise to surface contours C (Equation 45) from which the most salient feature positions F (Equation 64) are chosen as the next target positions P (Equation 66) for eye movements. Corollary discharges from the target positions are used to predictively remap key boundary and surface representations via gain fields (Section 3.6). In particular, a retinotopic binocular surface percept is remapped via gain fields (Equation 56) into attentional interneurons (Equation 55) that input to the spatial attention map at which a head-centered attentional shroud is chosen. The attentional shroud (Equation 51) habituates at an activity-dependent rate (Equation 61) and is inhibited by a burst of the parietal reset signal (Equation 62) that is rendered transient by its own habituative transmitter gate (Equation 63). This enables a shift in attention to occur to a different surface (see Sections 3.4–3.6 and Equations 36–66).

can represent an object in depth. In this generalization, processing stages for retinal adaptation as well as opponent and double-opponent processing in ON and OFF cells (Grossberg and Hong, 2006) feed into monocular and binocular laminar cortical boundary representations (Cao and Grossberg, 2005); see Sections 3 and 5 for details.

The surface representations that compete for spatial attention in shroud formation are called Filling-In Domains, or FIDOs (Grossberg, 1994). FACADE theory predicts that each of the depth-selective boundary representations that capture surface lightness and color at prescribed depths interacts with a complete set of opponent filling-in domains (light vs. dark, red vs. green, blue vs. yellow) that compete at each position. In addition, each FIDO's activity pattern is processed by an on-center off-surround shunting network that contrast-normalizes its input patterns (Grossberg, 1973, 1980). These two types of competition (opponent and spatial), acting together, define a double-opponent field of cells. There are multiple FIDOs, each sensitive to a different range of depths. These double-opponent FIDOs can represent conjunctions of depth and color across space. A unique conjunction of depth and color may pop out during visual search (Nakayama and Silverman, 1986) because it is the only active region on the FIDO corresponding to that depth and color. FACADE theory models its highest level of surface filling-in in cortical area V4, where visible surfaces are represented and 3D figure-ground separation is completed (e.g., Schiller and Lee, 1991).

These depth-selective double-opponent surface representations in V4 provide the computational substrates that compete for spatial attention in the model's parietal cortex. The reciprocal shroud-to-surface feedback may also be expected to be selective to conjunctions of depth and color. Such a mechanism may clarify various color-specific search data; e.g., Egeth et al. (1984) and

Wolfe et al. (1994) wherein human subjects may break up a conjunctive search task into a color priming operation followed by depth-selective pop-out.

The 3D ARTSCAN Search model simulates a single depth-selective double-opponent FIDO, for simplicity.

2.3.2. Predictive remapping maintains binocular fusion and shroud stability

In ARTSCAN, predictive remapping is used to maintain the stability of an attentional shroud as eye movements explore an attended object. This stability is needed to prevent the shroud from collapsing and disinhibiting the reset mechanism in response to every sufficiently large saccade that explores the object. In the current 3D ARTSCAN model, predictive remapping also has another role: it maintains binocular fusion of previously fused features as the eyes move within a certain spatial range to foveate a different set of features on the object. Thus, predictive remapping mechanisms that were previously predicted to operate in areas such as parietal cortex are here also suggested to operate as early as visual cortical area V1; see Sections 3.4, 3.5, and 5 for details.

The following sections summarize how the two types of predictive remapping are proposed to be related.

2.4. SURFACE CONTOUR SIGNALS INITIATE FIGURE-GROUND SEPARATION

Shroud stability is achieved in ARTSCAN using feedback signals between surfaces and boundaries in the following way: 3D boundary signals are topographically projected from where they are formed in the V2 interstripes to the surface representations in the V2 thin stripes (**Figure 1**). These boundaries act both as *filling-in generators* that initiate the filling-in of surface lightness and color when the corresponding boundary and surface signals are aligned, and as *filling-in barriers* that prevent the filling-in of lightness and color from crossing object boundaries (Grossberg, 1994). If the boundary is closed, it can contain, or *gate*, the filling-in of an object's lightness and color within it. If, however, the boundary has a sufficiently big gap in it, then surface lightness and color can spread through the gap and surround the boundary on both sides, thereby equalizing the contrasts on both sides of the boundary.

Feedback from surfaces in V2 thin stripes to boundaries in V2 interstripes is achieved by *surface contour* signals. Surface contour signals are generated by contrast-sensitive on-center off-surround networks that generate contour-sensitive output signals from the activities across each FIDO after surface filling-in occurs. The inhibitory connections in the network's off-surround act across position and within depth. As a result, each FIDO generates output signals via its own contrast-sensitive on-center off-surround network. Surface contour signals are the output signals that are generated by contrast changes across each FIDO.

Such contrast changes typically occur if the filled-in surface is surrounded by gating signals from a closed boundary, because a closed boundary can contain a FIDO's filling-in process. In particular, gating at closed boundary positions generates contrasts of filled-in lightnesses or colors at these positions by blocking the spread of lightnesses or colors across these positions. As a result, surface contour signals can be generated at the positions where the gating signals of closed boundaries occur. The positions at which surface contour signals in the surface stream are generated are thus a subset of the same positions as those of the corresponding boundaries in the boundary stream. These boundary and surface contour positions typically include positions where there are salient features on an object's surface.

Surface contour signals are not, however, generated at boundary positions near a big gap, or hole, in an object boundary, since filled-in lightnesses and colors can flow out of, and around, such a boundary break to cause approximately equal filled-in activities on both sides of the boundary. Since there is then zero contrast of filled-in activity across such a boundary, the contrast-sensitive on-center off-surround network does not generate an output signal at these positions, and hence no surface contour forms there.

The boundary positions that limit the filling-in process within the surface stream are thus a superset of the positions in the surface stream at which surface contours form after filling-in. As a result, surface contour output signals back to the boundary stream are received at a subset of boundary positions. In particular, gating signals that are generated by closed boundaries block the flow of filled-in brightness and/or color signals outside the regions that they surround. Closed boundaries hereby mark the positions where a contrast different across space in the filled-in brightness and/or color can occur. They are therefore also positions where surface contour feedback signals can arise.

The surface contour feedback signals from the surface stream to the boundary stream are delivered via an on-center off-surround network that acts within position and across depth. The on-center signals strengthen the closed boundaries that generated the successfully filled-in surfaces, whereas the off-surround signals inhibit spurious boundaries at the same positions but farther depths. Surface contour signals hereby strengthen the boundaries that lead to successfully filled-in surfaces, while inhibiting those that do not. By eliminating spurious boundaries, the off-surround signals initiate figure-ground separation by enabling occluding and partially occluded surfaces to be separated onto different depth planes, and partially occluded boundaries and surfaces to be amodally completed behind their occluders. See Grossberg (1994), Kelly and Grossberg (2000), and Fang and Grossberg (2009) for further discussion of figure-ground percepts and computer simulations of them.

2.5. ATTENDED SURFACE CONTOUR SIGNALS CREATE ATTENTION POINTERS TO SALIENT EYE MOVEMENT TARGET POSITIONS

Figure-ground separation needs to occur at an earlier processing stage than the learning of view-specific and view-invariant categories of an object, since if different objects were not pre-attentively separated from each other, the brain would have no basis for segregating the learning of views that belong to one object. Once figure-ground separation is initiated, ARTSCAN predicts how surface contour signals can be used to determine a sequence of eye movement target positions to salient features on an attended object surface, and thus to enable multiple view-specific categories of the object to be learned and associated with an emerging view-invariant object category.

This works as follows: the pre-attentive bottom-up inputs from the retina and LGN activate multiple surface representations in cortical area V4. These surfaces, in turn, attempt to topographically activate spatial attention to form a surface-fitting attentional shroud in parietal cortex. As they do so, they generate top-down excitatory topographic feedback to visual cortex and long-range inhibitory interactions in parietal cortex. Taken together, these interactions define a *recurrent* on-center off-surround network that is capable of contrast-enhancing the strongest shroud and inhibiting weaker ones. Positive feedback from a winning shroud in parietal cortex to its surface in V4 is thus predicted to increase the contrast gain of the attended surface, as has been reported in both psychophysical experiments (Carrasco et al., 2000) and neurophysiological recordings from cortical areas V4 (Reynolds et al., 1999, 2000; Reynolds and Desimone, 2003), possibly carried by the known connections from parietal areas to V4 (Cavada and Goldman-Rakic, 1989, 1991; Distler et al., 1993; Webster et al., 1994).

How do salient features on an attended surface attract eye movements? If figure-ground separation begins in cortical area V2, with surface contours as one triggering mechanism, then these eye movement commands need to be generated no earlier than V2. The surface contour signals themselves are plausible candidates from which to derive eye movement target commands because, being generated by a contrast-sensitive on-center off-surround network, they are stronger at contour discontinuities and other distinctive contour features that are typical end points of saccadic movements. When the contrast of an attended surface increases, the strength of its surface contour signals also increases (**Figure 1**). Corollary discharges of these surface contour signals are predicted to be computed within a parallel pathway that is mediated via cortical area V3A (Nakamura and Colby, 2000; Caplovitz and Tse, 2007), which occurs after V2, and to generate saccadic commands that are restricted to salient features of the attended surface (Theeuwes et al., 2010) until the shroud collapses and spatial attention shifts to enshroud another object. Consistent with this prediction, it is known that "neurons within V3A· · · process continuously moving contour curvature as a trackable feature…not to solve the "ventral problem" of determining object shape but in order to solve the "dorsal problem" of what is going where" (Caplovitz and Tse, 2007, p. 1179).

In particular, ARTSCAN proposed how surface contour signals within the corollary discharge pathway are contrast-enhanced to select the largest signal as the next position upon which spatial attention will focus and the next saccadic eye movement will move (**Figure 1**). These positions have properties of the "attention pointers" reported by Cavanagh et al. (2010).

2.6. PREDICTIVE SURFACE CONTOUR SIGNALS CONTROL GAIN FIELDS THAT MAINTAIN SHROUD STABILITY

Each eye movement target signal that is derived from a surface contour generates a gain field that maintains a stable shroud in head-centered coordinates as the eyes move (**Figure 5**). These outflow movement commands thus control predictive remapping that maintains attentional stability through time. The stable shroud, in turn, can maintain persistent inhibition of the category reset mechanism as the eyes explore the object and the brain learns multiple view-specific categories of it (**Figure 1**).

3. 3D ARTSCAN MODEL

The 3D ARTSCAN model unifies properties of the ARTSCAN, 3D LAMINART, and aFILM models in a way that is compatible with the pARTSCAN and ARTSCAN Search models. The model does not include the log-polar transformation of cortical magnification, however. This simplification reduces the computational burden in its simulations due to the need to transform binocular inputs into 3D boundary and surface representations that are preserved during eye movements.

3.1. RETINAL ADAPTATION

Two stages of retinal adaptation (**Figure 2**; Section 5.1 Equations 1–8) are implemented from the aFILM model of Grossberg and Hong (2006): light adaptation at the outer segment of the photoreceptors and spatial contrast adaptation at the inner segments of photoreceptors. In the outer segment of the photoreceptors, intracellular gating mechanisms such as calcium negative feedback occur (Koutalos and Yau, 1996). This process facilitates light adaptation *in vivo*, by shifting the operating range of the photoreceptor to adapt to the ambient luminance of the visual field. Spatial contrast adaptation at the inner segments of photoreceptors occurs through light adapted inputs from the outer segment, with negative feedback from the horizontal cells (HC) that modulate the influx of calcium ions and control the amount of glutamate release from the photoreceptor terminals (Fahrenfort et al., 1999). The HC network computes spatial contrast using gap junction connections (syncytium) between the HCs. The permeability of the gap junctions between HCs decreases as the difference of the inputs to the coupled photoreceptors increases, and the HCs in the light and dark image regions deliver different suppressive feedback signals to the inner segments of the photoreceptors to properly rescale the inputs that have too much contrast. For simplicity, only gap junction connections between nearest neighbor cells are considered.

During active scanning of natural images with eye movements, the scanned image intensities can vary over several orders of magnitude (Rieke and Rudd, 2009). The model retina uses these two different mechanisms to map widely different input intensities to sensitive, and therefore discriminable, portions of the response range.

3.2. LGN POLARITY-SENSITIVE ON AND OFF CELLS

The LGN ON and OFF cells normalize the adapted contrast and brightness information of the input pattern from the retina using on-center off-surround shunting networks which are solved at equilibrium for computational speed (**Figure 2** and Equations 9–12). LGN ON cells respond to image increments (Equation 13) whereas OFF cells respond to image decrements (Equation 14). These single-opponent cells generate output signals that compete at each position, thereby giving rise to double-opponent ON and OFF cells (Equations 15, 16).

3.3. BOUNDARY PROCESSING

The output signals of the double-opponent ON/OFF LGN cells are the inputs to simple cells that respond selectively to one

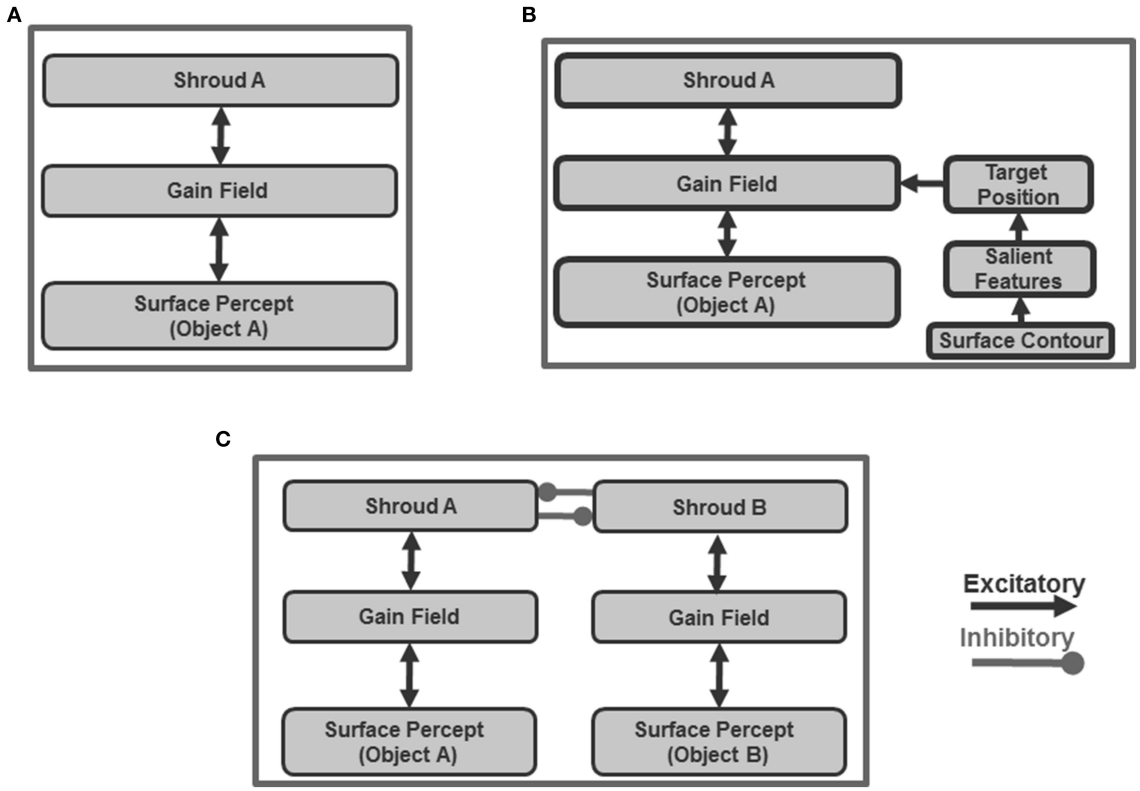


FIGURE 5 | Schematic for surface-shroud resonance through a feedback interaction between a retinotopic binocular surface and a head-centered spatial attentional shroud. (A) In the absence of any eye movement to a new target position, the gain fields maintain the stable object shroud of a given object surface. **(B)** When a surface contour is contrast-enhanced to localize salient features (Equation 45), and the position of the most salient feature is chosen as the next target position signal (Equation 67), the gain field is predictively remapped by the target position corollary discharge signal before the corresponding saccadic eye movement occurs (Equation 56), with the result that the shroud retains its stability across eye movements. While the shroud remains active and spatial attention remains focused on a single object surface, the eyes can explore different views of the object, and the What stream of ARTSCAN can learn multiple view-selective object categories and associatively link them to an emerging view-invariant object category. **(C)** If the currently attended shroud collapses, competition across the spatial attention layer (Equation 51)nables another shroud to win the competition and to focus object attention upon the corresponding object surface.

of four orientations (Equation 17). Simple cell output signals are pooled over all orientations and opposite contrast polarities to create polarity-insensitive complex cell boundaries (**Figure 2** and Equation 21). The simplification of pooling over orientation was done because the model is not used to simulate any polarity-specific interactions.

Both monocular and binocular boundaries are needed to generate depthful representations of object boundaries during biological vision (Nakayama and Shimojo, 1990; McKee et al., 1994; Smallman and McKee, 1995; Cao and Grossberg, 2005, 2012). The retinotopic monocular boundaries (**Figure 3** and Equation 22) are computed using bottom-up inputs from complex cells (Equation 21). Because they are computed in retinal coordinates, these boundaries are reset whenever the eyes move to fixate a different scenic position. The retinotopic monocular boundaries are also modulated by top-down signals from invariant monocular boundaries (Equation 26) that are not reset by an eye movement. This modulation facilitates predictive remapping. Invariance is achieved using a gain field (Equations 28–32); see **Figure 3**.

The invariant monocular boundaries (Equation 26) are derived from the retinotopic monocular boundaries (Equation 22), but are computed in head-centered coordinates that are invariant under eye movements. Before the eyes move, the invariant boundaries represent the same positions as the retinotopic boundaries (Equations 24, 25). The invariant monocular boundaries of a stationary object are, however, not reset when the eyes move. They derive their stability due to updated gain field signals that are derived from the next eye movement command even before the eyes actually move to the commanded position. Such predictive remapping of the invariant monocular boundaries to continuously represent the monocular boundaries in head-centered coordinates enables them to be maintained even while the retinotopic boundaries are reset.

The eye movement command is computed from surface contour signals (Sections 3.4–3.6) that are derived from the attended object surface (**Figures 1**, **4**) and that strengthen the boundaries that formed them. Moreover, when the contrast of a surface is increased by feedback from an attentional shroud, the surface contour signals increase, so the strength of the boundaries around the attended surface increase also.

Surface contour signals also activate a parallel, corollary discharge, pathway that projects to the salient features processing

stage (**Figure 4**). In order to compute the position of the next eye movement, these salient features signals are contrast-enhanced by an on-center off-surround network until the most active position is chosen as the next target position. The salient features of an attended surface have an advantage in this competition because they are amplified by shroud-to-surface-to-surface contour feedback.

This target position signal is used both to determine the target position of the next eye movement and to update gain fields that predictively remap retinotopic left and right monocular boundaries into invariant left and right monocular boundaries that remain continuously computed even during eye movements (**Figure 3**).

The invariant monocular boundaries (**Figure 3** and Equation 26) for a given object are fused to yield invariant binocular boundaries (**Figure 3** and Equation 33). Because of their computation from invariant monocular boundaries, the invariant binocular boundaries are also maintained as the eyes move. This maintained fusion is a main functional goal of the predictive remapping, since it enables the object percept to persist during eye movements. The fused binocular boundaries, in turn, modulate the activities of the invariant monocular boundaries and thus the activity of cells in the retinotopic boundary layer via top-down feedback through the gain field (**Figure 3**). This top-down modulatory feedback from the invariant binocular boundary to the invariant monocular boundary ensures that any change or collapse in the invariant binocular boundary activity is propagated back to the retinotopic boundaries (**Figure 3**).

In the brain, binocular fusion of monocular left and right boundaries tends to occur only between edges with the same contrast polarity (*same-sign hypothesis*; Howard and Rogers, 1995; Howe and Watanabe, 2003) and approximately the same magnitude of contrast (McKee et al., 1994). This constraint naturally arises when the brain fuses edges that derive from the same object in the world, and helps the brain to solve the classical *correspondence problem* (Julesz, 1971; Howard and Rogers, 1995). The model satisfies this constraint through interactions between excitatory and inhibitory cells (Equation 33) that are proposed to occur in layer 3B of cortical area V1 (Grossberg and Howe, 2003; Cao and Grossberg, 2005, 2012). These interactions endow the binocular cells with an *obligate property* (Poggio, 1991) whereby they respond preferentially to left and right eye inputs of approximately equal contrast (Equations 34, 35).

The original ARTSCAN model used gain fields only to predictively update the head-centered representations of attentional shrouds. The current model uses gain fields at several processing stages (**Figures 3**, **4**). They ensure that stable fusion of 3D binocular boundaries and surfaces is maintained in head-centered coordinates as the eyes move. The weights between the gain field neurons and the invariant boundary neurons are presumably learned. For simplicity, only the end product of the learning process, as suggested by Pouget and Snyder (2000), was used in the 3D ARTSCAN model.

3.4. SURFACE PROCESSING

The invariant binocular boundaries help to maintain the surface representations of stationary objects during eye movements. This is proposed to occur as follows:

Bottom-up inputs from double-opponent ON and OFF cells (**Figure 2** and Equations 15, 16) trigger monocular surface filling-in via a diffusion process (**Figure 4** and Equation 36), which is gated (Equation 37) by the retinotopic monocular object boundaries (Equation 22) that play the role of filling-in barriers (Grossberg and Todorović, 1988; Grossberg, 1994). The model computes filled-in binocular surfaces in separate double-opponent ON and OFF Filling-In Domains, or FIDOs (Equations 38–40). The final binocular percept is computed as the rectified sum of the ON and OFF FIDO activities [Equation (41) and **Figures 6–9** for simulation results]. This computation enables both light and dark filled-in surfaces to attract spatial attention in a surface-shroud resonance (see **Figure 4**).

The monocular and binocular FIDOs are computed in retinotopic coordinates, corresponding to the percept that objects that are seen with coarse spatial resolution when the fovea looks elsewhere are seen with cortically-magnified high acuity when they are themselves foveated. The surface contour signals that are derived from these filled-in surfaces are also computed in retinotopic coordinates. These surface contour signals are used to compute the eye movement signals that can command the eyes to move the correct direction and distance to foveate the commanded new fixation position. Aspects of how this happens have been simulated in neural models of saccadic eye movements (e.g., Grossberg et al., 1997; Gancarz and Grossberg, 1998, 1999; Silver et al., 2011).

On the other hand, the invariant binocular boundaries that maintain their fusion across eye movements are computed in head-centered coordinates, even though the monocular left and right boundaries on which they build are initially computed in retinotopic coordinates. Gain fields at several processing stages (**Figures 3**, **4**) cause predictive remapping between these several retinotopic and head-centered representations to maintain binocular fusion of the head-centered boundary representations while eye movements occur.

The head-centered invariant binocular boundaries (Equation 33) regulate surface filling-in within the two retinotopic monocular FIDOs (**Figure 4** and Equations 36, 37), which in turn form retinotopic binocularly-fused, or binocular, surface percepts (**Figure 4** and Equations 38–40). The head-centered binocular boundaries are converted into retinotopic binocular boundary signals (Equation 40) via gain fields (**Figure 4** and Equations 42–44) before they interact with the retinotopic monocular FIDOs. The retinotopic binocular surface percept can support a conscious percept of visible 3D form. Such a consciously seen surface percept in depth is maintained across eye movements due to the predictive remapping of their supporting boundaries by gain fields which occurs at several processing stages (**Figure 4** and Equation 38).

The retinotopic binocular surfaces generate surface contour output signals (**Figure 4** and Equation 45) through contrast-sensitive shunting on-center off-surround networks (Equations 46, 47). The surface contour signals (Equation 45) provide feedback (Equation 40) to the head-centered binocular boundaries (Equation 33) after being converted back to retinotopic

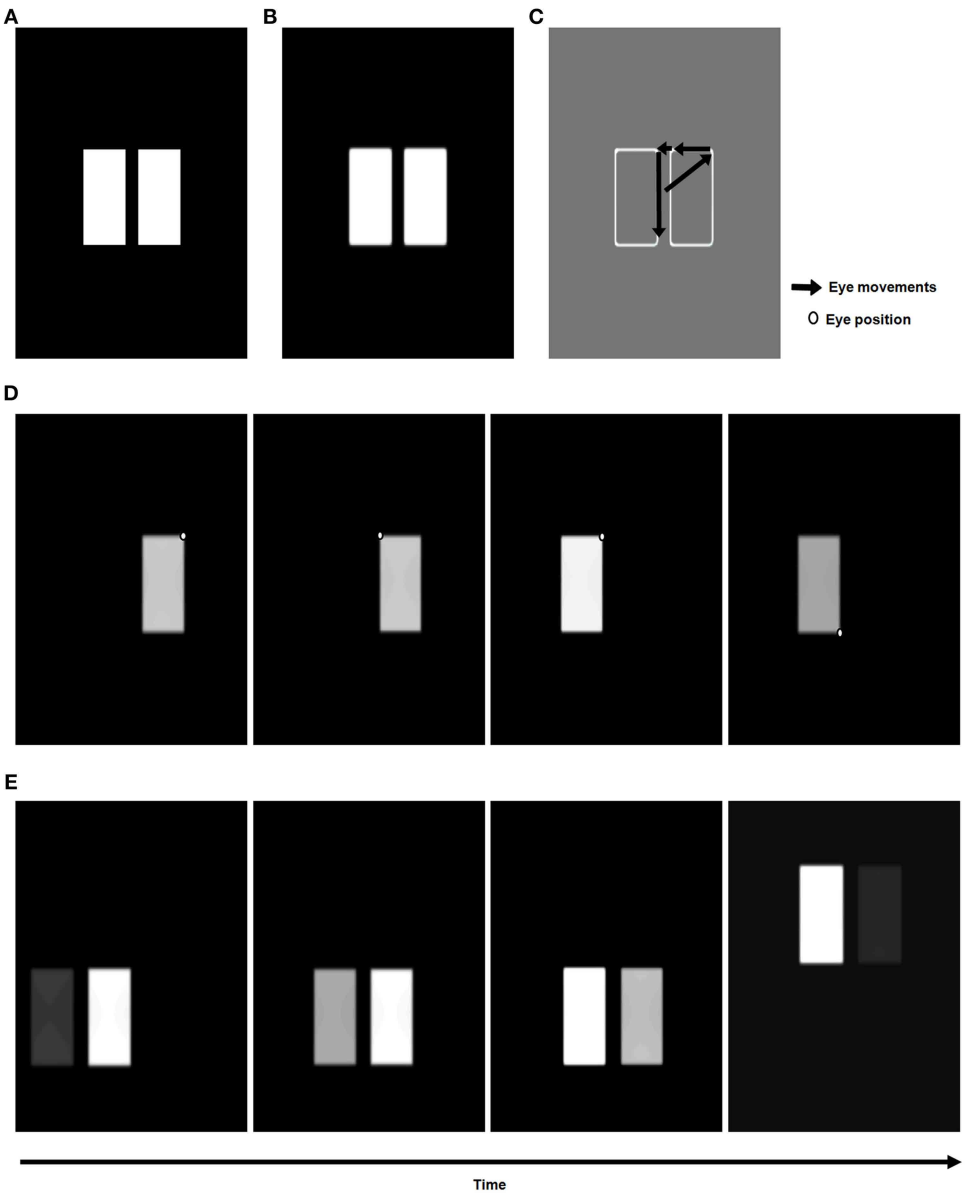


FIGURE 6 | Model simulations of the 3D ARTSCAN model with simple homogenous surfaces showing stability of binocular surface fusion. **(A)** The retinal input (I) (Equations 1–3) is a scene containing only two simple objects: two homogenously filled rectangles. This retinal image is presented monocularly to both the eyes. All simulation results are shown for far allelotropic shifts of $+3^o$. **(B)** In the absence of any eye movements, an initial binocular surface percept (S^b) (Equation 41) is formed through the mechanisms of the pre-attentive processing stage for boundaries and surfaces (**Figures 2**, **3**). **(C)** The surface contour map (C) (Equation 45) with a cumulative record of all the eye movements to target positions (Equation 66) made within and across the object surfaces is shown. **(D)** As an initial surface percept is formed, competition in the spatial attention map helps to choose a winning attentional shroud (A) (Equation 51). The shroud is represented in head-centered coordinates. The eye movements are initiated to salient target positions on the surface contour of a given object surface. In this simple stimulus, the salient features in the surface contours are always one of the corners of the rectangles. The first such surface shroud is activated with an eye movement to the top right corner of the rectangle on the right. Over time, a new target position (dots at rectangle corners) is chosen within or outside the object surface and the next saccade is made. **(E)** The fused binocular surface percept (Equation 41) after each eye movement to a salient feature is shown. Despite eye movements and the collapse of one surface shroud leading to another, the overall binocular surface percept is maintained in retinotopic coordinates. The active surface-shroud resonance enhances the brightness of the attended surface. See Section 4.1 for details.

coordinates by gain fields (**Figure 4** and Equations 48–50). The surface contour signals from a surface back to its generative boundaries strengthen consistent boundaries, inhibit irrelevant boundaries, and trigger figure-ground separation (**Figure 4**; Grossberg, 1994; Kelly and Grossberg, 2000). The feedback interaction between boundaries, surfaces, and surface contour signals is predicted to occur between V2 pale stripes and V2 thin stripes.

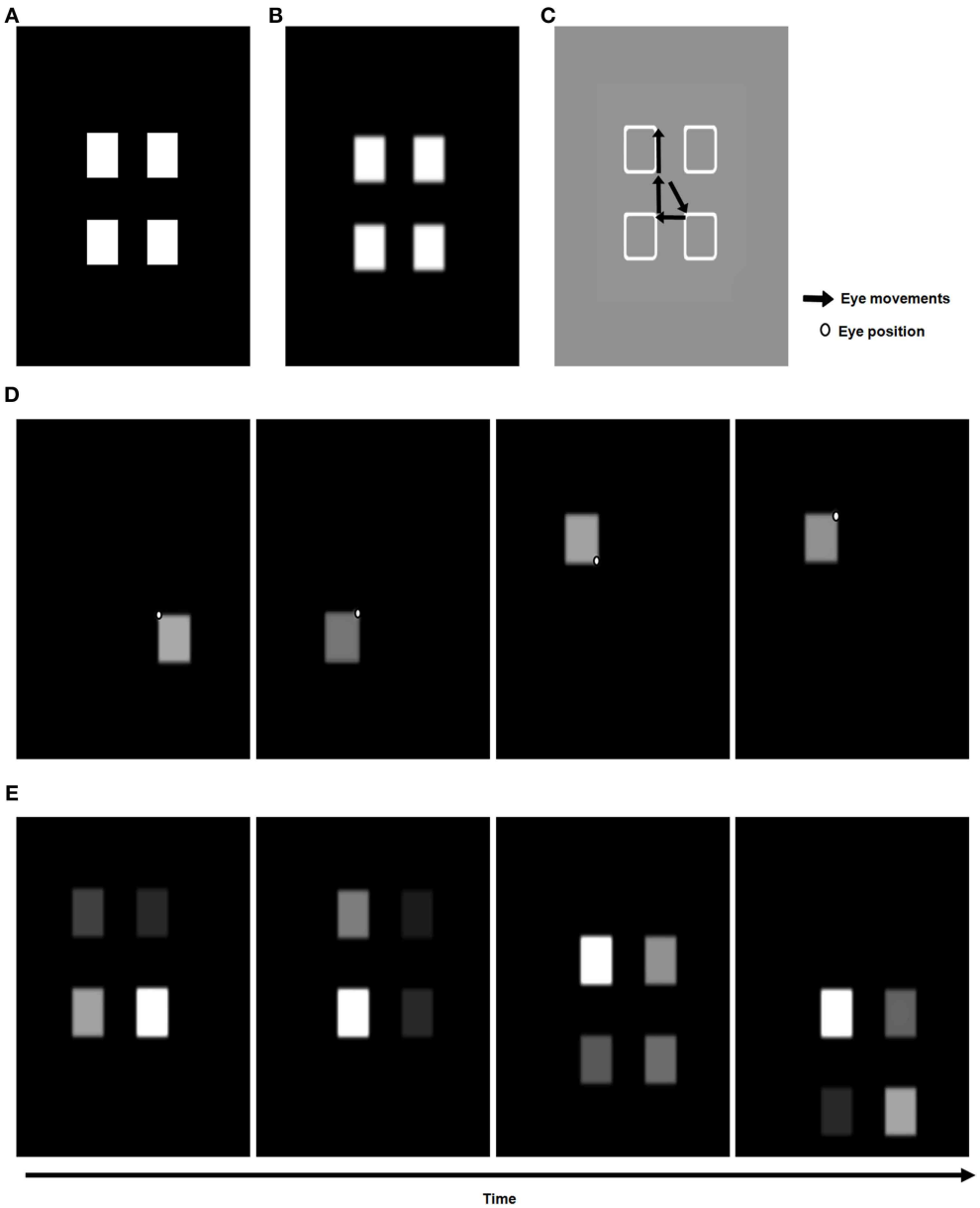


FIGURE 7 | Model simulations of the 3D ARTSCAN model showing stability of binocular surface fusion with four homogenous objects. This simulation illustrates that the model scales robustly without any parameter changes. The simulation environment and results are similar to those shown in **Figure 6**.

The coordinated action of all these gain fields acting between boundaries and surfaces, taken together with the surface-based spatial attentional shroud, achieves predictive remapping of the binocularly fused and attended surfaces. See Section 5 for details.

Although the surface filling-in here is modeled by a diffusion process, as in Cohen and Grossberg (1984) and Grossberg and Todorović (1988), Grossberg and Hong (2006) have modeled key properties of filling-in using long-range horizontal connections that operate several orders of magnitude faster than diffusion. Both processes yield similar results at equilibrium.

3.5. SPATIAL SHROUDS

A surface-shroud resonance fixes spatial attention on an object that is being explored with eye movements. The spatial attention neurons interact via recurrent on-center off-surround interactions (Equations 51–55) whose large off-surround enables selection of a winning attentional shroud. The recurrent on-center interactions enhance the winning shroud, and enable this shroud to remain active as other attentional neurons are persistently inhibited. Top-down attentional feedback from the resonating shroud (Equation 56) increases the contrast of the attended surface (Equation 39).

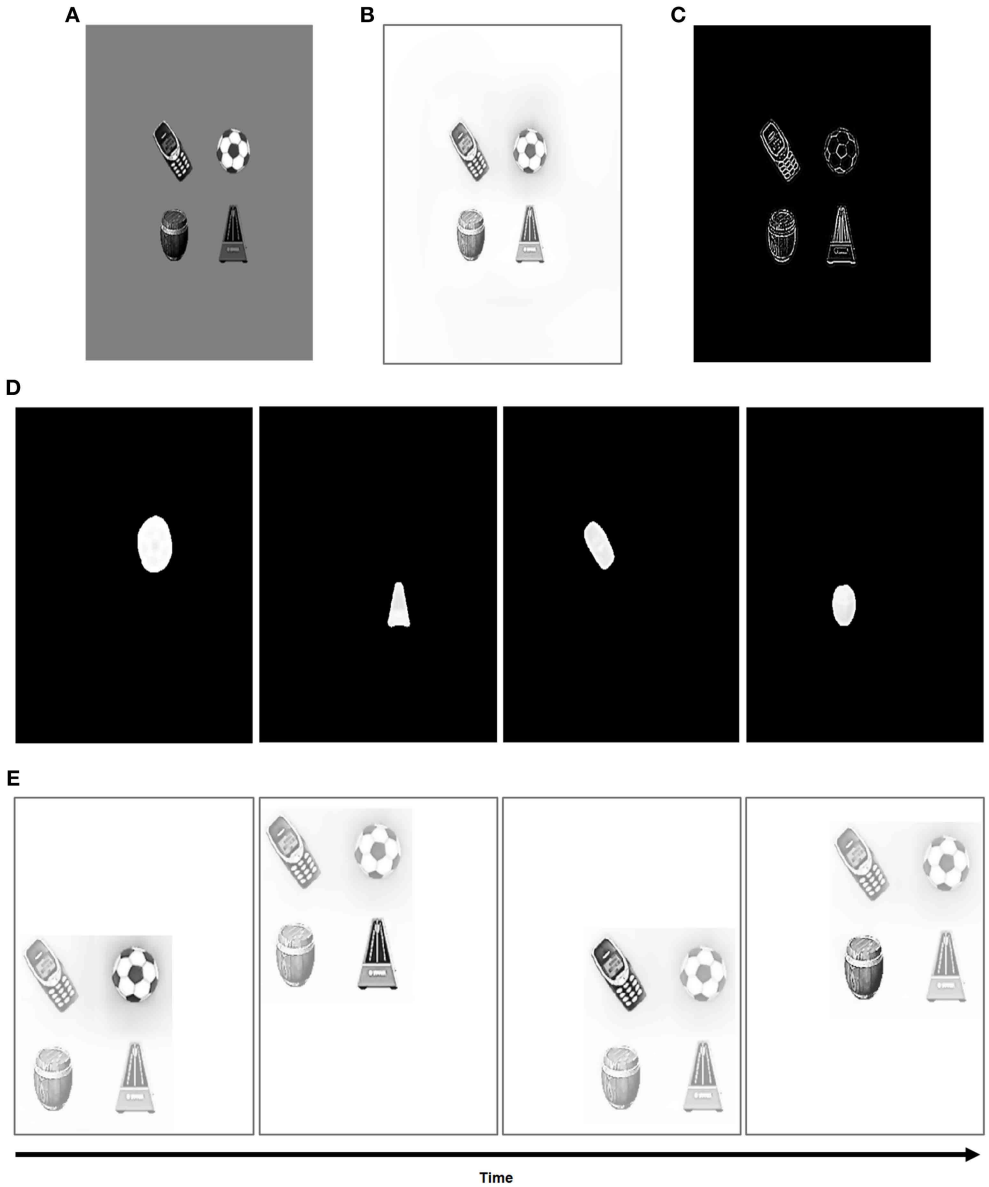


FIGURE 8 | Model simulations with natural objects showing binocular surface stability. The results are presented similar to those in **Figures 6**, **7**. The input consists of four non-overlapping grayscale objects with uniform and noiseless gray backgrounds from the Caltech 101 image database (Fei-Fei et al., 2004). The pre-attentive processing stages of the model enabled both the fusion and perceptual quality, including adaptation of ambient illumination, of the binocular surface percepts. Using ON and OFF channels for both boundary and surface representations (e.g., Equations 13–16) improved the perceptual quality of the attended surfaces. **(A)** Input I to the system. **(B)** Initial binocular surface percept S^b (Equation 41). **(C)** Surface contour map C (Equation 45). **(D)** Attentional shrouds A (Equation 51) over time. **(E)** The activity of the binocular surface percept (S^b) over time. Several saccades were made within each object's surface contour before moving to the next object. Detailed temporal dynamics of activity of attended shrouds and surfaces are shown in **Figures 10–13**.

Such a resonance habituates through time in an activity-dependent way (Equations 51, 61; Grossberg, 1972). Winning shrouds will thus eventually collapse, allowing new surfaces to be attended and causing inhibition of return (IOR). In addition, when a shroud collapses sufficiently during the first moments of a spatial attentional shift, a transient burst of activation by a reset mechanism (Equations 62, 63) helps to complete the collapse of the shroud (Equation 51), as well as to reset the invariant object category in the What stream.

As noted above, object surface input is combined with eye position signals via gain fields to generate a head-centric spatial attentional shroud in the parietal cortex (**Figures 4**, **5**). Such gain field modulation is known to occur in posterior parietal cortex (Andersen and Mountcastle, 1983; Andersen et al., 1985; Gancarz and Grossberg, 1999; Deneve and Pouget, 2003; Pouget et al.,

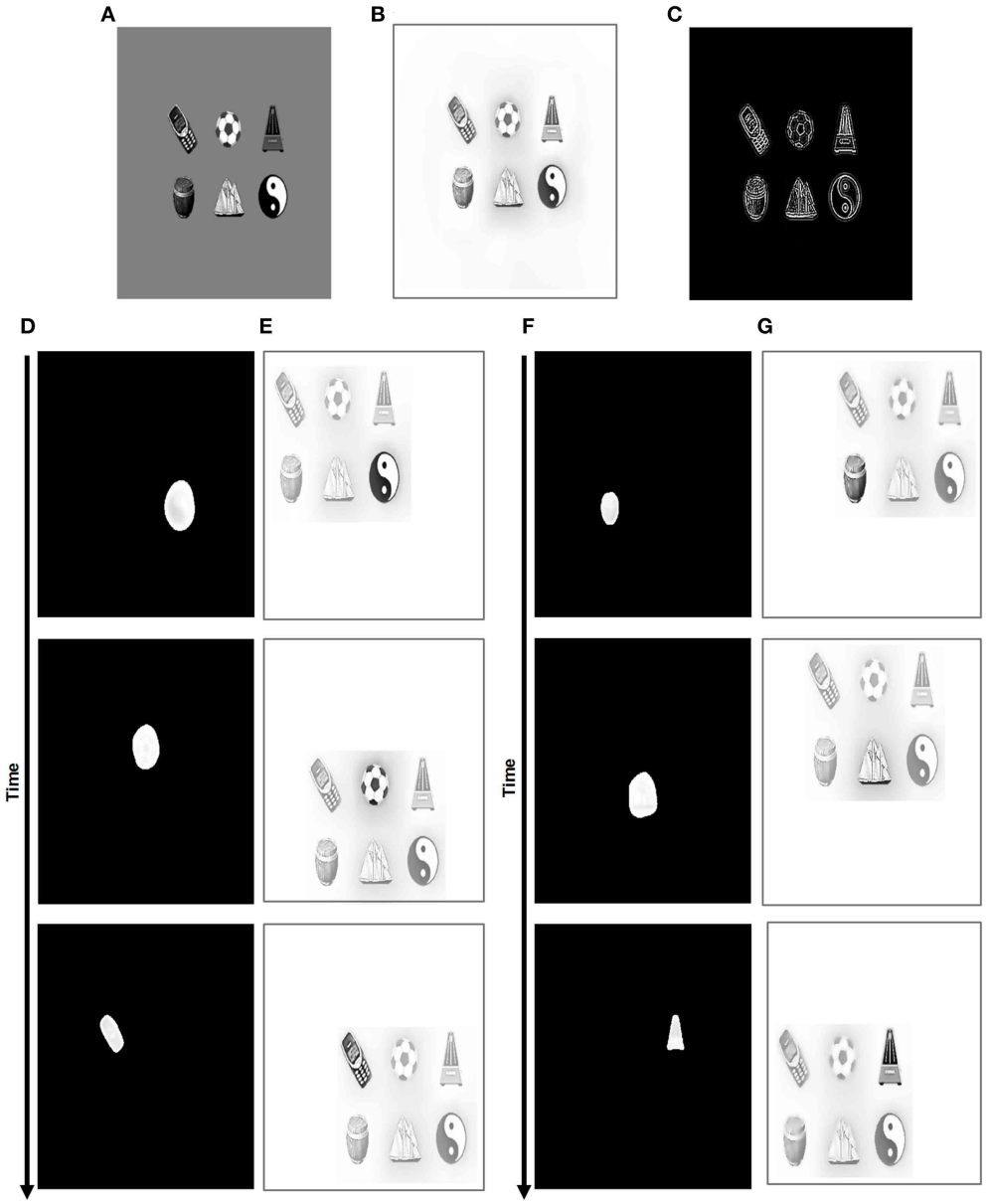


FIGURE 9 | Model simulations with an increased number of natural objects. The stimulus and results presented are similar to those in **Figure 8**, except that the number of objects in the scene is increased to six. **(A)** Input. **(B)** Initial binocular surface percept. **(C)** Surface contour map. (**D,F**) Attentional shrouds over time. (**E,G**) Activity of binocular surface percepts over time.

2003). The inputs from the gain fields (Equations 56–60) activate attentional interneurons (Equation 55) that interact through recurrent excitatory signals with attentional cells that excite and inhibit each other via a recurrent on-center off-surround network whose cells obey membrane equation, or shunting, laws (Equation 51).

3.6. EYE SIGNALS

The eye movement signals serve a major role in predictive remapping of boundaries, surfaces, and shrouds. They also determine the object views that will be attended, and thus which view-specific categories will be learned and associated with the emerging view-invariant object category. The eye movement signals are generated from the surface contour signals (Equation 45) that are derived from the currently active surface-shroud resonance. Surface contour signals tend to be larger at high curvature points and other salient boundary features due to the contrast-enhancing on-center off-surround interactions that generate them from filled-in surface lightnesses and colors. The surface contour signals are further contrast-enhanced to choose the position with the biggest activity, using a recurrent shunting on-center off-surround network (Equations 64–66). This transformation from surface contours to the next eye movement target position is predicted to occur in cortical area V3A

(Nakamura and Colby, 2000; Caplovitz and Tse, 2007). These eye movement signals are used to predictively update all the gain field signals (e.g., Equation 48), even before they generate the next saccadic eye movement. The chosen eye movement signal (Equation 66) habituates in an activity-dependent way (Equation 65) and hereby realizes an inhibition-of-return process that prevents perseveration on the same eye movement choice, thereby enabling exploration of multiple views of a given object. See Section 5 for details.

4. SIMULATION RESULTS

The entire input visual field is a 3000 × 3000 pixel grid with coordinates (i, j) and input intensity I_{ij}. Each pixel step corresponds to a distance of 0.01^o in visual space, so that each input spans $30^o \times 30^o$ in Cartesian space. All object surfaces in the stimulus are within 5^o on either side of the fixation point. Eye movements were controlled to be within 10^o of the entire visual field—that is, within the parafoveal region—in order for binocular fusion to be possible. In order to simulate the effects of binocular inputs, the simulations were performed with the monocular inputs shifted with respect to one another by $+3^o$ (allelotropic far shift). Thus, the inputs to the left and right eye are $I^l_{(i+3^o)j}$, and $I^r_{(i-3^o)j}$, respectively. Binocular fusion also works for other allelotropic shifts, far and near, within the range of binocular fusion, as demonstrated in Cao and Grossberg (2005). The range of values of the allelotropic shift s, and thus the number of depth planes simultaneously represented in the 3D ARTSCAN model, are $\{+8^o, +3^o, 0^o, -3^o, -8^o\}$. The model can readily be extended, without a change of mechanism, to represent any finite number of depth planes. In all the simulations, the initial fixation point was not on any object and was at the center of the visual field. The simulations show how the model's disparity sensitivity to the monocular left and right eye inputs leads to selective activation of the depth plane that is represented by the allelotropic far shift.

4.1. SIMULATIONS OF BINOCULAR FUSION OF HOMOGENEOUS SURFACES

The first simulation tested the ability of 3D ARTSCAN to maintain stable binocular fusion using rectangular-shaped objects as the eyes explored them in a scene. The input consisted of a scene with either two homogenously filled rectangles of equal size (**Figure 6A**) or four homogeneously filled squares (**Figure 7A**) on either side of the initial eye fixation point before any eye movements occurred. Each of the rectangles in **Figure 6A** is 300 × 400 pixels in size. The square stimuli in **Figure 7A** are each 200 × 200 pixels. The pixellated images are converted into a rectilinear grid in terms of degrees of visual angles as described earlier.

After the initial binocular surfaces are computed, the surface contour map (Equation 45) is also computed, and is shown in **Figures 6C**, **7C** before any eye movements occur. Due to the contrast-sensitive on-center off-surround interactions that generate surface contours from successfully filled-in surfaces, the positions of highest activity (salient features) occur at the corners of the rectangles. When the maximum activities are chosen by a subsequent on-center off-surround network (Equation 66), they determine the targets of the eye movements, which are shown as black arrows. In **Figure 6C**, the chosen salient feature initiates the first predictive eye movement to the top right corner of the rectangle on the right, consistent with the fact that the rectangle on the right is part of an active surface-shroud resonance (first panel, **Figure 6D**). Similarly, for the stimulus with four squares, the first eye movement is initiated to the top left corner of the bottom right square (**Figure 7C**) after the spatial attentional shroud is formed over the corresponding square surface (first panel, **Figure 7D**). As the eyes continue to move, the scene representation and perceptual stability of the fused binocular surfaces are maintained due to the predictive remapping of the boundaries and surfaces by the gain fields, which ensure that fusion is maintained as the eyes move to the next location. **Figures 6D**, **7D** show the activities of the head-center shrouds, and **Figures 6E**, **7E** show the activities of the corresponding surface representations, of the rectangles and squares through time. When spatial attention is focused on a particular surface as part of a surface-shroud resonance, its activity is enhanced. This is seen in the first panel of **Figure 6E**, where the rectangle on the right is more active (brighter) than the rectangle on the left. Similarly, the square on the bottom right is more active than others in **Figure 7E**. This is the fused binocular surface percept and is always in retinotopic coordinates. The attentional shrouds are computed in head-centered coordinates.

As the eyes freely scan the scene, they make several saccades within and across the different object surface contours. As this happens, spatial attention moves from one object, disengaging before engaging another object, based on the salient features in the surface contour map (see **Figure 5**). A temporal evolution of the spatial attention and binocular percepts are shown from left to right in **Figures 6D,E**, **7D,E**, respectively, for the two stimuli. Before the eyes can move from one object to the other, the currently active attentional shroud begins to collapse due to habituation (Equation 61), which leads to its reset (Equation 62). Multiple saccades move sequentially to the most salient positions on one object's surface contours before moving onto another object's surface contours.

These simulations establish a proof of concept that the extension of the ARTSCAN model to the 3D ARTSCAN model maintains stable fusion of binocular surfaces as the eyes explore them and other objects in their vicinity.

4.2. SIMULATIONS OF BINOCULAR FUSION OF NATURAL OBJECTS

Simulations were also carried out using 3D scenes with natural objects in them. For this set of simulations, grayscale images of objects from the Caltech 101 dataset (Fei-Fei et al., 2004) were used. The image backgrounds are a uniform gray and do not have any noise or texture. Each object is 100 × 100 pixels in size. The objects were tiled on the visual field, and two sets of stimuli with four (**Figure 8A**), and six (**Figure 9A**) objects were used to test the system's robustness and scalability to more realistic scenes. These pixellated images were rescaled to a rectilinear grid into degrees of visual field, as described earlier. The naturally occurring objects used in the simulations are "cell phone," "soccer ball," "metronome," "barrel," "yacht," and "yin yang."

The pre-processing stages for the natural objects are the same as for the rectangular and square stimuli in **Figures 6**, **7**. The

initial binocular surface percept that is represented in retinotopic coordinates is shown in **Figures 8B**, **9B** for the four and six image stimuli, respectively.

The surface contour maps for the natural objects, before any eye movements occur, are shown in **Figures 8C**, **9C**. These simulation figures show the results of when the eyes move from one object's surface contour to the other after the shifting of attentional shrouds. The maintenance of binocular fusion as the eyes move across a single object's surface, followed by shroud collapse and an eye movement to another object, are explained, with simulations, in the remainder of this section and in Section 4.3.

In **Figure 8**, the first eye movement is made to the soccer ball. Thus, the first spatial attentional shroud is linked to the soccer ball (first panel, **Figure 8D**). After several saccades explore the soccer ball using its surface contour map to determine salient saccadic target positions, the shroud begins to collapse and spatial attention begins to shift to the metronome as the next eye movement is made to a position chosen from the metronome's surface contour (second panel, **Figure 8D**). This process then proceeds to the cell phone (third panel, **Figure 8D**) and then finally to the barrel (fourth panel, **Figure 8D**). Several saccades are made within each object, thus exploring the object and learning invariant object categories for it (Fazl et al., 2009; Grossberg, 2009; Cao et al., 2011), before moving onto the next object. During all these saccadic eye movements within or across objects and shifts in attention across objects, all the binocular surfaces are maintained in fusion in retinotopic coordinates (**Figures 8E**, **9E,G**). Each panel that illustrates the binocular percept shows enhanced activity of the currently attended object surface.

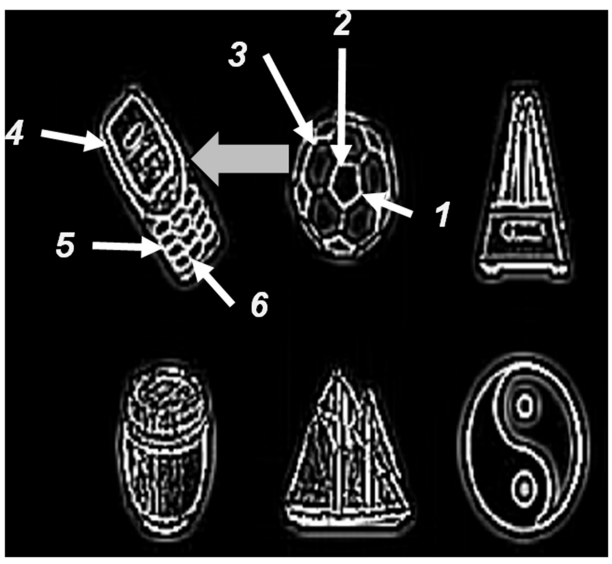


FIGURE 10 | Surface contour activity *C* (Equation 45) with attention first maintained on the soccer ball, followed by a then shift in attention to the cell phone. Saccades to target positions marked "*1*," "*2*," and "*3*" are made within the soccer ball. Saccades to target positions marked "*4*," "*5*," and "*6*" are made within the cell phone after a shift in attention. The thick gray arrow marks the shift in attention from the soccer ball to the cell phone following parietal reset (see Section 4.3 for details).

The same experiment was repeated with more stimuli (six instead of four) in the scene to test the scalability and robustness of the system; see **Figure 9**. Here, the first predictive eye movement is made to the yin yang symbol (first panel, **Figure 9D**) as its attentional shroud suppresses the shrouds of the other objects. After a few saccades on the yin yang surface contour, an eye movement is made to the soccer ball surface contour as spatial attention is disengaged from the yin yang and engaged with the soccer ball (second panel, **Figure 9D**). After this, an eye movement is made to the cell phone surface contour: spatial attention is disengaged from the soccer ball, and engaged with the cell phone (third panel, **Figure 9D**). This is then followed by an eye movement to the barrel, yacht, and finally to the metronome (panels in **Figure 9F**). Within each object, several saccades were made before moving onto the next object (see **Figure 10**).

The binocular surface percept remains fused in retinotopic coordinates while all this change occurs in spatial attention and eye movements. Here again, the perceptual contrast of the attended surface, which is in surface-shroud resonance, is enhanced (**Figures 8E**, **9E,G**). This simulation shows that system properties, using the same set of parameters, are robust in response to variable numbers of natural images. The invariant binocular boundaries were as well maintained in fusion by the predictive remapping signals. These dynamics are elaborated in Sections 4.3 and 4.4.

4.3. SIMULATIONS OF WITHIN OBJECT EYE MOVEMENTS AND ATTENTION SHIFTS BETWEEN OBJECTS

Sections 4.1–4.2 and **Figures 6–9** summarized simulations that illustrate how homogeneous surfaces (rectangles and squares) and natural objects induce surface representations that remain binocularly fused as attention shifts from one object to another during scanning eye movements. **Figure 10** describes the surface contours (Equation 45) before any eye movements occurred, as well as six of the eye movement target positions that were determined by the surface contours and which led to eye movements.

When attention is disengaged from the yin yang and shifts to the soccer ball, the fixated eye position (Equation 66) within the soccer ball is marked as "*1*" on the surface contour in **Figure 10**. The activities of the attentional shroud and the fused binocular surface after the eye position "*1*" is attained are shown in **Figures 9D,E** (second row), respectively. Following this, two more saccades numbered "*2*" and "*3*" are made to surface contour salient features of the soccer ball (**Figure 10**). While these saccadic explorations are made within the soccer ball, its shroud starts to collapse due to a combination of inhibition of return and habituation. This disinhibits and triggers the burst of the parietal reset signal (Equation 62), which was thus far inhibited by the active shroud of the soccer ball. This burst of the reset signal collapses the habituating attentional shroud on the soccer ball completely, thus initiating a shift in spatial attention (thick gray arrow) from the soccer ball to the cell phone. Once the spatial shift in attention to the cell phone occurs, the new eye position (Equation 66) within the cell phone is marked as "*4*" on the surface contour (**Figure 10**). Two saccades numbered

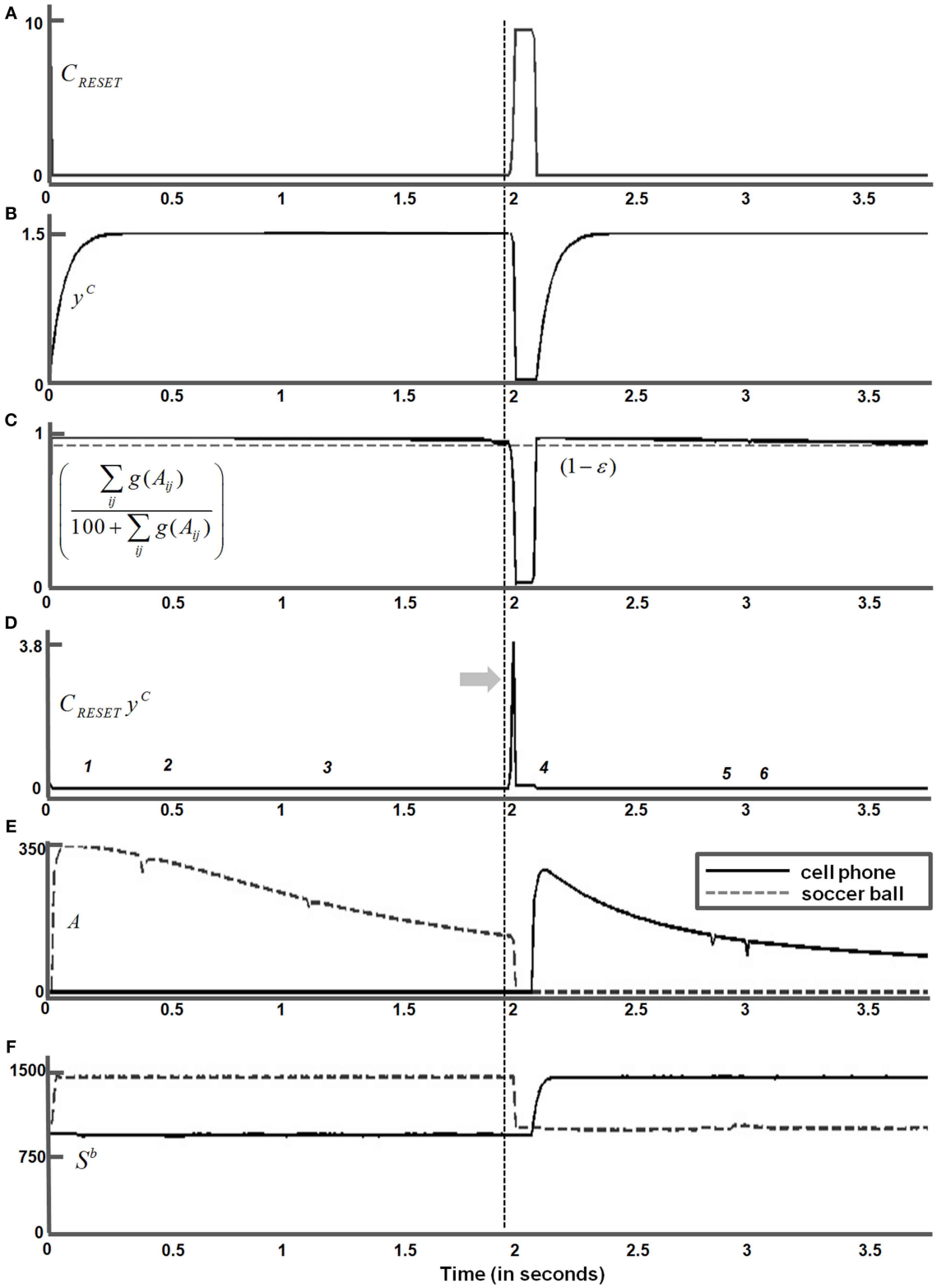


FIGURE 11 | Temporal dynamics after attention is engaged by the soccer ball and saccades are made within it, followed by a shift in attention to the cell phone and saccades within the cell phone. (A) Temporal evolution of the parietal reset signal C_{RESET} (**Figure 4** and Equation 62) for the paradigm described in **Figure 10**. When saccades are made within the attended object, C_{RESET} remains inhibited, thereby allowing for explorations of different views within the attended object that can be learned and associated with a view-invariant category of the object. A few moments after $\frac{\sum_{ij} g(A_{ij})}{100+\sum_{ij} g(A_{ij})}$ in **Figure C** crosses beneath the threshold $(1-\varepsilon)$, the parietal

(Continued)

FIGURE 11 | Continued
reset signal is disinhibited and inhibits the currently active shroud, thereby enabling a shift in spatial attention. The time when C_{RESET} turns on is marked by the dashed vertical line. When the next winning shroud starts to become active **(E)**, it inhibits the reset signal. **(B)** The habituative neurotransmitter y^C (Equation 63) is at its maximum activity when the reset signal is inhibited. When the reset signal is activated, the transmitter habituates in an activity-dependent way. The net reset signal $C_{RESET}y^C$ that inhibits the spatial attention map (Equation 51) is therefore transient. An attention shift to a new surface-shroud resonance can hereby develop after it shuts off. When the reset signal is inhibited by the newly active shroud, the habituative neurotransmitter gradually replenishes over time before the next reset event occurs. **(C)** The temporal evolution of the ratio of the attention function $\frac{\sum_{ij} g(A_{ij})}{100+\sum_{ij} g(A_{ij})}$ that is subtracted from the constant threshold $(1-\varepsilon) = 0.93$ to define the parietal reset signal. As long as the ratio of the attention function remains above the threshold, the reset signal remains inhibited. After the ratio crosses the threshold (marked by the dashed vertical line), the parietal reset signal is turned on. **(D)** The transient reset burst $C_{RESET}y^C$ inhibits the spatial attention map. **(E)** Temporal evolution of the attentional shrouds A (Equation 51) of the soccer ball and cell phone. The reset mechanism does not collapse the shroud when saccades (e.g., "*2-3*" or "*5-6*" in **Figures 10**, **11D**) are made within the surface of an active shroud. The small dips in activity of the active shroud correspond to saccades within the attended object. **(F)** Temporal evolution of the binocular surface percepts S^b (Equation 41). The attended binocular surface activity (dashed curve, soccer ball; solid curve, cell phone) is enhanced by surface-shroud resonance. See Section 4.3 for details.

"*5*," "*6*" are next made within the cell phone. The binocular surface percept and attentional shroud activity of the cell phone, for the position marked as "*6*" was shown previously (third panel, **Figures 9D,E**).

The temporal evolution of the parietal reset signal (**Figure 4** and Equation 62) during these six eye movements (**Figure 10**) is shown in **Figure 11A**. A reset signal occurs only when the soccer ball shroud collapses, thereby enabling a spatial attention shift to the cell phone. The eye movements within these objects do not cause a reset signal. The temporal profile of the habituative transmitter (**Figure 4** and Equation 63) that gates the parietal reset signal is shown in **Figure 11B**. The temporal evolution of the ratio $\frac{\sum_{ij} g(A_{ij})}{100+\sum_{ij} g(A_{ij})}$ that is subtracted from the constant threshold $(1-\varepsilon)$ to define the parietal reset signal C_{RESET} in Equation (62) is shown in **Figure 11C**. When $\frac{\sum_{ij} g(A_{ij})}{100+\sum_{ij} g(A_{ij})}$ becomes smaller than $(1-\varepsilon)$, C_{RESET} turns on at the time marked by the dashed vertical line, as in **Figure 11A**, and the habituative gate begins to decay in an activity-dependent way, as in **Figure 11B**. As a result, the net reset signal $C_{RESET}y^C$ in **Figure 11D** is a transient burst. This transient burst completely inhibits the active soccer ball shroud (dashed line) in **Figure 11E** via Equation (51). There is a time lag between the activation of successive shrouds, following the collapse of soccer ball shroud and the formation of the cell phone shroud (solid line), that corresponds to the time needed to shift spatial attention between the two objects (**Figure 11E**). The inhibition of the soccer ball shroud enables the cell phone shroud to win the competition for spatial attention. The binocular surface representation of the cell phone (**Figure 11F** and Equations 38–41) is then enhanced by top-down excitatory feedback from its shroud as a surface-shroud resonance develops. The newly activated shroud inhibits the tonically active reset signal (**Figure 11A**) and the habituative transmitter gradually recovers through time (**Figure 11B**). These dynamics repeat when next reset event occurs.

Figure 12 presents the evolution of the activities shown in **Figure 11** at finer temporal resolution at times just before, during, and after the occurrence of the reset event so that the reader can better appreciate these temporal details. When saccades (e.g., "*2–3*" or "*5–6*" in **Figure 10**) are made within the surface of an active shroud, they do not cause the reset mechanism to collapse the shroud. The small dips of activity in the active shrouds in **Figure 11E** correspond to such eye movements within an object. As a result of these saccadic explorations within an attended object, different view-specific categories of the object can be learned and associated with a view-invariant category of the object (see What stream of ARTSCAN in **Figure 1**).

Figure 13 shows the simulated activity profiles of the attentional shroud and binocular surface representations when saccades are made, as summarized in **Figure 10**, within an attended surface, and after shifts in attention to other surfaces. **Figure 13A** shows the profiles of the attentional shrouds which are represented in head-centered coordinates, and **Figure 13B** shows the profiles of the corresponding binocular surface percepts in retinotopic coordinates. The markings "*2*," "*3*," "*4*," "*5*," and the thick gray arrow on the sides of each pair of panels correspond to the eye positions after each saccade, and the shifts in attention described in **Figures 10–12**.

Figure 13C shows the average reaction time (RT) data in human subjects of Brown and Denny (2007). **Figure 13D** shows the average RTs to attend for the simulations shown in **Figure 9**. Average RTs in the simulations are computed on the spatial attention map(A) (Equation 51). The average reaction times for attending *within-object different position* (dark gray bar) after saccades are faster than the average response times for *between-object* (light gray bar) shifts of attention. The average reaction times for *within-object different position* after saccades were calculated as the time it takes the active shrouds to recover from the small dips in activity, corresponding to eye movement made within the object to a different target position (e.g., **Figure 11E**). The average reaction times for *between-object* shifts in attention were calculated as the time between the complete collapse of the previous shroud and the activation of the next shroud to half its maximum value (**Figures 11E**, **12E**). The investigations of Brown and Denny (2007) showed that between-object shifts of attention take longer than within-object shifts. This within-object advantage occurs because attention need not be disengaged from the object when eye movements to target positions are made inside it. Brown and Denny (2007) also found that shifting attention from an object to another object, or to another position with no object present, takes nearly the same amount of time (369 ± 10 vs. 376 ± 9 ms), concluding that the engagement of attention is not the time limiting step in object-based experiments.

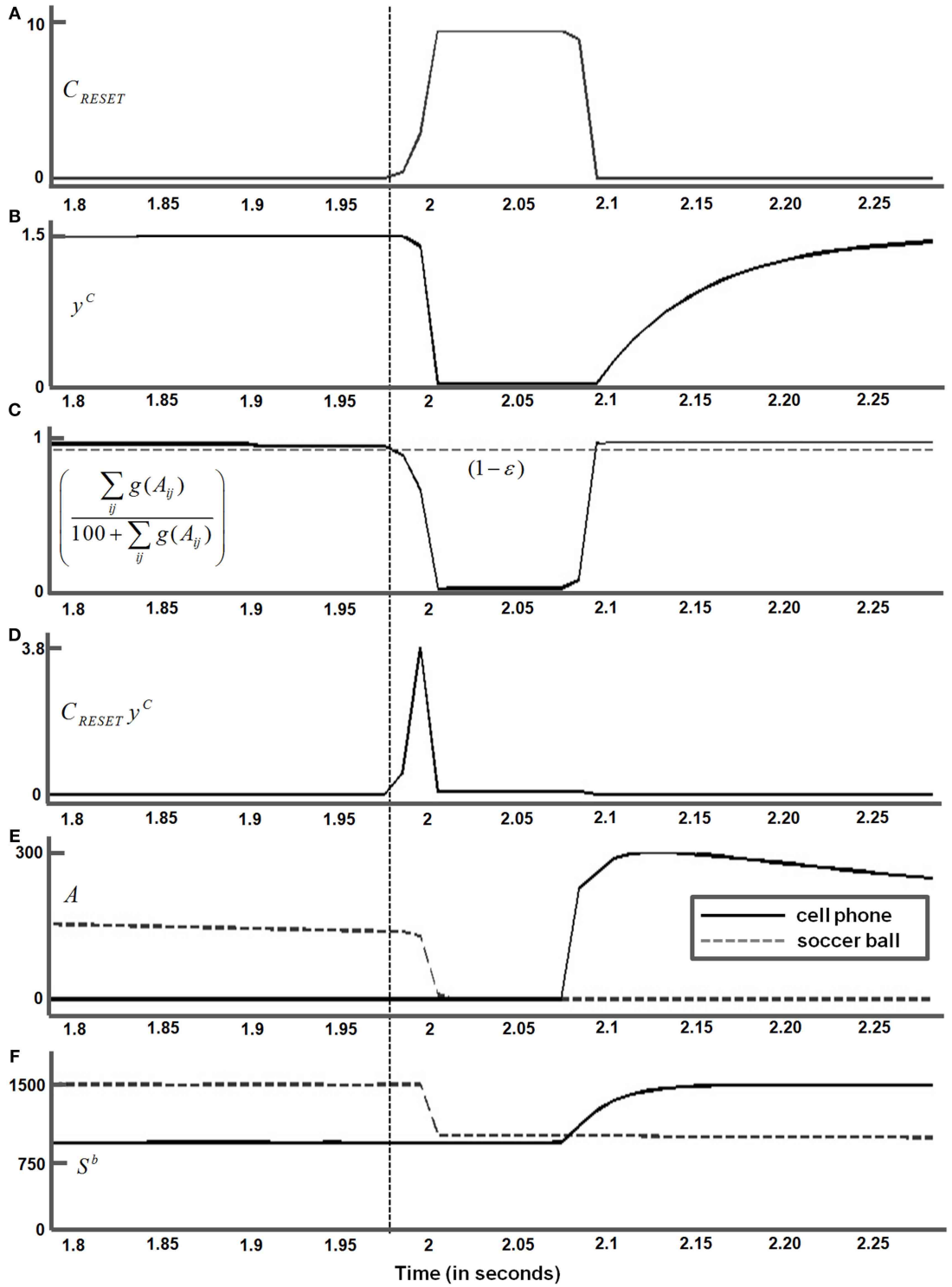


FIGURE 12 | Temporal dynamics of the plots in Figure 11, but at a finer temporal resolution before, during, and after the transient reset burst.

In the ARTSCAN model (cf. Fazl et al., 2009, **Figure 1**), the RTs for the corresponding simulations were scaled to be equal to the valid trials in the data. The dARTSCAN (cf. Foley et al., 2012) model has generalized ARTSCAN beyond its parietal spatial attentional capabilities to include prefrontal working memory storage, and has thereby extended the Fazl et al. (2009) simulations to quantitatively simulate all of the experimental cases described by Brown and Denny (2007). The 3D

ARTSCAN model replicates two of the trial conditions from the Brown and Denny (2007) experiment. The *within-object different position* (341 ± 9 ms, dark gray) and *between-object* (369 ± 10 ms, light gray) RTs in **Figure 13C** correspond to the invalid within, and invalid between, object trials of the experiment. The simulation RTs of *within-object different position* (40 ms, dark gray) and *between-object* (75 ms, light gray) presented in **Figure 13D** consistent with the data in **Figure 13C**. In

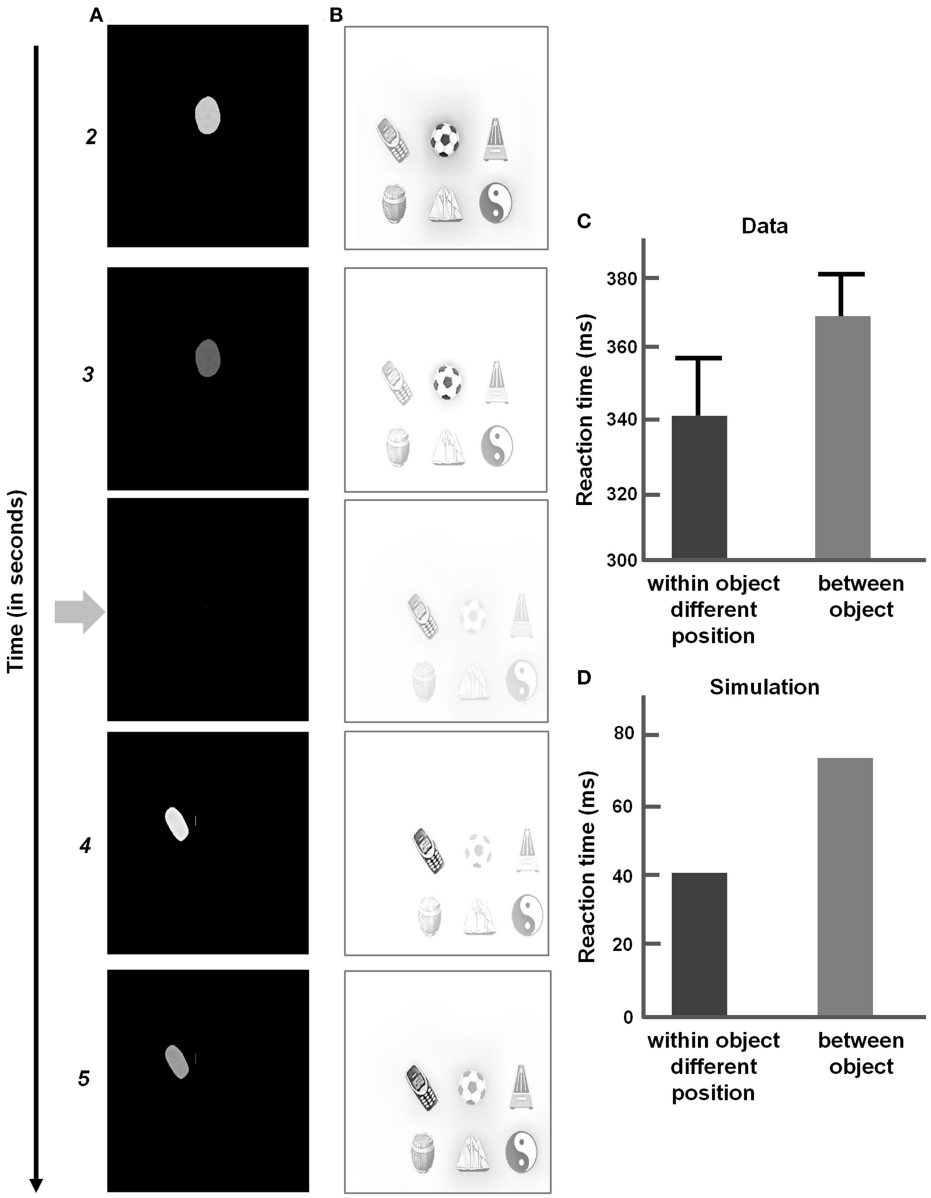


FIGURE 13 | Snapshots of the attentional shroud and the binocular surface percept during saccades within the soccer ball, followed by a shift in attention to the cell phone and a saccade within it. (A) Activities of attentional shrouds *A* (Equation 51) in head-centered coordinates after saccades to target positions "*2*," and "*3*" within the soccer ball, followed by an attentional shift to the cell phone (thick gray arrow), when no shroud is active, after which a cell phone shroud forms around target position "*4*," and then a saccade occurs within the cell phone to target position "*5*." **(B)** Corresponding activation patterns of the binocular surface percept (S^b) (Equation 41) in retinotopic coordinates. The eye positions and the attentional shift correspond to the paradigm explained in **Figure 10** and for the temporal profiles shown in **Figure 11** (see Section 4.3 for details). **(C)** Reaction time (RT) data from Brown and Denny (2007) for *within-object different position* (341 ± 9 ms, dark gray), and *between objects* (369 ± 10 ms, light gray) trials. **(D)** Simulations of RTs to object-based attention computed over the spatial attention map *A*. Average RTs to *within-object different position* (40ms, dark gray), and *between objects* (75 ms, light gray) are shown for the complete simulation run in **Figure 9**. RTs to attend to *within-object* different *positions* are faster than *between objects*, consistent with the data in **(C)** See Section 4.3 for an explanation of why the RT difference matches the data, but the total simulated RTs are 300 ms shorter.

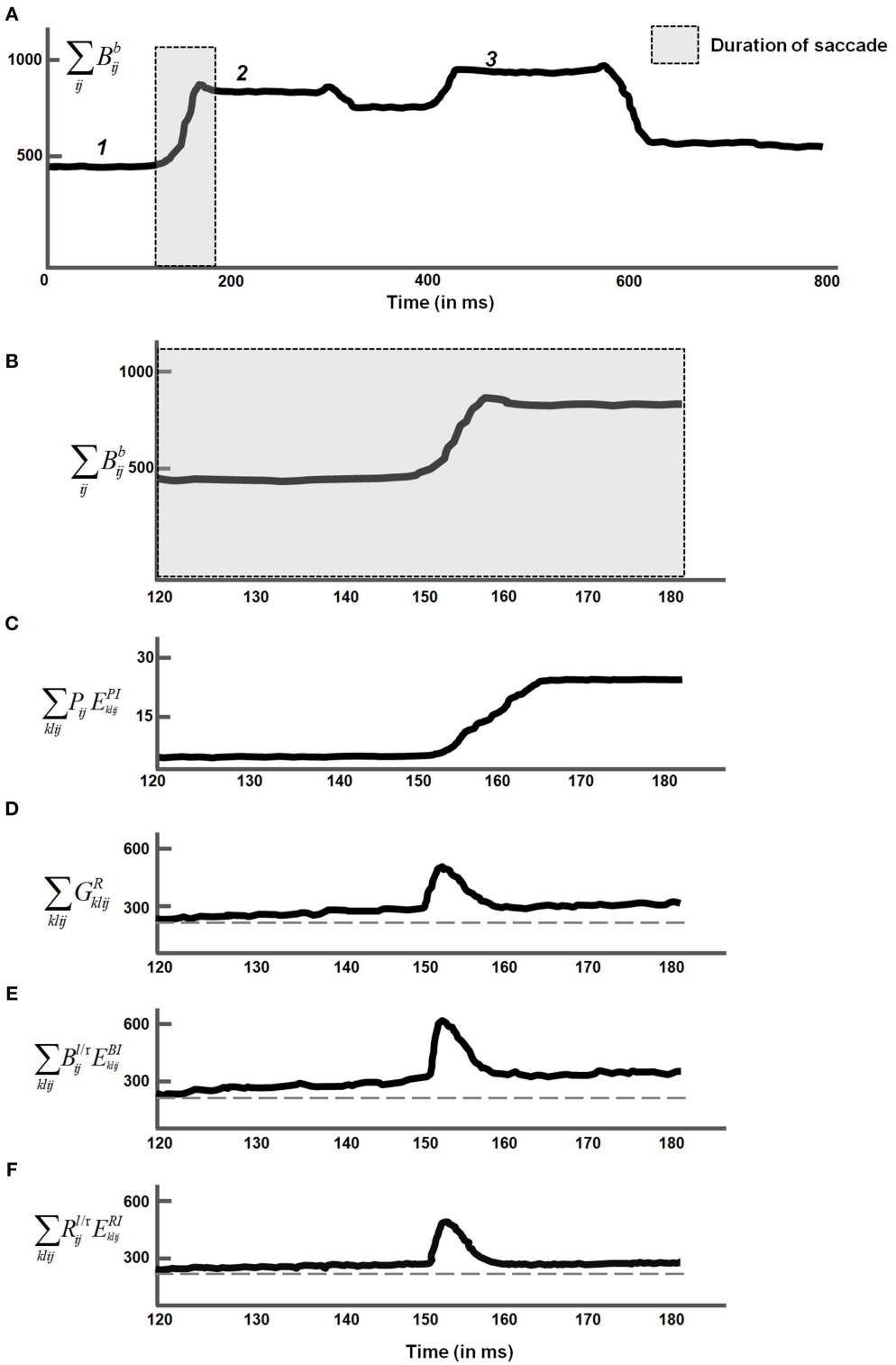


FIGURE 14 | Predictive remapping of fused invariant binocular boundaries. The input stimulus is the same as in **Figure 9** and the paradigm is from **Figure 10**. The maintained fusion of boundaries is demonstrated when saccades are made to target positions within one object, in this case, the soccer ball. For convenience, only ON channel (+) responses are shown. The OFF channel (−) responses look similar and thus the +/− superscripts are dropped for convenience. **(A)** Temporal evolution of the fused invariant binocular boundaries $\sum_{ij} B^b_{ij}$ (Equation 33) when saccades are made within the soccer ball. The markings "*1*," "*2*," and "*3*" correspond to the target positions on the surface contour map shown in **Figure 10**. The dashed gray box is the duration of the saccade (60 ms) for which the dynamics are presented in (**B–F**). **(B)** Temporal evolution of the invariant binocular boundaries $\sum_{ij} B^b_{ij}$ before, during, and after an eye movement to target position "*2*" in **Figure 10**

(Continued)

FIGURE 14 | Continued
following fusion of the invariant monocular boundaries. The dotted gray box shown covers the duration of the saccade shown in **(A)**. Even before the eye movement is completed, there is predictive remapping of the fused boundaries by the boundary gain fields. **(C–F)** show the boundary gain field activity for the left eye (*l*). The right eye profiles are the same. To achieve predictive remapping of the invariant left monocular boundary, the invariant left monocular boundary gain fields $G^{L^l}_{klij}$ (Equation 28) are activated by top-down inputs $\sum_{ij} B^l_{ij} E^{Bl}_{klij}$ from invariant left monocular boundaries (Equation 26), eye position signals $\sum_{klij} P_{ij} E^{Pl}_{klij}$ (Equation 66), and bottom-up inputs $\sum_{klij} R^l_{ij} E^{Rl}_{klij}$ from retinotopic left monocular boundaries (Equation 22). **(C)** Temporal profile of the eye position input $\sum_{klij} P_{ij} E^{Pl}_{klij}$. ((**D**) Temporal evolution of the summed invariant left monocular boundary gain field activity $\sum_{klij} G^{Rl}_{klij}$. **(E)** Temporal profile of the invariant left monocular boundary input $\sum_{klij} B^l_{ij} E^{Bl}_{klij}$. **(F)** Temporal evolution of the retinotopic left monocular boundary input $\sum_{klij} R^l_{ij} E^{Rl}_{klij}$. The gray dotted lines in **(D–F)** show the change in activity from baseline. See Section 4.4 for details.

ARTSCAN and dARTSCAN, trials were run explicitly instructing the system of the prime and cue, followed by a long interstimulus interval (ISI) before the target appears and a response is made with the appearance of the target. However, in 3D ARTSCAN, the cue and target selections are internally evaluated from the salient features on the surface contour map without any experimenter supervision, and only the response time is calculated from when the salient feature appears followed by an eye movement to the target position. The RTs shown here are thus 300 ms less than what was reported in Brown and Denny (2007).

4.4. SIMULATIONS OF PREDICTIVE REMAPPING OF BINOCULAR BOUNDARIES

Figures 14, **15** summarize simulations of predictive remapping by gain field modulation to maintain fusion of invariant binocular boundaries during eye movements. The inputs used in this analysis are the same as in previous sections (Sections 4.2–4.3 and **Figures 9**, **10**). The surface contour map from which eye position signals are generated is shown in **Figure 10**. The temporal dynamics of the predictive remapping of fused invariant binocular boundaries of all the objects are presented in **Figure 14** at the position marked "*2*" in **Figure 10** while saccadic eye movements are made to the target positions within the soccer ball to positions marked "*1*," "*2*," and "*3*."

Figure 14A shows the temporal profile of the summed response of the fused invariant binocular boundaries $\left(\sum_{ij} B^b_{ij}\right)$ (**Figures 3**, **4**, and Equation 33) for all the objects following a shift in attention from the yin yang to position "*1*" within the soccer ball. This is followed by two saccades to target positions "*2*" and "*3*" within the soccer ball. The duration of the saccade from position "*1*" to "*2*" is indicated by the gray dotted box, and is 60 ms. In all plots in **Figure 14**, only the ON channel profiles are shown. The OFF channel responses look similar. The $+/-$ superscripts are thus dropped for convenience. The summation of the invariant binocular boundary values $\left(\sum_{ij} B^b_{ij}\right)$ is plotted to show how the boundaries of all the objects are maintained in fusion while saccades are made to target positions within the soccer ball. This happens because the binocular boundaries are maintained in fusion in head-centered coordinates before the eye movement to the next target position, following predictive remapping of monocular boundaries in head-centered coordinates by monocular boundary gain fields (Equations 28–32). The monocular boundary gain fields are updated by predictive eye signals (Equations 64–66) that are derived from the surface contour map (Equation 45), as illustrated in the remainder of **Figure 14**. Additionally, the binocular boundaries of the attended object (the soccer ball) are strengthened by top-down feedback from the surface contour map (Equation 45) via gain fields (Equation 48). Thus, in **Figure 14A** it can be observed that there is an increase in summed activity of all the binocular boundaries by predictive buildup of the boundary gain fields acting on the monocular gain fields (their dynamics are explained in **Figures 14C–F**). Enhanced activity after the initial buildup for the invariant binocular boundaries of the attended surface (soccer ball) is maintained by its surface contour feedback (see **Figure 15** for illustration).

Figures 14B–F show a blown-up time scale (note the finer time scale) of these boundary dynamics achieved by a combination of the gain field activities and how they correlate with gain field predictive dynamics during the duration of the saccade. **Figure 14B** shows the temporal profile of the invariant binocular boundaries before, during, and after the eye movement from target position "*1*" to "*2*." This corresponds to the activity of the binocular boundaries shown in the gray dotted box in **Figure 14A**. Note the buildup and maintenance of the fused binocular boundary activity even before the eye movement (Equation 66) to the target position is completed, which only ends after 180 ms.

The invariant binocular boundaries B^b (Equation 33) are fused from invariant monocular boundaries $B^{l/r}_{ij}$ (Equation 26) that are derived from the retinotopic monocular boundaries $R^{l/r}_{ij}$ (Equation 22). This transformation from retinotopic to invariant monocular boundaries is achieved through predictive remapping by boundary gain fields (Equations 28–32), which are subsequently fused to yield the binocular boundaries (Equation 33). In **Figures 14C–F**, only the left monocular ON channel predictive remapping activities are presented. The summed activation patterns for the right monocular ON/OFF channels are exactly the same as that of the left images. In **Figures 14D–F**, the horizontal gray dashed lines are drawn to show how predictive remapping enhances the activities from before the eye movement to the target position.

Figure 14C plots the summed temporal activity of the eye position signal's P (Equation 66) gain modulation, defined as $\sum_{klij} P_{ij} E^{PI}_{klij}$ [in Equation (28)]. This modulates the boundary gain field in order to achieve predictive remapping of the invariant monocular boundary (see **Figure 3**). Only one target position is active at any given time and it can be observed that during the period of eye movement, there is a gradual buildup of this activity. Before the eye movement to a target position derived

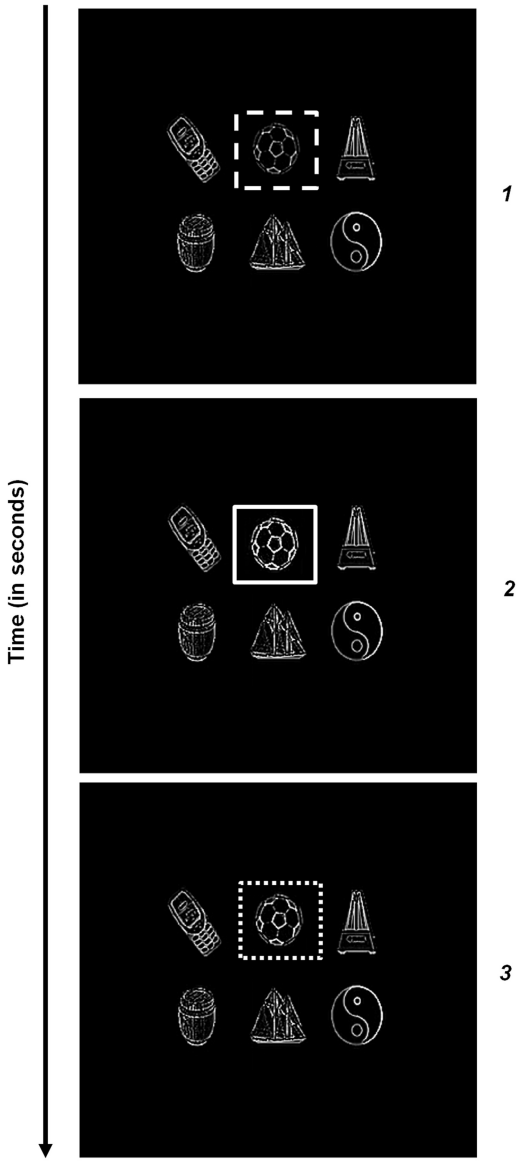


FIGURE 15 | Snapshot of fused invariant binocular boundaries B^b (Equation 33) of all the objects after saccades to target positions within the attended soccer ball. Only ON channel invariant binocular boundaries are shown. Following the paradigm in **Figure 10**, and the temporal profile in **Figure 14A**, the corresponding fused binocular boundaries are shown after the shift in attention to target position "*1*" followed by saccades to target positions "*2*," and "*3*" within the soccer ball. All the binocular boundaries are maintained in head-centered coordinates. The activities of the fused soccer ball boundaries are enhanced ("*1*," dashed box; "*2*," solid box; and "*3*," dotted box) as saccades are made to the corresponding target positions. Binocular boundaries of unattended objects remain fused as well. See Section 4.4 for details.

from the salient features is completed, the modulation from the predictive target position signal ensures that the invariant monocular boundaries are remapped to maintain the fusion of the binocular boundaries. The activity of this component is maintained at that level until the next eye movement occurs (here from target position "*2*" to "*3*").

The temporal evolution of the summed boundary gain field activity G^{R^l} (Equation 28) as $\sum_{klij} G^{R^l}_{klij}$, responsible for predictive remapping of the invariant monocular boundaries, is presented in **Figure 14D**. These boundary gain fields are modulated by the bottom-up inputs from retinotopic monocular boundaries (Equation 22), the target eye position signal (Equation 66), and feedback from the invariant monocular boundaries (Equation 26). These gain fields in turn modulate and predictively remap the invariant monocular boundaries (Equation 26) as well as the retinotopic monocular boundaries [Equation (22), also see **Figure 3**]. In **Figure 14D**, it can be observed that during the eye movement, there is a predictive buildup of the gain field activity. At the end of the eye movement, the overall gain field activity is enhanced from the initial value as marked by the dashed gray line. The transient increase in activity followed by plateauing is caused by a combination of top-down feedback from the invariant monocular boundaries and the bottom-up retinotopic monocular boundaries.

Figure 14E plots the summed temporal activity of the invariant left monocular boundaries' B^l (Equation 26) gain modulation expressed as $\sum_{klij} B^l_{ij} E^{BI}_{klij}$ (in Equation 28). Again there is a predictive buildup of this component and, after the transient activation, the activity plateaus. This transient activation is a combination of feedforward retinotopic inputs via the gain fields, followed by modulatory feedback from the fused invariant binocular boundaries to the invariant monocular boundaries. The gray horizontal line clearly shows an enhanced activation of the invariant monocular activation from its initial value before the saccade.

Figure 14F plots the summed temporal activity of the retinotopic left monocular boundaries' R^l (Equation 22) gain modulation $\sum_{klij} R^l_{ij} E^{RI}_{klij}$ [in Equation (28)]. During the eye movement to the target position "*2*," there is a buildup of this activity, followed by a transient activity before plateauing. The transient activity is caused by feedback from the invariant left monocular boundary via the boundary gain fields. The invariant left monocular boundaries in turn are modulated by invariant binocular boundaries (**Figure 3** and Equation 26). Thus, even before an eye movement is completed to the target position, the boundary gain fields predictively remap the invariant monocular boundaries. These invariant monocular boundaries are fused to yield invariant binocular boundaries, in which the binocular boundaries of the attended object are further strengthened by top-down feedback from their surface contour signals.

Figure 15 shows snapshots of activation profiles of the invariant fused binocular boundaries after a saccade occurs to those target positions ("*1*," dashed; "*2*," plain; and "*3*," dotted box) as shown in **Figure 10**. Again for convenience, only the ON channel invariant binocular boundaries are shown. It can be observed from the three snapshots in **Figure 15** that the binocular boundaries of all the six objects in the scene remain fused after every subsequent eye movement to the three different target positions within the soccer ball. They are also maintained in head-centered coordinates throughout the time when eye movements

are made to target positions within the soccer ball. Further, the activity of binocular boundaries of the attended soccer ball surface is enhanced with every eye movement due to surface contour feedback.

5. MATHEMATICAL EQUATIONS AND PARAMETERS

Unless specified otherwise, the equations are all solved dynamically. Symbol I is the input image and I_{ij} is the value of the input image in the visual field at position (i, j). The dynamic range of inputs I_{ij} is [0, 1]. The superscripts l/r are used to denote the boundary/surface processing in the left or right eyes, respectively. The superscripts $+/-$ are used to denote ON and OFF processing, respectively. The equations and parameters used for monocular cells that are responsive to the left or right eyes, and for ON and OFF cells are the same in the simulations, unless specified otherwise. The binocular cells/networks have a b superscript. The simulations are shown for a single depth with allelotropic shifts of $s = +3^o$ where the neurons are tuned for far disparity. The image input I_{ij} at position (i, j) gives rise to monocular inputs to the left and right eyes equal to $I^l_{(i+s)j}$, and $I^r_{(i-s)j}$, respectively, for all i and j that project to the retina. The simulations were carried out in MathWorks (R) MATLAB R2012a (TM) on a Linux GNOME x64 bit machine with Intel Quad-Core (TM)/3.10 GHz/7.7 GB of RAM.

5.1. RETINAL ADAPTATION

The retinal equations have been adapted from the aFILM model of Grossberg and Hong (2006). The potential $\phi^{l/r}_{ij}$ at position (i, j) of the outer segment of the retinal photoreceptor is simulated by the equation:

$$\phi^{l/r}_{ij}(t) = I^{l/r}_{ij} z^{l/r}_{ij}(t), \tag{1}$$

where $I^{l/r}_{ij}$ is the monocular input image and $z^{l/r}_{ij}(t)$ is a habituative gate that realizes an automatic gain control term simulating negative feedback mediated by Ca^{2+} ions, among others. It is defined as follows:

$$\frac{dz^{l/r}_{ij}}{dt} = \left(B_Z - z^{l/r}_{ij}\right) - z^{l/r}_{ij}\left(C_I I^{l/r}_{ij} + C_{I^*} I^*\right), \tag{2}$$

where $B_Z = 5$ is the asymptote to which $z^{l/r}_{ij}(t)$ accumulates, or recovers, in the absence of input, and the term $z^{l/r}_{ij}(C_I I^{l/r}_{ij} + C_{I^*} I^*)$ describes the inactivation of $z^{l/r}_{ij}$ by the present input, $I^{l/r}_{ij}$, and by a spatial average, I^*, of all the inputs that approximates the effect of recent image scanning by sequences of eye movements. Parameters $C_I = 2$, and $C_{I^*} = 6$. Solving Equations (1, 2) at equilibrium yields the equilibrium potential:

$$\phi^{l/r}_{ij} = \frac{B_Z I^{l/r}_{ij}}{1 + C_I I^{l/r}_{ij} + C_{I^*} I^*}. \tag{3}$$

In the simulations, $I^* = 0.5$ best approximates the effect of recent image scans.

The inner segment of the photoreceptor receives the signal $\phi^{l/r}_{ij}$ from the outer segment and gets feedback $H^{l/r}_{ij}$ from the horizontal cells (HC) at position (i, j). HC modulation of the output of the inner segment of the photoreceptor is modeled by:

$$\Phi^{l/r}_{ij} = \frac{\phi^{l/r}_{ij}}{B_h e^{H^{l/r}_{ij}}\left(B_s - \phi^{l/r}_{ij}\right) + 1}, \tag{4}$$

where $B_h = 0.05$ is a small constant, and $B_s = B_z \,/\, C_I = 2.5$. This constant value of B_s ensures that perfect shifts (viz., adaptation) of the $\log(I^{l/r}_{ij}) - \Phi^{l/r}_{ij}$ curve occur as $H^{l/r}_{ij}$ is varied. For more details, see Grossberg and Hong (2006). Many increasing functions of $H^{l/r}_{ij}$ will generate the shift property of $\Phi^{l/r}_{ij}$ as a function of $\log(I^{l/r}_{ij})$. Function $f(H_{ij}) = B_h e^{H^{l/r}_{ij}}$ was chosen because $e^{H^{l/r}_{ij}}$ makes the sensitivity curve shift in an accelerating manner with increasing $H^{l/r}_{ij}$, where $H^{l/r}_{ij}$ is the sigmoid output of the HC at (i, j) in response to its potential $h^{l/r}_{ij}$:

$$H^{l/r}_{ij} = \frac{a_H \left[h^{l/r}_{ij}\right]^2}{b^2_H + \left[h^{l/r}_{ij}\right]^2}, \tag{5}$$

where $a_H = 6$ and $b_H = 0.1$.

The potential of an HC connected to its neighbors through gap junctions is defined as follows.

$$\frac{dh^{l/r}_{ij}}{dt} = -h^{l/r}_{ij} + \sum_{pq \in N^H_{ij}} \Psi^{l/r}_{pqij}\left(h^{l/r}_{pq} - h^{l/r}_{ij}\right) + \Phi^{l/r}_{ij}, \tag{6}$$

where $\Psi^{l/r}_{pqij}$ is the permeability between cells at (i, j) and (p, q); namely:

$$\Psi^{l/r}_{pqij} = \frac{-1}{1 + exp\left[-\left(\left|\Phi^{l/r}_{ij} - \Phi^{l/r}_{pq}\right| - \beta_p\right)/\mu_p\right]} + 1, \tag{7}$$

where $\beta_p = 0.01$, and $\mu_p = 0.002$, and N^H_{ij} is the neighborhood of cells to which the HC at position (i, j) is connected:

$$N^H_{ij} = \left\{(p, q) : \sqrt{(p - i)^2 + (q - j)^2} \leq 13\right\}. \tag{8}$$

5.2. LGN POLARITY-SENSITIVE ON AND OFF CELLS

5.2.1. Center-surround processing

The retinally adapted signal $\Phi^{l/r}_{ij}$ is processed by on-center off-surround (ON cells) and off-center on-surround (OFF) cells that obey the membrane, or shunting, equations of neurophysiology. The activity $x^{l/r,+}_{ij}$ of the on-center off-surround (ON) network that receives input signals $\Phi^{l/r}_{ij}$ (Equation 4) from the inner

segment of the photoreceptors is defined as follows:

$$\frac{dx_{ij}^{l/r,+}}{dt} = -x_{ij}^{l/r,+} + \left(1 - x_{ij}^{l/r,+}\right)\left(0.6\Phi_{ij}^{l/r}\right) - \left(x_{ij}^{l/r,+} + 1\right)E_{ij}^{l/r} + \Theta^{l/r,+}. \quad (9)$$

In Equation (9), the term $0.6\Phi_{ij}^{l/r}$ is the on-center input, $E_{ij}^{l/r}$ is the off-surround input, and $\Theta^{l/r,+}$ is the resting activity. The off-surround obeys:

$$E_{ij}^{l/r} = \frac{0.6\left(\sum_{(p,q)\in N_{ij}^{E}} \Phi_{pq}^{l/r} E_{pqij}^{l/r}\right)}{\sum_{(p,q)\in N_{ij}^{E}} E_{pqij}^{l/r}}, \quad (10)$$

where N_{ij}^{E} is the off-surround neighborhood to which the cell at (i, j) is connected:

$$N_{ij}^{E} = \left\{(p,q) : \sqrt{(p-i)^2 + (q-j)^2} \le 6\right\}, \quad (11)$$

and $E_{pqij}^{l/r}$ is the inhibitory off-surround kernel:

$$E_{pqij}^{l/r} = \frac{0.6e^{\left(-\frac{(p-i)^2+(q-j)^2}{16}\right)}}{\sum_{(p,q)\in N_{ij}^{E}} e^{\left(-\frac{p^2+q^2}{16}\right)}}, \quad (12)$$

which is normalized by the terms in the denominator. With this LGN ON-center/OFF-surround processing, the single and double-opponent LGN polarity-sensitive cells can be computed as follows.

5.2.2. ON/OFF channels and double-opponent cells

As defined in Grossberg et al. (1995), the equilibrium, ON-cell activities of Equation (9) are thresholded to yield the output signals:

$$x_{ij}^{l/r,+} = \left[\frac{\Theta^{l/r,+} + 0.6\Phi_{ij}^{l/r} - E_{ij}^{l/r}}{1 + 0.6\Phi_{ij}^{l/r} + E_{ij}^{l/r}}\right]^{+}. \quad (13)$$

The corresponding equilibrium outputs of the off-center on-surround (OFF) network are:

$$x_{ij}^{l/r,-} = \left[\frac{\Theta^{l/r,-} + E_{ij}^{l/r} - 0.6\Phi_{ij}^{l/r}}{1 + 0.6\Phi_{ij}^{l/r} + E_{ij}^{l/r}}\right]^{+}. \quad (14)$$

By (14), the on-center and off-surround of an OFF cell is the off-surround and the on-center of the corresponding ON cell, respectively. The rest level parameters Θ^{+} and Θ^{-} were chosen with $\Theta^{-} > \Theta^{+}$ — in particular, $\Theta^{l/r,+} = 1.5$ and $\Theta^{l/r,-} = 4.5$, which allows the OFF cells to be tonically active in the presence of uniform inputs, including in the dark. The inhibitory interactions that define the ON and OFF cells in Equations (13, 14) are computed across space among other ON and OFF cells, respectively. In contrast, the next processing stage of, double-opponent cells is defined by subtracting the ON and OFF cell output output signals at each position, and then thresholding the result:

Double-opponent ON-cell:

$$X_{ij}^{l/r,+} = \left[x_{ij}^{l/r,+} - x_{ij}^{l/r,-}\right]^{+}. \quad (15)$$

Double-opponent OFF-cell:

$$X_{ij}^{l/r,-} = \left[x_{ij}^{l/r,-} - x_{ij}^{l/r,+}\right]^{+}. \quad (16)$$

5.3. BOUNDARY PROCESSING

5.3.1. Simple cells

The simple cell activities $T_{ij\theta}^{l/r}$ in model cortical area V1 receive their inputs from double-opponent LGN cells and are computed as in Raizada and Grossberg (2003). At each position (i, j), and for each of the four orientations $\theta = \{0°, 45°, 90°, 135°\}$, a Difference-of-Offset-Gaussian (DOOG) kernel was used to compute each simple cell's orientationally-tuned ON and OFF subregions. In response to an oriented contrast edge in an input image, a suitably oriented simple cell of correct polarity will have its ON subfield stimulated by a luminance increment and its OFF subfield stimulated by a luminance decrement. The simple cell activity $T_{ij\theta}^{l/r}$ for a given orientation θ, is the rectified sum of activities of each subfield, minus their difference:

$$T_{ij\theta}^{l/r} = \vartheta\left[U_{ij\theta}^{l/r} + V_{ij\theta}^{l/r} - \left|U_{ij\theta}^{l/r} - V_{ij\theta}^{l/r}\right|\right]^{+}, \quad (17)$$

where $\vartheta = 6$, and the term $U_{ij\theta}^{l/r}$ and $V_{ij\theta}^{l/r}$ in Equation (17) represent the ON and OFF subregions, respectively:

$$U_{ij\theta}^{l/r} = \sum_{mn}\left(\left[X_{mn}^{l/r,+}\right]^{+} - \left[X_{mn}^{l/r,-}\right]^{+}\right)\left[D_{mnij\theta}^{l/r}\right]^{+} \quad (18)$$

and

$$V_{ij\theta}^{l/r} = \sum_{mn}\left(\left[X_{mn}^{l/r,-}\right]^{+} - \left[X_{mn}^{l/r,+}\right]^{+}\right)\left[-D_{mnij\theta}^{l/r}\right]^{+}, \quad (19)$$

and $D_{mnij\theta}^{l/r}$ is the DOOG kernel:

$$D_{mnij\theta}^{l/r} = \frac{1}{2\pi\sigma_D^2}\left[\begin{array}{c}\exp\left(-\frac{(m-i+\delta\cos\theta)^2+(n-j+\delta\sin\theta)^2}{2\sigma_D^2}\right) \\ - \\ \exp\left(-\frac{(m-i-\delta\cos\theta)^2+(n-j-\delta\sin\theta)^2}{2\sigma_D^2}\right)\end{array}\right] \quad (20)$$

in which $\sigma_D = 0.5$ is the standard deviation of the kernel width.

5.3.2. Complex cells

The model boundary is not used to simulate any polarity-specific properties. Thus, for simplicity, the simple cell responses are

pooled across all four orientations to define the complex cell activities and output signals:

$$Z_{ij}^{l/r} = 0.25 \sum_{\theta} T_{ij\theta}^{l/r} \tag{21}$$

5.3.3. Monocular retinotopic boundaries

The monocular retinotopic boundary activities $R_{ij}^{l/r}$ (**Figure 2**) obey:

$$\frac{dR_{ij}^{l/r}}{dt} = -a_R R_{ij}^{l/r} + \left(b_R - R_{ij}^{l/r}\right)\left(Z_{ij}^{l/r} + c\sum_{klij} h\left(G_{klij}^{R^{l/r}}\right) E_{klij}^{IR}\right) - \left(R_{ij}^{l/r} + d_R\right)\left(\sum_{pq} Z_{pq}^{l/r} + d\sum_{klij} h\left(G_{klij}^{R^{l/r}}\right) E_{klij}^{IR}\right), \tag{22}$$

where the decay rate $\alpha_R = 5$, the shunting excitatory saturation activity $b_R = 10$, and the shunting inhibitory saturation activity $d_R = 2$. A bottom-up on-center $Z_{ij}^{l/r}$ off-surround $\sum_{pq} Z_{pq}^{l/r}$ network of inputs come from complex cell outputs $Z_{ij}^{l/r}$. Retinotopic monocular boundaries also receive top-down on-center off-surround signals $\sum_{klij} h(G_{klij}^{R^{l/r}}) E_{klij}^{IR}$ from invariant, or head-centered, monocular boundaries that are first transformed by gain fields. Functions $G_{klij}^{R^{l/r}}$ are the top-down gain field output signals from position (k, l) to (i, j), and E_{klij}^{IR} are the top-down connection weights from this gain field to the retinotopic boundary cells. These gain field functions and weights are defined in Equations (28–32). The feedback signal function h is threshold-linear:

$$h(a) = [a - 0.2]^{+}. \tag{23}$$

These top-down gain field signals are multiplied in Equation (22) by excitatory and inhibitory gains $c = 10$ and $d = 2$, respectively.

5.3.4. Invariant monocular boundaries

The invariant monocular boundary activities $B_{ij}^{l/r}$ receive bottom-up inputs via gain fields $G_{klij}^{R^{l/r}}$ that transform the retinotopic monocular boundaries into invariant monocular boundaries (**Figure 3**). Before an eye movement occurs, the dark-light monocular invariant boundary activity is defined to equal the corresponding retinotopic monocular boundary activity:

$$B_{ij}^{l/r,+} = R_{ij}^{l/r}, \tag{24}$$

and the light-dark monocular invariant boundary activity is defined as

$$B_{ij}^{l/r,-} = \begin{cases} [1 - B_{ij}^{l/r,+}]^{+} & \text{if } B_{ij}^{l/r,+} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

As eye movements occur, the invariant monocular boundaries receive retinotopic monocular boundary inputs (Equation 22) through the gain fields $G_{klij}^{R^{l/r}}$ described in Equations (28–32). Their left (L) $B_{ij}^{l,+/-}$ and right (R) $B_{ij}^{r,+/-}$ activities are defined as follows:

$$\frac{dB_{ij}^{l/r,+/-}}{dt} = -a_b B_{ij}^{l/r,+/-} + \left(1 - B_{ij}^{l/r,+/-}\right)\left(f\left(B_{ij}^{l/r,+/-}\right) + p_b \sum_{klij} h\left(G_{klij}^{R^{l/r}}\right) E_{klij}^{IB} + \lambda h\left(B_{ij}^{b,+/-}\right)\right) - B_{ij}^{l/r,+/-} \sum_{kl}\left(f\left(B_{kl}^{l/r,+/-}\right) + q_b \sum_{klij} h\left(G_{klij}^{R^{l/r}}\right) E_{klij}^{IB} + h\left(B_{kl}^{b,+/-}\right)\right), \tag{26}$$

where $a_b = 20$ is the decay rate, and

$$f(a) = \frac{a^2}{4 + 2a^2} \tag{27}$$

is the feedback sigmoid signal function that transforms the activities of the invariant monocular boundaries into a recurrent on-center off-surround network of feedback signals that maintain the persistent activity of the invariant boundaries in the network. Parameters $p_b = 16$ and $q_b = 16$ are excitatory and inhibitory gains that multiply the bottom-up excitatory and inhibitory signals, respectively, from the gain fields. Invariant monocular boundaries receive the same bottom-up excitatory and inhibitory signals $\sum_{klij} h(G_{klij}^{R^{l/r}}) E_{klij}^{IB}$ from retinotopic monocular boundaries that are first transformed by gain fields. Functions $G_{klij}^{R^{l/r}}$ are the bottom-up gain field output signals from position (k, l) to (i, j), and E_{klij}^{IB} are the bottom-up connection weights from this gain field to the retinotopic boundary cells. These gain field functions and weights are defined in (Equations 28–32). Parameter $\lambda = 1.5$ is a gain constant that multiplies the excitatory feedback signal $h(B_{ij}^{b,+/-})$ from the invariant binocular boundary $B_{ij}^{b,+/-}$ (Equation 33). The inhibitory feedback signal $h(B_{ij}^{b,+/-})$ has a gain of 1. Signal function $h$is the threshold-linear function defined in Equation (23).

5.3.5. Boundary gain fields

Boundary gain field activities $G_{klij}^{R^{l/r}}$ receive inputs from retinotopic monocular boundary signals $R_{ij}^{l/r}$ (Equation 22), predictive eye position signals P_{ij} (Equation 66), and invariant monocular boundary signals $B_{ij}^{l/r,+/-}$ (Equation 26 and **Figure 3**) in order to activate and maintain the invariant monocular boundaries $B_{ij}^{l/r,+/-}$ (Equation 26):

$$\frac{dG_{klij}^{R^{l/r}}}{dt} = \left(1 - G_{klij}^{R^{l/r}}\right)$$

$$\left(\sum_{ij} R_{ij}^{l/r} E_{klij}^{RI} + \sum_{ij} P_{ij} E_{klij}^{PI} + \sum_{ij} B_{ij}^{l/r,+/-} E_{klij}^{BI}\right) - (G_{klij}^{R^{l/r}} + 0.15) \sum_{klij} G_{klij}^{R^{l/r}}. \quad (28)$$

Gaussian kernels E_{klij}^{RI}, E_{klij}^{PI}, and E_{klij}^{BI} represent the gain field weights from each of these input sources:

$$E_{klij}^{RI} = \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma_{G_R^R}^2}\right); \sigma_{G_R^{RI}} = 2 \quad (29)$$

$$E_{klij}^{PI} = \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma_{G_R^P}^2}\right); \sigma_{G_R^{PI}} = 2 \quad (30)$$

$$E_{klij}^{BI} = \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma_{G_R^B}^2}\right); \sigma_{G_R^{BI}} = 3.5 \quad (31)$$

The top-down and bottom-up gain field weights are the same. Separate copies of these weights are defined for conceptual clarity:

$$E_{klij}^{BI} = E_{klij}^{IB}; \; E_{klij}^{PI} = E_{klij}^{IP}; \; E_{klij}^{RI} = E_{klij}^{IR} \quad (32)$$

5.3.6. Invariant binocular boundaries

The model considers how a 2D planar surface that is viewed in 3D is binocularly fused and how its 3D boundaries and surfaces are maintained during eye movements. It assumes a fixed, but otherwise arbitrary, binocular disparity of the left and right eye monocular boundaries corresponding to the object's image contours. The output signals $B_{ij}^{l/r}$ from the left and the right invariant monocular boundaries (**Figure 3** and Equation 26) are binocularly fused as follows to create the invariant binocular boundary activities B_{ij}^{b}:

$$\frac{dB_{ij}^{b,+/-}}{dt} = -\gamma_1 B_{ij}^{b,+/-} + \left(1 - B_{ij}^{b,+/-}\right) \left(\left[B_{(i+s)j}^{l,+/-} - \kappa\right]^+ + \left[B_{(i-s)j}^{r,+/-} - \kappa\right]^+\right) + \left(1 + 3.2 \sum_{klij} h\left(G_{klij}^{C}\right) J_{klij}^{CB}\right) - \alpha\left(\left[O_{ij}^{l,+/-}\right]^+ + \left[O_{ij}^{l,-/+}\right] + \left[O_{ij}^{r,+/-}\right]^+ + \left[O_{ij}^{r,-/+}\right]^+\right), \quad (33)$$

where $\gamma_1 = 0.1$ is the rate of decay of the membrane potential. In Equation (33), the binocular disparity is assumed to cause allelotropically shifted monocular boundary signals $B_{(i+s)j}^{l,+/-}$ and $B_{(i-s)j}^{r,+/-}$, with shift s, which are binocularly fused via the sum $[B_{(i+s)j}^{l,+/-} - \kappa]^+ + [B_{(i-s)j}^{r,+/-} - \kappa]^+$, where $\kappa = 0.4$ is the boundary signal threshold. The selectivity of binocular fusion is achieved by balancing these excitatory terms against the sum of inhibitory signals $\alpha([O_{ij}^{l,+/-}]^+ + [O_{ij}^{l,-/+}] + [O_{ij}^{r,+/-}]^+ + [O_{ij}^{r,-/+}]^+)$, where $\alpha = 7.2$ is the inhibitory gain. Together, these balanced excitatory and inhibitory terms help to realize the *obligate property* (Poggio, 1991; Grossberg and Howe, 2003), whereby these binocular cells respond only to left and right eye inputs of approximately equal size, one of the important prerequisites for solving the *correspondence problem* of binocular vision (Howard and Rogers, 1995, pp. 42, 43).

The left $O_{ij}^{l,+/-}$ and right $O_{ij}^{r,+/-}$ inhibitory interneuron cell activities that ensure the obligate property are defined by:

$$\frac{dO_{ij}^{l,+/-}}{dt} = -\gamma_2 O_{ij}^{l,+/-} + \left[B_{(i+s)j}^{l,+/-} - \kappa\right]^+ - \beta\left(\left[O_{ij}^{r,+/-}\right]^+ + \left[O_{ij}^{r,-/+}\right]^+ + \left[O_{ij}^{l,-/+}\right]^+\right) \quad (34)$$

and

$$\frac{dO_{ij}^{r,+/-}}{dt} = -\gamma_2 O_{ij}^{r,+/-} + \left[B_{(i-s)j}^{r,+/-} - \kappa\right]^+ - \beta\left(\left[O_{ij}^{l,+/-}\right]^+ + \left[O_{ij}^{l,-/+}\right]^+ + \left[O_{ij}^{r,-/+}\right]^+\right), \quad (35)$$

where the decay rate $\gamma_2 = 4.5$; $[B_{(i+s)j}^{l/r,+/-} - \kappa]^+$ are the excitatory signals from the monocular invariant boundaries that drive the inhibitory interneurons; and $\beta = 4$ is the gain of the recurrent inhibitory signals $\beta([O_{ij}^{r,+/-}]^+ + [O_{ij}^{r,-/+}]^+ + [O_{ij}^{l,-/+}]^+)$ among the inhibitory interneurons that are needed to ensure the obligate property (Grossberg and Howe, 2003). In Equations (33–35), the subscript s denotes the allelotropic, or positional, shift between the left and the right eyes that depends on the disparity to which the model neurons are tuned. In the simulations, results are shown for an allelotropic shift of $s = +3^o$ to illustrate neurons that are tuned to a far disparity. The simulations also work for other binocular disparities and the allelotropic shifts that they induce. The obligate cell theorem from Grossberg and Howe (2003) was used to solve Equations 33–35 at equilibrium to speed up the simulations.

The invariant binocular boundaries in Equation (33) also receive feedback $\sum_{klij} h(G_{klij}^{C}) J_{klij}^{CB}$ from the surface contour signals (Equation 45) that are generated from filled-in surfaces to their inducing boundaries. These surface contour signals enhance the corresponding closed boundaries, a crucial step in figure-ground separation whereby partially occluded object surfaces are separated in depth (Grossberg, 1994; Kelly and Grossberg, 2000). Since the fused binocular boundary is invariant, and thus computed in head-centered coordinates, but the surface contour is computed in retinotopic coordinates, the feedback from the surface contour is mediated through a gain field G^C to execute this coordinate change (**Figure 4**). The activity of the surface contour gain field G^C and the gain field kernel J^{CB} are defined in Equations (48, 49).

5.4. SURFACE PROCESSING

5.4.1. Monocular retinotopic surface capture and filling-in

The monocular retinotopic surface filling-in activities $S_{ij}^{l/r,+/-}$ are computed from the brightness information that is driven by monocular retinotopic double-opponent ON and OFF cell activities $X_{ij}^{l/r,+/-}$ (**Figure 2** and Equations 15, 16):

$$\frac{dS_{ij}^{l/r,+/-}}{dt} = -40S_{ij}^{l/r,+/-} + \sum_{pq \in N_{ij}} P_{pqij}^{l/r}\left(S_{pq}^{l/r,+/-} - S_{ij}^{l/r,+/-}\right) + X_{ij}^{l/r,+/-}. \quad (36)$$

The activities $S_{ij}^{l/r,+/-}$ diffuse via nearest-neighbor interactions via term $\sum_{pq \in N_{ij}} P_{pqij}^{l/r}(S_{pq}^{l/r,+/-} - S_{ij}^{l/r,+/-})$, where N_{ij} is the set of nearest neighbors around cell (i, j), and the permeability coefficients

$$P_{pqij}^{l/r} = \frac{10^4}{0.01 + 20\left(K_{pq}^{b,+/-} + K_{ij}^{b,+/-}\right)} \quad (37)$$

are determined by binocular boundary gating signals $K_{pq}^{b,+/-}$ and $K_{ij}^{b,+/-}$ at positions (p, q) and (i, j), respectively. Since the binocular boundaries are computed in head-centered co-ordinates, whereas the monocular surfaces are computed in retinotopic coordinates, the boundary gating signals need to also be computed in retinotopic coordinates. This is accomplished by converting the binocular boundaries into retinotopic coordinates (**Figure 4**) using a predictive gain field:

$$K_{ij}^{b,+/-} = \sum_{kl} h\left(G_{klij}^{S,+/-}\right) Q_{klij}^{BS} \quad (38)$$

that is defined in Equations (42–44).

5.4.2. Binocular retinotopic surface capture and filling in

The binocular surface representations are preserved during eye movements, even though they are computed in retinotopic coordinates, due to the action of predictive gain fields that control the binocular filling-in process. In particular, the retinotopic surface filling-in activities $S_{ij}^{b,+/-}$ are activated by the rectified sum $\left[S_{ij}^{l,+/-}\right]^+ + \left[S_{ij}^{r,+/-}\right]^+$ of the monocular retinotopic surface activities captured by the invariant binocular boundary (Equation 36) corresponding to the same retinotopic position (i, j):

$$\frac{dS_{ij}^{b,+/-}}{dt} = -28S_{ij}^{b,+/-} + \sum_{pq \in N_{ij}} N_{pqij}\left(S_{pq}^{b,+/-} - S_{ij}^{b,+/-}\right) + \left[S_{ij}^{l,+/-}\right]^+ + \left[S_{ij}^{r,+/-}\right]^+ + 9\sum_{kl} h\left(G_{klij}^{A}\right) M_{klij}^{IS} \quad (39)$$

The binocular surface activities undergo diffusion $\sum_{pq \in N_{ij}} N_{pqij}^{(} S_{pq}^{b,+/-} - S_{ij}^{b,+/-})$ in response to these input signals. The diffusion takes place among their nearest-neighbor cells N_{ij}, whose permeabilities

$$N_{pqij} = \frac{10^4}{0.01 + 20\left(K_{pq}^{b,+/-} + K_{ij}^{b,+/-}\right)} \quad (40)$$

are determined by binocular boundary gating signals $K_{pq}^{b,+/-}$ and $K_{ij}^{b,+/-}$ at positions (p, q) and (i, j), respectively. Similar to the monocular surfaces, binocular surfaces are as well computed in retinotopic coordinates. However, the binocular boundaries are computed in head-centered co-ordinates and thus the boundary gating signals need to also be computed in retinotopic coordinates. This is accomplished by converting the binocular boundaries into retinotopic coordinates (**Figure 4**) using a predictive gain field. The retinotopic boundary gating signals $K_{ij}^{b,+/-}$ were defined earlier in Equation (38). The gain fields for accomplishing this conversion are defined in Equations (42–44).

The binocular surface representation also receives top-down excitatory feedback from spatial attention (**Figure 4**) to induce and maintain a surface-shroud resonance. Spatial attention is in head-centered coordinates, whereas the binocular surface representation is retinotopic. Hence the spatial attentional feedback $\sum_{kl} h(G_{klij}^{A}) M_{klij}^{IS}$ in Equation (39) is also computed in retinotopic coordinates using the predictive gain field G_{klij}^{A} that is defined by Equations (56–60).

$S_{ij}^{b,+/-}$ is the fused binocular surface representation that is maintained in retinotopic coordinates despite eye movements across the visual scene. These ON and OFF binocular FIDO activities are rectified and combined to yield the final binocular surface percept:

$$S^b = \left[S^{b,+}\right]^+ + \left[S^{b,-}\right]^+ \quad (41)$$

In the simulation results, S^b is shown as the final binocular surface percept. This rectified summation of the ON and OFF domains enables surface-shroud resonance by attracting spatial attention on both light and dark filled-in surfaces. However, all the different representations, not just of brightness information, but also of brightness and color in depth, can be held as separate representations. The ensemble of all such parallel representations is what is learned, recognized, and categorized as belonging to a particular object in the What stream.

5.4.3. Surface gain fields

The gain fields that enable binocular invariant boundaries to gate binocular and monocular surface percepts are defined as follows. Surface gain fields receive inputs from binocular invariant boundaries and predictive eye position signals (**Figure 4**):

$$\frac{dG_{klij}^{S,+/-}}{dt} = \left(1 - G_{klij}^{S,+/-}\right)\left(\sum_{ij} B_{ij}^{b,+/-} Q_{klij}^{BS} + \sum_{ij} P_{ij} Q_{klij}^{PS}\right) - \left(G_{klij}^{S,+/-} + 0.37\right)\sum_{klij} G_{klij}^{S,+/-} \quad (42)$$

where $B_{ij}^{b,+/-}$ is the invariant binocular boundary activity defined in (Equation 33), and P_{ij} is the predictive eye position described in Equation (66). Gaussian kernels Q_{klij}^{BS} and Q_{klij}^{PS} multiply the invariant binocular boundary signals and the eye position signals, respectively:

$$Q_{klij}^{PS} = \exp\left(-\frac{(k-i)^2+(l-j)^2}{2\sigma_{G_S^{PS}}^2}\right);\ \sigma_{G_S^{PS}} = 1.2 \quad (43)$$

$$Q_{klij}^{BS} = \exp\left(-\frac{(k-i)^2+(l-j)^2}{2\sigma_{G_S^{BS}}^2}\right);\ \sigma_{G_S^{BS}} = 1.4 \quad (44)$$

5.4.4. Surface contour activity

The binocular surface activities S_{pq}^b (Equation 41) are contrast-enhanced by on-center off-surround output networks to generate surface contour signals that modulate the invariant binocular boundaries (**Figure 3** and Equation 33) and, through them, the corresponding retinotopic boundaries (Equation 22). Surface contour signals (**Figure 4**) are also used to determine the predictive target position signal (Equation 66) that maintains the stability of boundaries, surfaces, and attentional shrouds in head-centered coordinates via gain fields (**Figures 1**, **3**, **4**), even before the next eye movement is made, and to generate this eye movement signal. Surface contour signals occur only at positions corresponding to the boundary contours of the surface. The contour signals C_{ij} obey:

$$C_{ij} = \left[\frac{\sum_{pq} S_{pq}^b\left(\Lambda_{pqij}^+ - \Lambda_{pqij}^-\right)}{0.04 + \sum_{pq} S_{pq}^b\left(\Lambda_{pqij}^+ + \Lambda_{pqij}^-\right)}\right]^+ + \left[\frac{\sum_{pq} S_{pq}^b\left(\Lambda_{pqij}^- - \Lambda_{pqij}^+\right)}{0.04 + \sum_{pq} S_{pq}^b\left(\Lambda_{pqij}^+ + \Lambda_{pqij}^-\right)}\right]^+, \quad (45)$$

where Λ_{pqij}^+ and Λ_{pqij}^- are the contrast-enhancing S^b on-center and off-surround kernels, respectively:

$$\Lambda_{pqij}^+ = \frac{1}{3.61}\exp\left(-\frac{(p-i)^2+(q-j)^2}{2\sigma_{\Lambda^+}^2}\right);\ \sigma_{\Lambda^+} = 0.5 \quad (46)$$

$$\Lambda_{pqij}^- = \frac{1}{12.27}\exp\left(-\frac{(p-i)^2+(q-j)^2}{2\sigma_{\Lambda^+}^2}\right);\ \sigma_{\Lambda^-} = 2 \quad (47)$$

5.4.5. Gain fields from surface contour to invariant binocular boundary

Since the surface contour is in retinotopic coordinates and the fused binocular boundary that it modulates is in head-centered coordinates, a gain field G_{klij}^C transforms the input from surface contour to binocular boundary (**Figure 4**):

$$\frac{dG_{klij}^C}{dt} = \left(1.8 - G_{klij}^C\right)\left(\sum_{ij} C_{ij}J_{klij}^{CB} + \sum_{ij} P_{ij}J_{klij}^{PB} +\right) - \left(G_{klij}^C + 0.7\right)\sum_{klij} G_{klij}^C, \quad (48)$$

where C_{ij} is the surface contour activity defined in Equation (45), and P_{ij} is the predictive target position signal described in Equation (66). Terms J_{klij}^{CB}, and J_{klij}^{PB} in Equation (48) represent the Gaussian gain field kernels that transform the surface contour and the target position signals, respectively:

$$J_{klij}^{CB} = \exp\left(-\frac{(k-i)^2+(l-j)^2}{2\sigma_{G_C^{CB}}^2}\right);\ \sigma_{G_C^{CB}} = 2.6 \quad (49)$$

$$J_{klij}^{PB} = \exp\left(-\frac{(k-i)^2+(l-j)^2}{2\sigma_{G_C^{PB}}^2}\right);\ \sigma_{G_C^{CB}} = 1.2 \quad (50)$$

5.5. SPATIAL SHROUDS

5.5.1. Spatial attention activity

The spatial attention cell activities A_{ij} that support attentional shrouds obey:

$$\frac{1}{10}\frac{dA_{ij}}{dt} = -0.2A_{ij} + \left(2 - A_{ij}\right)\left(A_{ij}^I + \sum_{mn} g(A_{mn})\Omega_{mnij}^+\right)y_{ij}^A - A_{ij}\left(\sum_{mn}\left(A_{mn}^I + g\left(A_{mn}\right)\Omega_{mnij}\right) + C_{RESET}{}^{y^c}\right) \quad (51)$$

These cell activities receive bottom-up excitatory inputs A_{ij}^I from the corresponding attention interneurons (see Equation 55). They also receive recurrent on-center signals $\sum_{mn} g(A_{mn})\Omega_{mnij}^+$ and off-surround signals $g(A_{mn})\Omega_{mnij}^-$ from other spatial attention cells, where g is a sigmoid signal function that converts cell activities into output signals:

$$g(a) = \frac{7}{1+e^{-25a+11}}. \quad (52)$$

Kernels Ω_{mnij}^+, and Ω_{mnij}^- are the on-center and off-surround Gaussian weights, respectively, from position (m, n) to position (i, j):

$$\Omega_{mnij}^+ = 0.04\exp\left(-\frac{(m-i)^2+(n-j)^2}{2\sigma_{\Omega^+}^2}\right);\ \sigma_{\Omega^+} = 0.5 \quad (53)$$

$$\Omega_{mnij}^- = 2.2\exp\left(-\frac{(m-i)^2+(n-j)^2}{2\sigma_{\Omega^-}^2}\right);\ \sigma_{\Omega^-} = 100 \quad (54)$$

The excitatory inputs and recurrent signals in Equation (51) are multiplied by habituative attentional transmitter gates y_{ij}^A (Equation 61) that enable inhibition-of-return (IOR). The system also receives a parietal reset signal C_{RESET} (Equation 62)

that inhibits the currently active shroud. The reset signal C_{RESET} is multiplied by a habituative transmitter gate y^C (Equation 63) which ensures that the net reset signal $C_{RESET}y^C$ is transient.

5.5.2. Attentional interneuron cell activity

Attentional interneuronal activities A^I_{ij} input to the spatial attention cell activities in Equation (51), receive reciprocal top-down feedback from the spatial attention cells (**Figures 4**, **5**), and are themselves activated by bottom-up signals from the binocular filled-in surfaces (Equation 41) to form surface-shroud resonances:

$$\frac{dA^I_{ij}}{dt} = -0.9A^I_{ij} + 1.2\sum_{kl} h\left(G^A_{klij}\right) M^{IA}_{klij} + g\left(A_{ij}\right). \quad (55)$$

Because the binocular filled-in surfaces are computed in retinotopic coordinates, whereas the attentional shrouds are computed in head-center coordinates, gain fields are needed to transform their inputs between them. In Equation (55), $\sum_{kl} h(G^A_{klij})Q^{IA}_{klij}$ is the bottom-up input from the spatial attention gain fields.

5.5.3. Gain fields for spatial attentional shrouds

The gain fields G^A_{klij} from binocular surface to attentional interneuron (**Figures 4**, **5**) obey:

$$\frac{dG^A_{klij}}{dt} = \left(1 - G^A_{klij}\right)\left(\sum_{ij} S^b_{ij}M^{SI}_{klij} + \sum_{ij} P_{ij}M^{PI}_{klij} + \sum_{ij} A^I_{ij}M^{AI}_{klij}\right) - \left(G^A_{klij} + 0.37\right)\sum_{klij} G^A_{klij}, \quad (56)$$

where S^b_{ij} is the binocular surface representation (Equation 41), P_{ij} is the target position signal (Equation 66), and A^I_{ij} is the attentional interneuronal activity (Equation 55). The Gaussian gain field kernels M^{SI}_{klij}, M^{PI}_{klij}, M^{AI}_{klij} obey:

$$M^{SI}_{klij} = \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2_{G^{SI}_A}}\right); \sigma_{G^{SI}_A} = 3.2 \quad (57)$$

$$M^{PI}_{klij} = \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2_{G^{PI}_A}}\right); \sigma_{G^{PI}_A} = 1.3 \quad (58)$$

$$M^{AI}_{klij} = \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2_{G^{AI}_A}}\right); \sigma_{G^{AI}_A} = 5 \quad (59)$$

In the simulations, the top-down and bottom-up gain field weights are symmetrical:

$$M^{SI}_{klij} = M^{IS}_{klij}; M^{PI}_{klij} = M^{IP}_{klij}; M^{AI}_{klij} = M^{IA}_{klij} \quad (60)$$

5.5.4. Habituative attentional transmitter gates

The habituative attentional transmitter gate (Equation 51) obeys:

$$\frac{dy^A_{ij}}{dt} = \eta_A\left(\left(1.5 - y^A_{ij}\right) - 10^3 A^I_{ij}y^A_{ij}\right), \quad (61)$$

where $\eta_A = 10^{-5}$ is a slow rate of decay, $(1.5 - y^A_{ij})$ says that the gate y^A_{ij} passively accumulates to a maximal activity of 1.5, and $-10^3A^I_{ij}y^A_{ij}$ describes the activity-dependent habituation of y^A_{ij}.

5.5.5. Shroud-mediated parietal reset and habituation

The parietal reset neurons are tonically active and their activities are inhibited by inputs from all the active cells across the spatial attention map. Their activity is disinhibited when an attentional shroud collapses, and generates a transient activity burst that inhibits, and resets, the spatial attention map. This reset mechanism (Chang et al., 2014) obeys:

$$C_{RESET} = 10\left[1 - \varepsilon - \frac{\sum_{ij} g(A_{ij})}{100 + \sum_{ij} g(A_{ij})}\right]^+, \quad (62)$$

where $\varepsilon = 0.07$ is a small threshold, A_{ij} (Equation 51) is the activity of spatial attention at position *(i,j)* and g is defined in Equation (52).

The reset habituative transmitter y^C that gates the parietal reset signal obeys:

$$\frac{dy^C}{dt} = 10\left(0.75\left(1.5 - y^C\right) - 4C_{RESET}y^C\right). \quad (63)$$

As in Equation (61), this habituative gate also consists of a passive accumulation term $0.75(1.5 - y^C)$ and an activity-dependent habituation term $-4C_{RESET}y^C$.

5.6. EYE SIGNALS

5.6.1. Eye movement signals to salient features and inhibition of return

Surface contour cell activities (Equation 45) are contrast-enhanced using a recurrent on-center off-surround network to choose the activity F_{ij} of the most salient feature, and thus the target position *(i,j)* for the next saccadic eye movement. A movement habituative transmitter gate weakens this choice in an activity-dependent way, thereby providing an inhibition-of-return mechanism which ensures that the same target position is not perseveratively chosen.

Salient feature F_{ij} at position (i, j) obeys:

$$\frac{dF_{ij}}{dt} = -15F_{ij} + \left(2 - F_{ij}\right)\left(\left[C_{ij}\right]^+ + 250F^2_{ij}\right)y^F_{ij} - 0.04F_{ij}\sum_{ij}\left(\left[C_{ij}\right]^+ + F^2_{ij}\right), \quad (64)$$

where $C_{ij}$is the surface contour activity (Equation 45), and y_{ij}^F is the movement habituative gate::

$$\frac{dy_{ij}^F}{dt} = \eta_F \left((2 - 10^5 y_{ij}^F \left(\left[C_{ij} \right]^+ + 250 F_{ij}^2 \right) \right), \tag{65}$$

where $\eta_F = 10^{-4}$ is rate of decay. Note that this rate of decay is an order of magnitude larger than η_A, the rate of habituative decay for the spatial shrouds (Equation 61). Thus, the attentional shroud collapses much slower than inhibition-of-return of individual saccades that search the corresponding object (Chang et al., 2014). This rate difference enables multiple saccades within the attended surface to be explored and to thereby trigger learning of view-specific categories that encode multiple views of the attended object.

5.6.2. Target position signal

The target position signal at (i, j) obeys:

$$P_{ij} = \begin{cases} 1 & \text{for } F_{ij} = \max_{ij} \left(F_{ij} \right) \ \forall \ (i, j) \\ 0 & \text{otherwise.} \end{cases} \tag{66}$$

This determines the next predictive eye position signal from the highest activity position, or salient feature, on the surface contour map (Equation 45). All the gain field cells for boundaries, surfaces, and spatial attention processing have access to this positional signal (cf. Pouget and Snyder, 2000).

6. DISCUSSION

This article builds on the ARTSCAN and pARTSCAN models of how spatial attention in the Where stream modulates invariant object learning, recognition, and eye movement exploration of multiple object views in the What stream (Grossberg, 2007, 2009; Fazl et al., 2009; Cao et al., 2011; Foley et al., 2012; Chang et al., 2014). The 3D ARTSCAN model that is described herein extends these insights to explain how these processes can work in response to 3D objects and scenes. Together, these interacting processes model how mechanisms for maintaining stable binocular percepts of 3D objects are related to mechanisms for learning to invariantly categorize and recognize these objects.

A key insight of the current model concerns how predictive remapping through eye position-dependent gain fields maintains perceptual stability of binocularly fused images and scenes during saccadic eye movements. Additional processes of the 3D LAMINART model, a laminar cortical embodiment and further development of the FACADE model of 3D vision and figure-ground segregation (Grossberg, 1994, 1999; Kelly and Grossberg, 2000; Raizada and Grossberg, 2003; Grossberg and Swaminathan, 2004; Cao and Grossberg, 2005, 2012; Grossberg and Yazdanbakhsh, 2005; Fang and Grossberg, 2009), may be joined to the ARTSCAN model to clarify how more complex properties of 3D scenes than are simulated herein retain their perceptual stability under free viewing conditions.

6.1. FACADE AND 3D ARTSCAN

FACADE theory proposes how visible 3D surfaces are captured by binocularly fused 3D boundaries. Surface capture is achieved when depth-selective filling-in of surface brightness and color is triggered by these boundaries through their function as *filling-in generators* (Grossberg, 1994). Boundaries also function as *filling-in barriers* that restrict filling-in within surface regions that the boundaries surround. The filled-in features can be derived either from bottom-up object brightness and color contrasts or from top-down attentional spotlights. An attentional spotlight can, for example, arise when top-down spatial attentional signals from parietal cortex modulate filled-in object surfaces in a depth-selective manner within visual cortical areas such as V4.

The 3D ARTSCAN model shows, in addition, how binocularly fused boundaries can use eye position-dependent gain fields to maintain fusion and an invariant head-centered representation during eye movements (**Figure 3**). These invariant boundaries can capture left and right eye monocular surface features in a depth-selective way (**Figure 4**). The captured monocular surfaces can, in turn, form and maintain binocular surfaces (**Figure 4**). An attended binocular surface is modulated by an attentional shroud, with gain fields again ensuring that the interactions are dimensionally consistent (**Figure 4**). Thus, during filling-in, surface contrasts are activated either bottom-up from the binocularly combined monocular surfaces after they are captured in depth by the binocular boundaries, or top-down from the surface's attentional shroud.

FACADE model retinal lightness adaptation, spatial contrast adaptation, and double opponent processing (Grossberg and Hong, 2006) are among the useful pre-processing stages that are incorporated in the 3D ARTSCAN model. The 3D ARTSCAN model does not, however, yet process chromatic natural scenes, such as in the aFILM simulations of anchoring (Hong and Grossberg, 2004; Grossberg and Hong, 2006); or orientationally-selective depth-selective boundary completion processes, such as in the 3D LAMINART model simulations of binocular stereograms (Fang and Grossberg, 2009), the LIGHTSHAFT model simulations of 3D shape-from-texture (Grossberg et al., 2007), and the FACADE model simulations of da Vinci stereopsis (Grossberg and McLoughlin, 1997; Cao and Grossberg, 2005, 2012); or moving-form-in-depth processes, such as in the 3D FORMOTION model simulations of coherent and incoherent plaid motion, speed perception, and the aperture problem (Chey et al., 1997, 1998), transformational apparent motion (Baloch and Grossberg, 1997), the chopsticks and rotating ellipse illusions (Berzhanskaya et al., 2007), and the barberpole illusion, line capture, and motion transparency (Grossberg et al., 2001). All of these other studies are computationally consistent with the 3D ARTSCAN model and hence their competences can be incorporated in future model extensions.

6.2. ATTENTIONAL SHROUDS AND SURFACE-SHROUD RESONANCES: SEEING AND KNOWING

The 3D ARTSCAN model also does not explicitly study invariant object category learning and recognition, although the concept of attentional shrouds in the ARTSCAN and pARTSCAN models, which plays a key role in modulating invariant category learning in those models, also clarifies in the current study how an object in depth maintains its perceptual stability and attentional focus during eye movements (**Figures 1**, **4**).

The original use of the attentional shroud concept is closer to its perceptual role in 3D ARTSCAN than it is to its learned categorization role in ARTSCAN and pARTSCAN. In particular, the concept of an attentional shroud was introduced by Tyler and Kontsevich (1995) to clarify how spatial attention could morph itself to the shape of an object in depth, and how, in response to a transparent display, only one depth at a time might be perceived. Likova and Tyler (2003), also noted that "depth surface reconstruction is the key process in the accuracy of the interpolated profile from both depth and luminance signals" (see p. 2655), and thus that shroud formation involves surface fillng-in. However, they did not provide a design rationale or mechanistic explanation of these empirical facts.

The 3D ARTSCAN model does explain and simulate mechanistically how such depth-selective shrouds may form in the brain (**Figure 4**). Moreover, as noted above, the ARTSCAN family of models proposes how shrouds can form in response to either exogenously activated attention, via bottom-up inputs from objects in a scene, or endogenously activated attention, via a top-down route. In the 3D ARTSCAN model, once the attentional shroud fits itself to binocular surface input signals, the 3D surface-shroud resonance (**Figures 4**, **5**) is the dynamical state corresponding to "paying spatial attention" to the object surface. Such a 3D surface-shroud resonance is a mechanistic revision and explanation of the proposal of Tyler and Kontsevich (1995, p. 138) that "stereoscopic-attentional process therefore would be much more valuable if it could be wrapped around the form of any spatial object, rather than being restricted to frontoparallel planes… more vivid representation of this process is to think of it as an attentional shroud, wrapping the dense locus of activated disparity detectors as a cloth wraps a structured object." The 3D ARTSCAN model extends this view by proposing that it is the *3D surface-shroud resonance* which embodies a unified representation of consciously perceived object structure, not just the shroud taken alone, as in the Tyler and Kontsevich (1995) proposal. Boundary-category resonances and surface-category resonances are other aspects of object structure, whereby 3D boundary and surface representations interact reciprocally with their corresponding object category representations to invariantly categorize and recognize these object properties. Said more simply, these various resonances can synchronously represent seeing an object and knowing what it is.

6.3. COMPARISON WITH OTHER MODELS

To study object-based attention, LaBerge and Brown (1989) modeled attention as a gradient across the visual field with the peak at the expected target location. This gradient hypothesis could explain attention shifts better than a moving spotlight of attention, especially when spatial attention can form over more than one object. They also discussed how such a system could help in object recognition, especially in the identification of a visual shape in a cluttered scene. The model proved better than non-gradient based models of attention in explaining data on pre-cueing of locations in the visual field and of words.

Within the 3D ARTSCAN model, gradient properties can arise due to bottom-up properties of filling-in, the spatially distributed kernel that carries surface-to-shroud inputs, and the non-uniform distribution of shroud activity due to inhibition-of-return and activity-dependent habituation (Equations 51–66). Gradient properties can also be induced when a prefrontally-mediated top-down attentional spotlight, as modeled by Foley et al. (2012), remains on through time due to persistent volitional gain control (Brown et al., 2004; Grossberg, 2012, 2013) and combines with bottom-up shroud-maintaining mechanisms.

Logan (1996) integrated space-based and object-based approaches to visual attention by combining the COntour DEtector (CODE) theory of perceptual grouping by proximity (Van Oeffelen and Vos, 1982, 1983) with the Theory of Visual Attention (TVA) (Bundesen, 1990). In this unified Code Theory of Visual Attention (CTVA), CODE provides input to TVA, thereby accounting for spatially based between-object selection, while TVA converts the input to output, thereby accounting for feature- and category-based within-object selection. CODE clusters nearby items into emergent perceptual groupings that are both perceptual objects and regions of space, thereby integrating object-based and space-based approaches to attention. The theory assumes that attention chooses among perceptual objects by sampling the features that occur within an above-threshold region. The features of different items within this region are sampled with a probability that equals the area of the distribution of the item that falls within the region. This sampling probability is called the *feature catch*.

ARTSCAN also combines space-based and object-based visual attention. The space-based attention concerns how an object-fitting attentional shroud (cf. an "above-threshold region") controls both the learning of invariant object categories and their recognition, including when recognition may break down due to the inability of a shroud to form around a target object, as is predicted to happen during perceptual crowding (Foley et al., 2012). At least three types of grouping occur in the ARTSCAN framework: The first concerns the kind of feature-based grouping of perceptual boundaries that explains Gestalt grouping laws (e.g., Grossberg and Pinna, 2012). The second concerns the surface grouping that occurs during a surface-shroud resonance. And the third concerns how these emergent boundary and surface representations are bound into view-specific categories, and how view-specific categories are, in turn, bound into invariant object categories. Object attention enters ARTSCAN in two ways: Adaptive Resonance Theory top-down expectations control the learning of ARTSCAN categories by focusing object attention upon predictive combinations of object features. Object attention also plays a key role in controlling a primed search for a desired object, as during a solution of the Where's Waldo problem, which is modeled by the ARTSCAN Search model (Chang et al., 2014). These various processes occur on multiple spatial and temporal scales, and clarify some of the complexities that occur when object and spatial attentional processes interact.

Visual attention and search models, such as Guided Search (Wolfe et al., 1989; Wolfe, 2007), and Saliency Map (Itti and Koch, 2001) models, have their genesis in Feature Integration Theory (Treisman and Gelade, 1980). In these models, the units are local features or positions. The models are thus *pixel-based*.

The model mechanisms are based on competition between parallel visual representations, whereby a strong local salient feature wins and directs shifts in attention and eye movements to it (Deubel and Schneider, 1996; Deubel et al., 2002). In particular, in Saliency Map models, (e.g., Itti and Koch, 2001) different feature maps, such as brightness, orientation, color, or motion are computed in parallel visual representations. In each feature map, the strongest feature is selected by competition using an on-center, off-surround mechanism. The winning outputs of all these feature maps are then combined into a single map to build the saliency map. This saliency map predicts the probability with which a certain spatial positions will attract an observer's attention and eye movements.

Unlike pixel-based models, 3D ARTSCAN, as well as its ARTSCAN, pARTSCAN, dARTSCAN, and ARTSCAN Search variants, are *object-based* (Pylyshyn, 1989, 2001; Kahneman et al., 1992; Vergilino-Perez and Findlay, 2004) to enable the models to learn to attend, categorize, recognition, and search for objects in a scene. In these models, the competition for focusing attention, whether spatial (leading to a surface-shroud resonance) or object (leading to a feature-category resonance) is *regional* rather than local (Duncan, 1984).

The pre-processing of the 3D ARTSCAN model can be readily enhanced, as noted above, to include features such as color, orientation, and motion, as in the pixel-based models, but these features are bound into invariant binocular boundaries and retinotopic binocular surfaces which are the perceptual units that compete for spatial and object attention.

3D ARTSCAN can search a 3D scene to learn and recognize objects in it based on the salience of its boundary and surface properties, but it currently does so without accumulating evidence about contextual information. In contrast, in response to seeing a refrigerator and a stove, humans would expect to next see a sink more probably than a beach. 3D ARTSCAN does not learn such contextual expectations. In addition, 3D ARTSCAN, just like ARTSCAN and pARTSCAN before it, is devoted to *object*, rather than *scene*, perception, attention, learning, and recognition. 3D ARTSCAN is, however, one of a family of ART-based models (Carpenter and Grossberg, 1991, 1993) that do have these capabilities, and that can be combined in an enhanced future 3D ARTSCAN model.

For example, the ARTSCENE model (Grossberg and Huang, 2009) uses attentional shrouds to learn and recognize the gist of a scene as a large-scale texture category. ARTSCENE can also accumulate scenic evidence by using shrouds to iteratively focus attention on salient regions of the scene, and thereby learn texture categories at a finer scale, which can be combined by voting to improve scene recognition. However, ARTSCENE does not have a contextual memory of this accumulated scenic evidence through time.

Contextual cueing (e.g., Jiang and Chun, 2001; Olson and Chun, 2002) is modeled in the ARTSCENE Search model (Huang and Grossberg, 2010), which shows how spatial and object working memories can learn to accumulate and remember sequential contextual information to facilitate efficient search for an expected goal object, in the manner of the refrigerator/stove/sink example. In the ARTSCENE Search model, the object working memory involves perirhinal cortex interacting with prefrontal cortex, and the spatial working memory involves parahippocampal cortex, again interacting with prefrontal cortex. These brain regions also interact with inferotemporal and parietal cortices, respectively, among other brain areas, to determine where the eyes will look next. Thus, in ARTSCENE Search, each eye movement enables currently attended objects to be seen and recognized, while also triggering new category learning and working memory storage that can better predict goal objects in the future.

Another search variant that was mentioned above: the ARTSCAN Search model (Chang et al., 2014), uses pARTSCAN mechanisms to learn and recognize view- and positionally-invariant object categories using Where-to-What stream interactions. In addition, ARTSCAN Search can also search a scene for a valued goal object using What-to-Where stream interactions. Such a search may be activated by a top-down cognitive prime or motivational prime. The model hereby proposes a neurobiologically-grounded solution of the Where's Waldo problem.

6.4. ATTENTIONAL GAIN CONTROL AND NORMALIZATION: A CONVERGENCE ACROSS MODELS

Recent models of attention have focused on studying the effects of attention on neuronal responses in visual cortical areas such as MT and V4 (e.g., Ghose, 2009; Lee and Maunsell, 2009; Reynolds and Heeger, 2009). These models explored how attention enhances processing of selected areas of the visual field, and concluded that divisive normalization using center-surround processing causes the effects of attention on V4 neurons. Top-down attentional priming had earlier been modeled in the FACADE, ART, and 3D LAMINART models using top-down, modulatory on-center, off-surround networks acting on cells that obey the membrane, or shunting, equations of neurophysiology (e.g., Carpenter and Grossberg, 1987, 1991, 1993; Gove et al., 1995; Grunewald and Grossberg, 1998; Grossberg et al., 2001; Berzhanskaya et al., 2007; Bhatt et al., 2007). In ART, such a top-down circuit for attention is called the ART Matching Rule. These ART results, in turn, built on the fact that cells which obey shunting dynamics in on-center off-surround anatomies automatically compute the property of divisive normalization. Grossberg (1973) provided an early mathematical proof of this normalization property, and Grossberg (1980) contained an early review.

More recently, there has been a convergence across models of how to mathematically instantiate the ART Matching Rule attentional circuit. For example, the "normalization model of attention" (Reynolds and Heeger, 2009) simulates several types of experiments on attention using the same equation for self-normalizing attention that the distributed ARTEXture (dARTEX) model (Bhatt et al., 2007, Equation A5) used to simulate human psychophysical data about Orientation-Based Texture Segmentation (OBTS, Ben-Shahar and Zucker, 2004). Whereas Reynolds and Heeger (2009) described an algebraic form-factor for attention, Bhatt et al. (2007) described and simulated the attentional dynamics whose steady state reduces to that form factor. Although the 3D ARTSCAN model uses shunting competitive dynamics to define its attentional modulation at multiple processing stages, it is difficult to summarize their net effect in a single

steady-state equation due to the role of gain fields between surface and shroud representations to maintain perceptual stability during eye movements (see Equations 38–61).

6.5. BALANCING OBJECT EXPLORATION vs. PERSEVERATION: INHIBITION-OF-RETURN

The brain can learn view-invariant object categories by exploring multiple salient features on each object. But why are not successive eye movement positions instead chosen randomly, thereby preventing efficient intra-object exploration? Indeed, psychophysical data support the idea that the eyes prefer to move within the same object for awhile (Theeuwes et al., 2010), rather than randomly. The stability of the surface-shroud resonance while the eyes explore an object's surface helps to explain how this happens. Such a resonance maintains spatial attention on a given object for awhile, while also enhancing the activity of the attended surface's surface contours. The most active position on a surface contour is chosen as the next saccadic target position on the attended object (Fazl et al., 2009), a transformation that is predicted to take place using cortical area V3A (**Figure 1**).

The brain must also solve the problem of not perseveratively choosing the same maximally activated position over and over again. Inhibition of return (IOR) is an important mechanism for any model of attention (List and Robertson, 2007), or, for that matter, any model of sequential performance. Perseverative performance of maximally active eye movement representations is prevented by their activity-dependent habituation as they are chosen to determine next eye movement target position (see Equations 64–66). This choice-dependent inhibitory feedback enables the 3D ARTSCAN model to choose the next most active position as the next saccadic target location. The combination of a self-normalizing activity map, selection of the maximal activity for the next output, and choice-dependent inhibitory feedback was introduced in Grossberg (1978a,b; see also Grossberg and Kuperstein, 1986) and has been used in many subsequent models, notably Koch and Ullman (1985).

6.6. PREDICTIVE REMAPPING VIA EYE COMMAND-MEDIATED GAIN FIELDS

Visual stability and object constancy requires the visual system to keep track of the spatiotopic or allocentric positions of several objects in a scene during saccades (Mathot and Theeuwes, 2010a,b). Retinotopic coordinates generate different representations of the same scene when it is viewed at different centers of gaze. This fact has led many investigators to conclude that retinotopic representations are predictively remapped by eye movement commands, with eye position-sensitive gain fields as a key remapping mechanism (Von Holst and Mittelstaedt, 1950; Von Helmholtz, 1867; Duhamel et al., 1992; Gottlieb et al., 1998; Tolias et al., 2001; Melcher, 2007, 2008, 2009; Saygin and Sereno, 2008; Mathot and Theeuwes, 2010a,b). Corollary discharges of outflow movement signals that act before the eyes stabilize on their next movement target are used to update the gain fields.

Several fMRI studies suggest that various visual representations in the Where, or dorsal, cortical stream that are sensitive to visual attention are computed in retinotopic coordinates. At least one area in anterior parietal cortex has been found using fMRI to be responsive to head-centered, or some sort of spatiotopic or absolute, coordinates (Sereno and Huang, 2006). Perisaccadic remapping of receptive fields has been reported in electrophysiological studies in frontal eye fields (Goldberg and Bruce, 1990), in parietal areas, including LIP (Andersen et al., 1990; Duhamel et al., 1992), and in V4 (Tolias et al., 2001). Interestingly, in these regions, after saccades, no new transient activity is caused when targets are attended (see Mathot and Theeuwes, 2010a for a review).

Psychophysical experiments have suggested that predictive remapping is mediated by predictive shifts of attention to the positions of intended targets. Cavanagh et al. (2010) called these shifts "attention pointers" (see Section 2.5). Predictive remapping of visual attention enables improved attentional performance that enhances perceptual processing at target positions and speeds up the eye movements to the new target's position (Rolfs et al., 2011). In the 3D ARTSCAN and related ARTSCAN models, the maximally active position on a surface contour is chosen as the next saccadic target position before the eye movement occurs, and causes a predictive updating of gain fields to maintain the stability of a currently active shroud and of the 3D surface percept during intra-object movements, and to facilitate the shift of spatial attention to a newly attended object (Sections 2.5 and 2.6). It therefore seems that the maximally active surface contour position, as described in the Fazl et al. (2009) ARTSCAN article, predicted key properties of the Cavanagh et al. (2010) attention pointer data. One way to test if this proposed connection is mechanistically sound is to link it to other ARTSCAN predictions. For example, are attention pointers computed in cortical area V3A (**Figure 1**), as is compatible with the data of Caplovitz and Tse (2007, p. 1179) showing "neurons within V3A. . . process continuously moving contour curvature as a trackable feature. . . not to solve the 'ventral problem' of determining object shape but in order to solve the 'dorsal problem' of what is going where"?

6.7. RETINOTOPIC vs. SPATIOTOPIC REPRESENTATIONS

A recent behavioral study using fMRI in higher visual areas proposed that, in the dorsal visual stream and the intraparietal sulcus, all object locations are represented in retinotopic coordinates as their native coordinate system (Golomb and Kanwisher, 2012). These authors found little to no evidence of spatiotopic object position and suggested that a spatiotopic, or head-centered, ability to interact with objects in the world might be achieved by spatiotopic object positions that are "computed indirectly and continually reconstructed with each eye movement" (Golomb and Kanwisher, 2012, p. 2794), presumably using gain fields. One concern about an fMRI test of spatiotopic representation is that such a representation may be masked by the more rapidly changing retinotopic representations, especially given the kind of theoretical analyses presented here which suggest a preponderance of retinotopic representations, such as retinotopic boundary, surface, surface contour, and eye command representations, that are nested among a smaller number of spatiotopic representations, such as binocular boundary and attentional shroud

representations (**Figures 2–4**). Finer neurophysiological methods will likely be needed to sort out these retinotopic and spatiotopic differences, as they have begun to in past research.

Some behavioral experiments report a brief retinotopic facilitation (priming) effect followed by a sustained spatiotopic IOR effect (Posner and Petersen, 1990). The kind of stimuli in these experiments include attending to events in a given visual position, covert shifts in attention or orienting to a new position upon cuing, visual search (Posner and Cohen, 1984; Posner, 1988), as well as letter and word matching (Posner, 1978). Some behavioral measures for such data are collated from reaction times to efficiently respond to activities in the cued location (Posner, 1988), enhanced scalp electrical activity (Mangoun and Hillyard, 1987), higher discharge rates of neurons in several areas of the monkey brain (Mountcastle, 1978; Wurtz et al., 1980; Petersen et al., 1987), spared abilities of patients with lesions and monkeys with chemical lesions in different areas of the brain (Posner and Cohen, 1984; Posner et al., 1984; Posner, 1988), and how each area and hemispheric differences affects the ability to engage in attention, orient or remain alert to a target (Gazzaniga, 1970; Sergent, 1982; Robertson and Delis, 1986).

The brief facilitation was due to the activation of retinotopic units representing the stimulus, in which case, the selection of a response occurs more quickly than when not expecting a target to occur or when targets occur without warning. This selection of a response, though, is based upon a lower quality of information about the classification of the target stimulus, resulting in an increase in error rate to respond to the stimulus. This increase in errors, while not affecting the build-up of information in the retinotopic system, affects the rate at which attention can respond to the stimulus leading to a sustained spatiotopic IOR. 3D ARTSCAN mechanisms are compatible with such data, since the retinotopic representations are used to build spatiotopic representations, and shroud IOR mechanisms are computed in spatiotopic coordinates.

Various experiments find persistent spatiotopic facilitation along with short-term retinotopic facilitation in certain task conditions (Golomb et al., 2008, 2010a,b). Thus, contextual relevance of tasks may play a role in whether object locations are coded in retinotopic or head-centered/spatiotopic coordinates systems. For example, in Golomb et al. (2008), the manipulation of the Stimulus Onset Asynchrony of the probe stimulus enabled the tracking of when the transition between retinotopic and spatiotopic coordinates occurs. In one of the experiments to sustain a stable spatiotopic representation, immediately after a saccade, attention is primarily maintained at the previously relevant retinotopic coordinates of the cue. However, after 100–200 ms, the task-relevant spatiotopic coordinates start to dominate and the retinotopic facilitation decays. On the other hand, when the experiment was modified to make the retinotopic location the task-relevant location and the spatiotopic location task-irrelevant, the retinotopic location was facilitated over the entire delay period of 75–600 ms probed. This kind of manipulation gives insight into the temporal dynamics of spatial attention and the mechanisms by which attention is maintained across saccades.

6.8. REMAPPING OF BORDER-OWNERSHIP IN V2 AND ATTENTIVE ENHANCEMENT IN V1

The electrophysiological experiments of O'Herron and von der Heydt (2013) on border-ownership neurons in visual cortical area V2 of monkeys showed that there is remapping of border-ownership signals when the retinal image moves either due to saccades or object movements. A border-ownership neuron responds to borders with differing firing rates depending on whether the border is owned by a figure on one side or the other. The difference in firing rates to the two conditions is defined as the border-ownership signal. An ambiguous edge was used as a probe in both cases. In the saccade paradigm, the edge of a figure (square) is presented outside the cell receptive field (RF) in the first phase. This is substituted by the ambiguous edge in the second phase. In the third phase, a saccade is induced to move the RF into the ambiguous edge. The V2 neuron did not respond during the first two phases, but responded when the saccade brought the RF onto the edge. The difference in the response was related to neither the direction of the saccade nor the location of the figure relative to the RF, but to the initial border-ownership. The border-ownership defined by the figure edge was inherited by the ambiguous edge and transferred across cortex at the time of saccade. In the object movement paradigm, the displays used in the first two phases were the same as for the saccades paradigm. In the third phase, instead of moving the fixation point (as was done in the saccade condition), the figure edge along with the object were moved to have the edge land in the RF of the neuron. The results were similar to those of the saccade experiment in terms of the amplitudes of the transferred signals. The response onset and rise of the border-ownership signal in the object movement were more abrupt and aligned to the edge movement. For the saccade condition, they were aligned with the movement of the fixation point and the response onset varied with saccade latency. This remapping of border-ownership was observed in both the paradigms at the V2 population level as well.

Border-ownership modulation of neurons in area V2 is akin to the remapping often observed in neurons in areas controlling visual attention and planning of eye movements, in which a stimulus activates a neuron whose RF has not yet seen the stimulus (e.g., Duhamel et al., 1992), showing that remapping may occur in low-level visual areas as well.

The FACADE and 3D LAMINART models have simulated a number of figure-ground percepts using model neural mechanisms in V2. These percepts include Bregman-Kanizsa figure-ground separation and various lightness percepts, including the Munker-White, Benary cross, and checkerboard percepts (Kelly and Grossberg, 2000), percepts of Kanizsa stratification, transparency, and 3D neon color spreading (Grossberg and Yazdanbakhsh, 2005), and bistable percepts, including their modulation by attention, such as the percept of a Necker cube (Grossberg and Swaminathan, 2004) and binocular rivalry (Grossberg et al., 2008). Because these models can be consistently added to the pre-processing levels in 3D LAMINART, they can be explained in this model in a manner consistent with the figure-ground remapping results.

A study involving a curve tracing task, with multi-unit activity recorded from monkey visual cortical area V1, established remapping of response modulation for attentive enhancement (Khayat et al., 2004). In this work, the monkeys performed a curve tracing task, and had to make two successive saccades along a single curve to which they were attending, while ignoring another curve. Response enhancement for the neurons representing the selected curve was observed. After the first saccade, there was enhancement in the response of the neurons representing the curve in the new retinal locations. Response modulation appeared in neurons that had not been activated initially, and the attentive enhancement was remapped, or transferred across cortex. This response modulation to attentive enhancement in V1 is strikingly similar to the predictive remapping often observed in neurons in LIP and other areas that control visual attention and planning of predictive eye movements and requires the selective attention of one stimulus over the other for response modulation.

The two studies summarized above appear to differ in the role of attention in remapping, but are complementary and can be integrated within the 3D ARTSCAN model. To achieve such remapping, both the systems need to compute the displacement vector of the shift. In predictive remapping, this displacement information is provided by the outflow command of the eye movement centers, which update gain fields that drive the remapping. The similarity of the results for saccades or object movement in the border-ownership in V2, and the response modulation in V1 to attentive enhancement, are consistent with the remapping via gain fields, that is used in the 3D ARTSCAN model, and lend further support to the FAÇADE theory claim that figure-ground mechanisms for boundary formation, and thus for their remapping, can occur at early stages of visual cortex. Despite frequent saccades or displacement on the retina, early remapping is essential to maintain assignment of local features to an external object. Such congruity serves as a crucial step toward building object invariance, and enabling the integration of details of the object into a coherent percept.

ACKNOWLEDGMENTS

Supported in part by CELEST, an NSF Science of Learning Center (SBE-0354378), and by the SyNAPSE program of DARPA (HR0011-09-03-0001).

REFERENCES

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., and Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *J. Neurosci.* 10, 1176–1196.

Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science* 230, 456–458. doi: 10.1126/science.4048942

Andersen, R. A., and Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light- sensitive neurons of the posterior parietal cortex. *J. Neurosci.* 3, 532–548.

Baloch, A. A., and Grossberg, S. (1997). A neural model of high-level motion processing: line motion and formotion dynamics. *Vision Res*, 37, 3037–3059. doi: 10.1016/S0042-6989(97)00103-X

Baylis, G. C., and Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nat. Neurosci.* 4, 937–942. doi: 10.1038/nn0901-937

Beardslee, D. C., and Wertheimer, M. (1958). "Readings in perception," in *The University Series in* Psychology (Princeton, NJ: Van Nostrand), 94–203, 751.

Beauvillain, C., Vergilino-Perez, D., and Dukic, T. (2005). Spatial object representation and its use in planning eye movements. *Exp. Brain Res.* 165, 315–327. doi: 10.1007/s00221-005-2308-8

Ben-Shahar, O., and Zucker, S. (2004). Sensitivity to curvatures in orientation-based texture segmentation. *Vision Res.* 44, 257–277. doi: 10.1016/j.visres.2003.08.018

Berzhanskaya, J., Grossberg, S., and Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spat. Vis.* 20, 337–395. doi: 10.1163/156856807780919000

Bhatt, R., Carpenter, G. A., and Grossberg, S. (2007). Texture segregation by visual cortex: Perceptual grouping, attention, and learning. *Vision Res.* 47, 3173–3211. doi: 10.1016/j.visres.2007.07.013

Brown, J. M., and Denny, H. I. (2007). Shifting attention into and out of objects: evaluating the processes underlying the object advantage. *Percept. Psychophys.* 69, 608–618. doi: 10.3758/BF03193918

Brown, J. W., Bullock, D., and Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Netw.* 17, 471–510. doi: 10.1016/j.neunet.2003.08.006

Bundesen, C. (1990). A theory of visual attention. *Psychol. Rev.* 97, 523–547. doi: 10.1037/0033-295X.97.4.523

Cabeza, R., Ciaramelli, E., Olson, I. R., and Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nat. Rev. Neurosci.* 9, 613–625. doi: 10.1038/nrn2459

Cao, Y., and Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: closure and da Vinci stereopsis. *Spat. Vis.* 18, 515–578. doi: 10.1163/156856805774406756

Cao, Y., and Grossberg, S. (2012). Stereopsis and 3D surface perception by spiking neurons in laminar cortical circuits: a method of converting neural rate models into spiking models. *Neural Netw.* 26, 75–98. doi: 10.1016/j.neunet.2011.10.010

Cao, Y., Grossberg, S., and Markowitz, J. (2011). How does the brain rapidly learn and reorganize view- and positionally-invariant object representations in inferior temporal cortex? *Neural Netw.* 24, 1050–1061. doi: 10.1016/j.neunet.2011.04.004

Caplovitz, G. P., and Tse, P. U. (2007). V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cereb. Cortex*, 17, 1179–1189. doi: 10.1093/cercor/bhl029

Carlson-Radvansky, L. A. (1999). Memory for relational information across eye movements. *Percept. Psychophys.* 61, 919–934. doi: 10.3758/BF03206906

Carpenter, G. A., and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern-recognition machine. *Comput. Vision Graph. Image Process.* 37, 54–115. doi: 10.1016/S0734-189X(87)80014-2

Carpenter, G. A., and Grossberg, S. (1991). *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press.

Carpenter, G. A., and Grossberg, S. (1993). Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends Neurosci.* 16, 131–137. doi: 10.1016/0166-2236(93)90118-6

Carrasco, M., Penpeci-Talgar, C., and Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Res.* 40, 1203–1215. doi: 10.1016/S0042-6989(00)00024-9

Cavada, C., and Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: II. evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J. Comp. Neurol.* 287, 422–445. doi: 10.1002/cne.902870403

Cavada, C., and Goldman-Rakic, P. S. (1991). Topographic segregation of corticostriatal projections from posterior parietal subdivisions in the macaque monkey. *Neuroscience* 42, 683–696. doi: 10.1016/0306-4522(91)90037-O

Cavanagh, P., Hunt, A. R., Afraz, A., and Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends Cogn. Sci.* 14, 147–153. doi: 10.1016/j.tics.2010.01.007

Cavanagh, P., Labianca, A. T., and Thornton, I. M. (2001). Attention-based visual routines: sprites. *Cognition* 80, 47–60. doi: 10.1016/S0010-0277(00)00153-0

Chang, H.-C., Grossberg, S., and Cao, Y. (2014). Where's Waldo? How perceptual, cognitive, and emotional brain processes cooperate during learning to categorize and find desired objects in a cluttered scene. *Front. Integr. Neurosci.*8:43. doi: 10.3389/fnint.2014.00043

Chey, J., Grossberg, S., and Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Res.* 38, 2769–2786. doi: 10.1016/S0042-6989(97)00372-6

Chey, J., Grossberg, S., and Mingolla, M. (1997). Neural dynamics of motion grouping: from aperture ambiguity to object speed and direction. *J. Opt. Soc. Am.* 14, 2570–2594. doi: 10.1364/JOSAA.14.002570

Chiu, Y. C., and Yantis, S. (2009). A domain-independent source of cognitive control for task sets: shifting spatial attention and switching categorization rules. *J. Neurosci.* 29, 3930–3938. doi: 10.1523/JNEUROSCI.5737-08.2009

Cohen, M. A., and Grossberg, S. (1984). Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. *Percept. Psychophys.* 36, 428–456. doi: 10.3758/BF03207497

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietalcortex. *Nat. Neurosci.* 3, 292–297. doi: 10.1038/73009

Deneve, S., and Pouget, A. (2003). Basis functions for object-centered representations. *Neuron,* 37, 347–359. doi: 10.1016/S0896-6273(02)01184-4

Deubel, H., and Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res.* 36, 1827–1837. doi: 10.1016/0042-6989(95)00294-4

Deubel, H., Schneider, W. X., and Bridgeman, B. (2002). Transsaccadic memory of position and form. *Prog. Brain Res.* 140, 165–180. doi: 10.1016/S0079-6123(02)40049-0

Distler, C., Boussaoud, D., Desimone, R., and Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *J. Comp. Neurol.* 334, 125–150. doi: 10.1002/cne.903340111

Driver, J., and Baylis, G. C. (1996). Edge-assignment and figure-ground segmentation in short term visual matching. *Cogn. Psychol.* 31, 248–306. doi: 10.1006/cogp.1996.0018

Duhamel, J. R., Colby, C. L., and Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255, 90–92. doi: 10.1126/science.1553535

Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* 113, 501–517. doi: 10.1037/0096-3445.113.4.501

Egeth, H. E., Virzi, R. A., and Garbart, H. (1984). Searching for conjunctively defined targets. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 32–39. doi: 10.1037/0096-1523.10.1.32

Elder, J. H., and Zucker, S. W. (1998). Evidence for boundary-specific grouping. *Vision Res.* 38, 143–152. doi: 10.1016/S0042-6989(97)00138-7

Elder, J., and Zucker, S. (1993). The effect of contour closure on the rapid discrimination of 2- dimensional shapes. *Vision Res.* 33, 981–991. doi: 10.1016/0042-6989(93)90080-G

Fahrenfort, I., Habets, R. L., Spekreijse, H., and Kamermans, M. (1999). Intrinsic cone adaptation modulates feedback efficiency from horizontal cells to cones. *J. Gen. Physiol.* 114, 511–524. doi: 10.1085/jgp.114.4.511

Fang, L., and Grossberg, S. (2009). From stereogram to surface: how the brain sees the world in depth. *Spat. Vis.* 22, 45–82. doi: 10.1163/156856809786618484

Fazl, A., Grossberg, S., and Mingolla, E. (2009). View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. *Cogn. Psychol.* 58, 1–48. doi: 10.1016/j.cogpsych.2008.05.001

Fecteau, J. H., and Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nat. Rev. Neurosci.* 4, 435–443. doi: 10.1038/nrn1114

Fei-Fei, L., Fergus, R., and Perona, P. (2004). Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *IEEE CVPR* 2004, 178. doi: 10.1109/CVPR.2004.109

Foley, N. C., Grossberg, S., and Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: object cueing, useful-field-of-view, and crowding. *Cogn. Psychol.* 65, 77–117. doi: 10.1016/j.cogpsych.2012.02.001

Gancarz, G., and Grossberg, G. (1998). A neural model of the saccade generator in the reticular formation. *Neural Netw.* 11, 1159–1174. doi: 10.1016/S0893-6080(98)00096-3

Gancarz, G., and Grossberg, S. (1999). A neural model of saccadic eye movement control explains task-specific adaptation. *Vision Res.* 39, 3123–3143. doi: 10.1016/S0042-6989(99)00049-8

Gazzaniga, M. S. (1970). *The Bisected Brain*. New York, NY: Appleton.

Ghose, G. M. (2009). Attentional modulation of visual responses by flexible input gain. *J. Neurophysiol.* 101, 2089–2106. doi: 10.1152/jn.90654.2008

Goldberg, M. E., and Bruce, C. J. (1990). Primate frontal eye fields. *III. Maintenance of a spatially accurate saccade signal. Journal of Neurophysiology*, 64, 489–508.

Golomb, J. D., Chun, M. M., and Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *J. Neurosci.* 28, 10654–10662. doi: 10.1523/JNEUROSCI.2525-08.2008

Golomb, J. D., and Kanwisher, N. (2012). Higher-level visual cortex represents retinotopic, not spatiotopic, object location. *Cereb. Cortex* 22, 2794–2810. doi: 10.1093/cercor/bhr357

Golomb, J. D., Nguyen-Phuc, A. Y., Mazer, J. A., McCarthy, G., and Chun, M. M. (2010a). Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. *J. Neurosci.* 30, 10493–10506. doi: 10.1523/JNEUROSCI.1546-10.2010

Golomb, J. D., Pulido, V. Z., Albrecht, A. R., Chun, M. M., and Mazer, J. A. (2010b). Robustness of the retinotopic attentional trace after eye movements. *J. Vis.* 10, 11–12. doi: 10.1167/10.3.19

Gottlieb, J. P., Kusunoki, M., and Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature* 391, 481–484. doi: 10.1038/35135

Gove, A., Grossberg, S., and Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Vis. Neurosci.* 12, 1027–1052. doi: 10.1017/S0952523800006702

Grossberg, S. (1972). A neural theory of punishment and avoidance, II: quantitative theory. *Math. Biosci.*15, 253–285. doi: 10.1016/0025-5564(72)90038-7

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 213–257.

Grossberg, S. (1978a). "A theory of human memory: self-organization and performance of sensory-motor codes, maps, and plans," in *Progress in Theoretical Biology*, Vol. 5, eds R. Rosen and F. Snell (New York, NY: Academic Press), 233–374.

Grossberg, S. (1978b). Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? *J. Math. Psychol.* 3, 199–219.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychol. Rev.* 87, 1–51. doi: 10.1037/0033-295X.87.1.1

Grossberg, S. (1984). Some psychophysiological and pharmacological correlates of a developmental, cognitive and motivational theory. *Ann. N.Y. Acad. Sci.* 425, 58–151. doi: 10.1111/j.1749-6632.1984.tb23523.x

Grossberg, S. (1987a). Cortical dynamics of 3-dimensional form, color, and brightness perception.1. monocular theory. *Percept. Psychophys.* 41, 87–116. doi: 10.3758/BF03204874

Grossberg, S. (1987b). Cortical dynamics of 3-dimensional form, color, and brightness perception.2. binocular theory. *Percept. Psychophys.* 41, 117–158. doi: 10.3758/BF03204875

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Percept. Psychophys.* 55, 48–121. doi: 10.3758/BF03206880

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two dimensional figures. *Psychol. Rev.* 104, 618–658. doi: 10.1037/0033-295X.104.3.618

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spat. Vis.* 12, 163–185. doi: 10.1163/156856899X00102

Grossberg, S. (2003). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behav. Cogn. Neurosci. Rev.* 2, 47–76. doi: 10.1177/1534582303002001003

Grossberg, S. (2007). "Towards a unified theory of neocortex: laminar cortical circuits for vision and cognition, in *For Computational Neuroscience: From Neurons to Theory and Back Again*, eds P. Cisek, T. Drew, and J. Kalaska (Amsterdam: Elsevier), 79–104.

Grossberg, S. (2013). "Cortical dynamics of attentive object recognition, scene understanding, and decision making," in *Human Information Processing: Vision, Memory, and Attention*, eds C. Chubb, B. A. Dosher, Z.-L. Lu, and R. M. Shiffrin (Washington, DC: American Psychological Association), 165–191.

Grossberg, S. (2009). Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion, and action. *Philos. Trans. R. B Soc. Lond.* 364, 1223–1234. doi: 10.1098/rstb.2008.0307

Grossberg, S. (2012). Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world. *Neural Netw.* 37, 1–47. doi: 10.1016/j.neunet.2012.09.017

Grossberg, S., and Hong, S. (2006). A neural model of surface perception: lightness, anchoring, and filling-in. *Spat. Vis.* 19, 263–321. doi: 10.1163/156856806776923399

Grossberg, S., and Howe, P. D. (2003). A laminar cortical model of stereopsis and three dimensional surface perception. *Vision Res.* 43, 801–829. doi: 10.1016/S0042-6989(03)00011-7

Grossberg, S., and Huang, T.-R. (2009). ARTSCENE: A neural system for natural scene classification. *J. Vis*. 9:6 doi: 10.1167/9.4.6

Grossberg, S., Hwang, S., and Mingolla, E. (2002). Thalamocortical dynamics of the McCollough effect: boundary-surface alignment through perceptual learning. *Vision Res.* 42, 1259–1286. doi: 10.1016/S0042-6989(02)00055-X

Grossberg, S., and Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Res.* 39, 3796–3816. doi: 10.1016/S0042-6989(99)00095-4

Grossberg, S., Kuhlmann, L., and Mingolla, E. (2007). A neural model of 3D shape-from-texture: multiple-scale filtering, boundary grouping, and surface filling-in. *Vision Res.* 47, 634–672. doi: 10.1016/j.visres.2006.10.024

Grossberg, S., and Kuperstein, M. (1986). *Neural Dynamics of Adaptive Sensory-Motor Control: Ballistic Eye Movements.* Amsterdam; New York; North-Holland.

Grossberg, S., and McLoughlin, N. (1997). Cortical dynamics of 3-D surface perception: binocular and half-occluded scenic images. *Neural Netw*. 1997, 1583–1605. doi: 10.1016/S0893-6080(97)00065-8

Grossberg, S., and Mingolla, E. (1985a). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* 92, 173–211. doi: 10.1037/0033-295X.92.2.173

Grossberg, S., and Mingolla, E. (1985b). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Percept. Psychophys*. 38, 141–171. doi: 10.3758/BF03198851

Grossberg, S., and Mingolla, E. (1987). Neural dynamics of surface perception - boundary webs, illuminants, and shape-from-shading. *Comput. Vision Graph. Image Process*. 37, 116–165. doi: 10.1016/S0734-189X(87)80015-4

Grossberg, S., Mingolla, E., and Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Res.* 41, 2521–2553. doi: 10.1016/S0042-6989(01)00131-6

Grossberg, S., Mingolla, E., and Williamson, J. (1995). Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Netw.* 8, 1005–1028. doi: 10.1016/0893-6080(95)00079-8

Grossberg, S., and Pinna, B. (2012). Neural dynamics of Gestalt principles of perceptual organization: from grouping to shape and meaning. *Gestalt Theory* 34, 399–482.

Grossberg, S., Roberts, K., Aguilar, M., and Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *J. Neurosci.* 17, 9706–9725.

Grossberg, S., Srihasam, K., and Bullock, D. (2012). Neural dynamics of saccadic and smooth pursuit eye movement coordination during visual tracking of unpredictably moving targets. *Neural Netw*. 27, 1–20. doi: 10.1016/j.neunet.2011.10.011

Grossberg, S., and Swaminathan, G. (2004). A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention, and bistability. *Vision Res.* 44, 1147–1187. doi: 10.1016/j.visres.2003.12.009

Grossberg, S., and Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Percept. Psychophys*. 43, 241–277. doi: 10.3758/BF03207869

Grossberg, S., and Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: stratification, transparency, and neon color spreading. *Vision Res.* 45, 1725–1743. doi: 10.1016/j.visres.2005.01.006

Grossberg, S., Yazdanbakhsh, A., Cao, Y., and Swaminathan, G. (2008). How does binocular rivalry emerge from cortical mechanisms of 3-D vision? *Vision Res*. 48, 2232–2250. doi: 10.1016/j.visres.2008.06.024

Grunewald, A., and Grossberg, S. (1998). Self-organization of binocular disparity tuning by reciprocal corticogeniculate interactions. *J. Cogn. Neurosci.* 10, 199–215. doi: 10.1162/089892998562654

He, Z. J., and Nakayama, K. (1992). Surfaces versus features in visual search. *Nature* 359, 231–233. doi: 10.1038/359231a0

He, Z. J., and Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proc. Natl. Acad. Sci. U.S.A.* 92, 11155–11159. doi: 10.1073/pnas.92.24.11155

Henderson, J. M., and Hollingworth, A. (2003). Global transsaccadic change blindness during scene perception. *Psychol. Sci*. 14, 493–497. doi: 10.1111/1467-9280.02459

Hong, S., and Grossberg, S. (2004). A neuromorphic model for achromatic and chromatic surface representation of natural images. *Neural Netw*. 17, 787–808. doi: 10.1016/j.neunet.2004.02.007

Howard, I. P., and Rogers, B. J. (1995). *Binocular Vision and Stereopsis*. New York, NY: Oxford University Press.

Howe, P. D. L., and Watanabe, T. (2003). Measuring the depth induced by an opposite-luminance (but not anticorrelated) stereogram. *Perception* 32, 415–421. doi: 10.1068/p5026

Huang, T.-R., and Grossberg, S. (2010). Cortical dynamics of contextually cued attentive visual learning and search: Spatial and object evidence accumulation. *Psychol. Rev.* 117, 1080–1112. doi: 10.1037/a0020664

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cogn. Psychol.* 23, 420–456. doi: 10.1016/0010-0285(91)90015-G

Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci*. 2, 194–203. doi: 10.1038/35058500

Jiang, Y., and Chun, M. M. (2001). Selective attention modulates implicit learning. *Q. J. Exp. Psychol*. 54A, 1105–1124. doi: 10.1080/713756001

Julesz, B. (1971). *Foundations of Cyclopean Perception.* Chicago, IL: University of Chicago Press.

Kahneman, D., and Henik, A. (1981). "Perceptual organization and attention," in *Perceptual organization* eds M. Kubovy and J. R. Pomerantz (Hillsdale, NJ: Erlbaum), 181–211.

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219. doi: 10.1016/0010-0285(92)90007-O

Kelly, F., and Grossberg, S. (2000). Neural dynamics of 3-D surface perception: figure-ground separation and lightness perception. *Percept. Psychophys*. 62, 1596–1618. doi: 10.3758/BF03212158

Khayat, P. S., Spekreijse, H., and Roelfsema, P. R. (2004). Correlates of transsaccadic integration in the primary visual cortex of the monkey. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12712–12717. doi: 10.1073/pnas.0301935101

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.

Kourtzi, Z., and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science* 293, 1506–1509. doi: 10.1126/science.1061133

Koutalos, Y., and Yau, K. W. (1996). Regulation of sensitivity in vertebrate rod photoreceptors by calcium. *Trends Neurosci.* 19, 73–81. doi: 10.1016/0166-2236(96)89624-X

LaBerge, D. (1995). *Attentional Processing: The Brain's Art of Mindfulness*. Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674183940

LaBerge, D., and Brown, V. (1989). Theory of attentional operations in shape identification. *Psychol. Rev.* 96, 101–124. doi: 10.1037/0033-295X.96.1.101

Lamme, V. A. F., Rodriguez-Rodriguez, V., and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries, and surfaces in primary visual cortex of the macaque monkey. *Cereb. Cortex* 9, 406–413. doi: 10.1093/cercor/9.4.406

Lee, J., and Maunsell, J. H. (2009). A normalization model of attentional modulation of single unit responses. *PLoS ONE* 4:e4651. doi: 10.1371/journal.pone.0004651

Likova, L. T., and Tyler, C. W. (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D surface representation. *Vision Res.* 43, 2649–2657. doi: 10.1016/S0042-6989(02)00575-8

List, A., and Robertson, L. C. (2007). Inhibition of return and object-based attentional selection. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1322–1334. doi: 10.1037/0096-1523.33.6.1322

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychol. Rev.* 103, 603–649. doi: 10.1037/0033-295X.103.4.603

Mangoun, G. R., and Hillyard, S. A. (1987). The spatial allocation of attention as indexed by event-related brain potentials. *Hum. Factors* 29, 195–211.

Mathot, S., and Theeuwes, J. (2010a). Evidence for the predictive remapping of visual attention. *Exp. Brain Res*. 200, 117–122. doi: 10.1007/s00221-009-2055-3

Mathot, S., and Theeuwes, J. (2010b). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychol. Sci.* 21, 1793–1798. doi: 10.1177/0956797610388813

McKee, S. P., Bravo, M. J., Taylor, D. G., and Legge, G. E. (1994). Stereo matching precedes dichoptic masking. *Vision Res.* 34, 1047-1060.

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nat. Neurosci.* 10, 903–907. doi: 10.1038/nn1917

Melcher, D. (2008). Dynamic, object-based remapping of visual features in trans-saccadic perception. *J. Vis.* 8, 1–17. doi: 10.1167/8.14.2

Melcher, D. (2009). Selective attention and the active remapping of object features in trans- saccadic perception. *Vision Res* 49, 1249–1255. doi: 10.1016/j.visres.2008.03.014

Mountcastle, V. B. (1978). Brain mechanisms of directed attention. *J. R. Soc. Med.* 71, 14–27.

Nakamura, K., and Colby, C. L. (2000). Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area V3A. *J. Neurophysiol.* 84, 677–692.

Nakayama, K., and Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Res.* 30, 1811-1825.

Nakayama, K., and Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature* 320, 264–265.

O'Herron, P. J., and von der Heydt, R. (2013). Remapping of border ownership in the visual cortex. *J. Neurosci.* 33, 1964–1974. doi: 10.1523/JNEUROSCI.2797-12.2013

Olson, I. R., and Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Vis. Cogn.* 9, 273–302. doi: 10.1080/13506280042000162

Paradiso, M. A., and Nakayama, K. (1991). Brightness Perception and Filling-In. *Vision Res.* 31, 1221–1236.

Petersen, S. E., Robinson, D. L., and Morris, J. D. (1987). Contributions of the pulvinar to visual spatial attention. *Neuropsychology* 25, 97–105.

Poggio, G. F. (1991). "Physiological basis of stereoscopic vision," in *Vision and Visual Dysfunction. Binocular Vision*, ed J. R. Cronly-Dillon (Boston, MA: CRC. Press), 224–238.

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Englewood Heights, NJ: Erlbaum.

Posner, M. I. (1988). "Structures and functions of selective attention," in *Master Lectures in Clinical Neuropsychology*,eds. T. Boll and B. Bryant (Washington, D.C: American Psychological Association), 173–202.

Posner, M. I., and Cohen, Y. (1984). "Components of performance," in *Attention and Performance, X*, eds H. Bouma and D. Bowhuis (Hillsdale, NJ: Erlbaum), 531–556.

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42.

Posner, M. I., Walker, J. A., Friedrich, F. J., and Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention. *J. Neurosci.* 4, 1863–1874.

Pouget, A., Dayan, P., and Zemel, R. S. (2003). Inference and computation with population codes. *Annu. Rev. Neurosci.* 26, 381–410. doi: 10.1146/annurev.neuro.26.041002.131112

Pouget, A., and Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nat. Neurosci.* 3(Suppl.), 1192–1198. doi: 10.1038/81469

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. *Cognition* 32, 65–97.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition* 80, 127–158. doi: 10.1016/S0010-0277(00)00156-6

Raizada, R. D., and Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cereb. Cortex* 13, 100–113. doi: 10.1093/cercor/13.1.100

Reynolds, J. H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* 19, 1736–1753.

Reynolds, J. H., and Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron* 37, 853–863. doi: 10.1016/S0896-6273(03)00097-7

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 68–185. doi: 10.1016/j.neuron.2009.01.002

Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron* 26, 703–714. doi: 10.1016/S0896-6273(00)81206-4

Rieke, F., and Rudd, M. E. (2009). The challenges natural images pose for visual adaptation. *Neuron* 64, 605–616. doi: 10.1016/j.neuron.2009.11.028

Robertson, L., and Delis, D. C. (1986). Part-whole processing in unilateral brain damaged patients: dysfunction of hierarchical organization. *Neuropsychology* 24, 363–370.

Rogers-Ramachandran, D. C., and Ramachandran, V. S. (1998). Psychophysical evidence for boundary and surface systems in human vision. *Vision Res.* 38, 71–77.

Rolfs, M., Jonikaitis, D., Deubel, H., and Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nat. Neurosci.* 14, 252–256. doi: 10.1038/nn.2711

Rubin, E. (1921). "Visuell wahrgenommene figuren; studien in psychologischer analyse," in *Copenhagen Gyldendalske Boghandel.*

Saygin, A. P., and Sereno, M. I. (2008). Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. *Cereb. Cortex* 18, 2158–2168. doi: 10.1093/cercor/bhm242

Schiller, P. H., and Lee, K. (1991). The role of the primate extrastriate area V4 in vision. *Science* 251, 1251–1253.

Sereno, M. I., and Huang, R. S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nat. Neurosci.* 9, 1337–1343. doi: 10.1038/nn1777

Sergent, J. (1982). The cerebral balance of power: confrontation or cooperation? *J. Exp. Psychol. Hum. Percept. Perform.* 8, 253–272. doi: 10.1037/0096-1523.8.2.253

Silver, M. R., Grossberg, S., Bullock, D., Histed, M. H., and Miller, E. K. (2011). A neural model of sequential movement planning and control of eye movements: item-order-rank working memory and saccade selection by the supplementary eye fields. *Neural Netw.* 26, 29–58. doi: 10.1016/j.neunet.2011.10.004

Smallman, H. S., and McKee, S. P. (1995). A contrast ratio constraint on stereo matching. *Proc. Biol. Sci.* 260, 265-271. doi: 10.1098/rspb.1995.0090

Sommer, M. A., and Wurtz, R. H. (2006). Influence of the thalamus on spatial vision processing in frontal cortex. *Nature* 444, 374–377. doi: 10.1038/nature05279

Srihasam, K., Bullock, D., and Grossberg, S. (2009). Target selection by frontal cortex during coordinated saccadic and smooth pursuit eye movements. *J. Cogn. Neurosci.* 21, 1611–1627. doi: 10.1162/jocn.2009.21139

Theeuwes, J., Mathôt, S., and Kingstone, A. (2010). Object-based eye movements: The eyes prefer to stay within the same object. *Atten. Percept. Psychophys.* 72, 12–21. doi: 10.3758/APP.72.3.597

Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., and Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron* 29, 757–767. doi: 10.1016/S0896-6273(01)00250-1

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Tyler, C. W., and Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: stereo attention and surface perception in depth reconstruction. *Perception* 24, 127–153. doi: 10.1068/p240127

Umeno, M. M., and Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye fields, I: predictive visual responses. *J. Neurophysiol.* 78, 1373–1383.

Van Oeffelen, M. P., and Vos, P. G. (1982). Configurational effects on the enumeration of dots: counting by groups. *Mem. Cogn.* 10, 396–404. doi: 10.3758/BF03202432

Van Oeffelen, M. P., and Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recogn.* 16, 341–348. doi: 10.1016/0031-3203(83)90040-7

Vergilino-Perez, D., and Findlay, J. M. (2004). Object structure and saccade planning. *Brain Res. Cogn. Brain Res.* 20, 525–528. doi: 10.1016/j.cogbrainres.2004.03.022

Von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik.* Leipzig: Voss.

Von Holst, E., and Mittelstaedt, H. (1950). Das reafferenz princip: wedlselwirkungen zwishchen Zentrainervensystem und Peripherie. *Die Naturwissenschaften* 37, 464–476. doi: 10.1007/BF00622503

Webster, M. J., Bachevalier, J., and Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb. Cortex* 4, 470–483. doi: 10.1093/cercor/4.5.470

Wolfe, J. M., (Ed.). (2007). *Guided search 4.0: Current Progress with a Model of Visual Search*. New York, NY: Oxford.

Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433. doi: 10.1037/0096-1523.15.3.419

Wolfe, J. M., Friedman-Hill, S. R., and Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Percept. Psychophys.* 55, 537–550. doi: 10.3758/BF03205311

Wurtz, R. H., Goldberg, M. E., and Robinson, D. L. (1980). Behavioral modulation of visual responses in monkeys. *Prog. Psychobiol. Physiol. Psychol.* 9, 42–83.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., et al. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nat. Neurosci.* 5, 995–1002. doi: 10.1038/nn921

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: 04 June 2014; accepted: 28 November 2014; published online: 14 January 2015.

Citation: Grossberg S, Srinivasan K and Yazdanbakhsh A (2015) Binocular fusion and invariant category learning due to predictive remapping during scanning of a depthful scene with eye movements. Front. Psychol. ***5****:1457. doi: 10.3389/fpsyg.2014.01457*

This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.

Copyright © 2015 Grossberg, Srinivasan and Yazdanbakhsh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 02 July 2014
doi: 10.3389/fpsyg.2014.00612

Perceptions of document relevance

Peter Bruza* and Vivien Chang

Science and Engineering Faculty, Information Systems School, Queensland University of Technology, Brisbane, QLD, Australia

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Emmanuel Pothos, City University London, UK
Falk Scholer, RMIT University, Australia

****Correspondence:***
Peter Bruza, Science and Engineering Faculty, Information Systems School, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia
e-mail: p.bruza@qut.edu.au

This article presents a study of how humans perceive and judge the relevance of documents. Humans are adept at making reasonably robust and quick decisions about what information is relevant to them, despite the ever increasing complexity and volume of their surrounding information environment. The literature on document relevance has identified various dimensions of relevance (e.g., topicality, novelty, etc.), however little is understood about how these dimensions may interact. We performed a crowdsourced study of how human subjects judge two relevance dimensions in relation to document snippets retrieved from an internet search engine. The order of the judgment was controlled. For those judgments exhibiting an order effect, a *q*–test was performed to determine whether the order effects can be explained by a quantum decision model based on incompatible decision perspectives. Some evidence of incompatibility was found which suggests incompatible decision perspectives is appropriate for explaining interacting dimensions of relevance in such instances.

Keywords: document relevance, quantum cognition, information retrieval, cognitive modeling, user modeling

1. INTRODUCTION

This article aims to shed light on how humans judge the relevance of documents. We will, however, take a modern view of what a document is. Nowadays individuals and groups interact with one another in a variety of information environments of ever increasing complexity. They are accessing search engines, sharing messages on Facebook, browsing short messages on their mobile devices from microblog sites like Twitter. In this setting, a document is usually very short, e.g., a Twitter post, or in some cases it is not a document at all, but rather a document surrogate, such as the query-biased summaries (snippets) of documents displayed in rankings produced by search engines.

Document relevance has been carefully studied over more than three decades within the fields of information science usually by identifying or employing known inter-subjective dimensions of relevance (Schamber et al., 1990; Barry, 1994; Mizzaro, 1997; Borlund, 2003). For example, Barry and Schamber (1998) identified the dimensions "presentation quality," "currency," "reliability," "verifiability," "geographic proximity," "specificity," "dynamism" and "accessibility" in a comprehensive study. A recent study examined how users determined which list of search results they preferred over another using five dimensions of relevance: "topicality," "freshness" (currency), "authority" (credibility), "caption quality," and "diversity" (Kim et al., 2013). Other dimensions have also been identified with respect to a particular genre document. For example, Chu (2012) identified the dimensions "specificity," "ease of use" and "breadth" in the context of legal documents.

Whilst it is widely accepted that there are a variety of dimensions at play when it comes to judging relevance, little is known of how these dimensions may interact. The aim of this article is to adopt a decision theoretic perspective and test a novel cognitive decision model in which potential interactions between dimensions are a consequence of incompatible decision perspectives which impose an order effect on relevance judgments. Incompatible perspectives are a recent development in a field called "quantum cognition" (See, for example, Conte et al., 2007; Aerts, 2009; Bruza et al., 2009; Pothos and Busemeyer, 2009; Atmanspacher and Filk, 2010; Khrennikov, 2010; Busemeyer et al., 2011; Conte et al., 2011; Trueblood and Busemeyer, 2011; beim Graben et al., 2012; Busemeyer and Bruza, 2012; Conte, 2012; Dzhafarov and Kujala, 2012; Aerts et al., 2013; Blutner et al., 2013; Haven and Khrennikov, 2013). This field aims to apply the formalism of quantum theory in order to more adequately model cognitive phenemona. For example, decades of research have uncovered a whole spectrum of human judgment that deviates substantially from what would be normatively correct according to logic and probability theory. An example of the latter is the following:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrationsÓ. Which is more probable:
> (a) Linda is a bank teller, or
> (b) Linda is a bank teller and is active in the feminist movement?

In this now famous experiment proposed by Tversky and Kahneman (1983), human subjects consistently rate option (b) as more probable than (a). However, according to probability theory, the probability of a conjunction of events must be less than or equal to the probability of a constituent event. Thereofore, according to the axioms of probabilty theory (b) is less probable than (a). Probability judgment errors of this nature have since become known as the "conjunction fallacy."

The key to explaining the conjunction fallacy using a quantum model is the *incompatibility* between the perspective that Linda is a bank teller and her being a feminist. Consider **Figure 1A**. The perspective "Linda is a feminist" is represented as a two dimensional vector space where the basis vector F corresponds to the decision "Linda is a feminist" and $\bar{F}$ corresponds to "Linda *not* being a feminist." A similar two dimensional vector space corresponds to the perspective of Linda being a bank teller B, or not $\bar{B}$. Initially, the cognitive state of the subject is represented by the vector Ψ, which is suspended between both sets of basis vectors. This situation represents the subject being undecided about whether Linda is a bank teller or a feminist. Suppose the subject now decides that Linda is a feminist. This decision is modeled by Ψ "collapsing" onto the basis vector labeled F. (The probability of the decision corresponds to the square of the length of the projection of the cognitive state Ψ onto the basis vector F, denoted $\|\mathbf{P_F}\psi\|^2$). Observe how the subject is now necessarily uncertain about Linda being a bank teller because the basis vector F is suspended between the two basis vectors B and $\bar{B}$ by the angle θ. The hall mark of incompatibility is the state of indecision from one perspective (e.g., the bank teller perspective) when a decision is taken from another (e.g., the feminist perspective). This indecision means the decision maker can't form the joint probability of Linda being both a feminist and a bank teller, $\Pr(F,\ B)$ (Busemeyer et al., 2011). (This is crucially different to the situation in standard probability theory in which events are compatible, and thus the joint probability is always defined).

The consequence of incompatibility is an interference term denoted Int. The partial derivation below shows that this term Int appears when the decision of whether Linda is a feminist is made in relation to the incompatible subspace corresponding to the decision perspective of her being a bank teller (represented by projector $\mathbf{P_B}$ and its dual $\mathbf{P_B}^{\perp}$):

$$p(F) = \|\mathbf{P_F}\psi\|^2 \quad (1)$$

$$= \|(\mathbf{P_F} \cdot \mathbf{I})\psi\|^2 \quad (2)$$

$$= \|(\mathbf{P_F} \cdot (\mathbf{P_B} + \mathbf{P_B^{\perp}})\psi\|^2 \quad (3)$$

$$= \|\mathbf{P_F}\mathbf{P_B}\psi\|^2 + \|\mathbf{P_F}\mathbf{P_B^{\perp}}\psi\|^2 + \text{Int} \quad (4)$$

The intuition behind Equation 4 is that the law of total probability is being modified by the interference term. In probability theory this would be expressed as follows: $p(F) = p(F, B) + p(F, \bar{B}) +$ Int. When the interference term is zero, the law of total probability holds. This happens when the decision perspectives are compatible.

Incompatible decision perspectives are a recent development in cognitive modeling and their striking characteristic is the use of "quantum" probabilities. By quantum probabilities, we mean that the decision event space is modeled as a vector space rather than a Boolean algebra of sets. A key differentiator is the use of the interference term. When this term is non-zero, violations of the law of total probability occur. The interference term has been used in models of the perception of gestalt images (Conte et al., 2007; Khrennikov, 2010), models of the conjunction and other decision fallacies (Busemeyer et al., 2011; Conte et al., 2011), modeling violations of rational decison theory (Bordley, 1998; Pothos and Busemeyer, 2009; Khrennikov, 2010), modeling belief dynamics (Trueblood and Busemeyer, 2011) and conceptual processing (Gabora and Aerts, 2002; Gabora et al., 2008; Aerts, 2009; Aerts et al., 2013; Blutner et al., 2013). Broader works relate the formal structures used in quantum theory to cognition and other areas (Bruza et al., 2009; Khrennikov, 2010; Busemeyer and Bruza, 2012; Conte, 2012; Haven and Khrennikov, 2013).

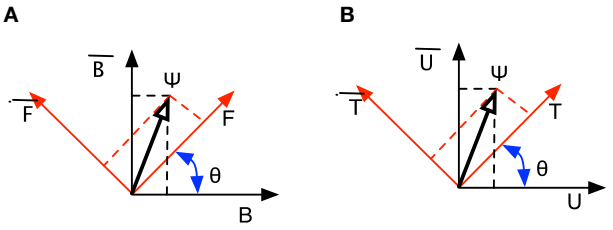


FIGURE 1 | Incompatible perspectives in a relevance judgment. (A) Incompatible perspectives in the Linda example. **(B)** Incompatible perspectives in judging document relevance.

Consider **Figure 1B** which has the same structure as the Linda problem depicted in **Figure 1A**. This figure comprises two perspectives regarding a decision of document relevance. Assuming that a human subject perceives a document's relevance via different perspectives in relation to their given information need, the "topicality" perspective is represented as a two dimensional vector space where the basis vector T corresponds to the decision "the information is topically related to the information need" and $\bar{T}$ corresponds to "the information is *not* topically related to the information need." A similar two dimensional vector space corresponds to the perspective of the information being understandable U, or not $\bar{U}$, to the human subject. Initially, the cognitive state of the human subject is represented by the vector Ψ, which is suspended between both sets of basis vectors. This situation represents the subject being undecided about whether the information being perused is topical or understandable. Suppose the subject now decides that the information is topical. This decision is modeled by Ψ "collapsing" onto the basis vector labeled T. Once again, the probability of the decision corresponds to the square of the length of the projection of the cognitive state Ψ onto the basis vector T, denoted $\|\mathbf{P}_T\psi\|^2$.

Observe how the subject is now necessarily uncertain about whether the information is understandable because the basis vector T is suspended between the two basis vectors U and $\bar{U}$ by the angle θ. The intuition behind incompatibility in this case is that the subject may be confident in deciding the information is topically relevant but remain in two minds about whether they understand the information, for example, if the snippet is interspersed with specialized technical vocabulary as in **Figure 2**. An important consequence of incompatible decision perspectives is an order effect. In the context of the example, this means the probability of judging that the information is relevant differs when first considering "topicality" followed by "understandability" compared to when these decisions are reversed. This is because when decision perspectives are incompatible, projections do *not* commute, i.e., $\|\mathbf{P}_U\mathbf{P}_T\psi\|^2 \neq \|\mathbf{P}_T\mathbf{P}_U\psi\|^2$.

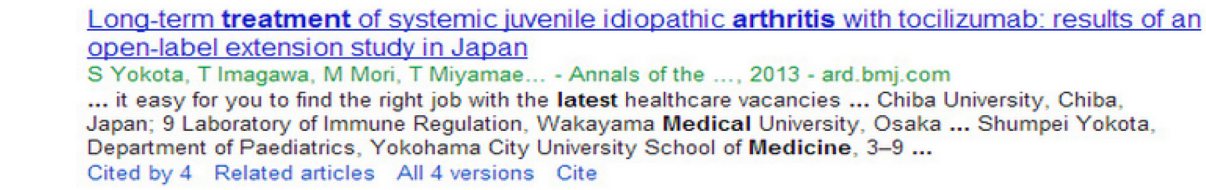


FIGURE 2 | Example document snippet.

The preceding should not be taken to imply that all relevance judgments are modeled in terms of incompatible decision perspectives. In some cases, the perspectives may be compatible. For example, the subject can make a decision that the document is topically relevant and then also be certain in regard to their decision about the document's understandability. In formal terms, compatible decision perspectives entail that the projectors commute, i.e., $\|\mathbf{P}_U\mathbf{P}_T\psi\|^2 = \|\mathbf{P}_T\mathbf{P}_U\psi\|^2$.

The focus is this artice is to explore whether there is evidence for incompatible decision perspectives. The question then becomes how to determine whether the model presented in **Figure 1B** explains decisions of document relevance. Wang and Busemeyer (2013) have recently proposed an innovative solution to this question. They proved that if there is an order effect and a so called $q-$test holds, then a model based on incompatible decision perspectives like those depicted in **Figure 1B** is a valid cognitive decision model . In terms of our example, the $q-$test has the following form based on yes(y)/no(n) answers regarding "topicality" and "understandability":

$$p(TyUn) + p(TnUy) = p(UyTn) + p(UnTy) \tag{5}$$

Let $p_{TU} = p(TyUn) + p(TnUy)$ define the probability of different answers when "topicality" T is asked first, followed by "understandability" U. Conversely, let $p_{UT} = p(UyTn) + p(UnTy)$ be the probability of different answers when the order of questions is first "understandability" followed by "topicality." The $q-$test has the following form:

$$q = p_{AB} - p_{BA} = 0 \tag{6}$$

The advantage of the $q-$test is that it is a parameter free test. It has successfully been applied to motivate a quantum model in relation to order effects in political survey data (Wang and Busemeyer, 2013). In this article, we will examine: (1) whether there are order effects in relation to decisions pertaining to specific dimensions of relevance, and (2) whether a quantum model based on incompatible decision perspectives explains these order effects.

2. MATERIALS AND METHODS

2.1. SUBJECTS

Relevance judgments were crowdsourced by the internet based Amazon's Mechanical Turk platform. Crowdsourcing is the outsourcing of tasks to an undefined, large group of people. In the case of Amazon's Mechanical Turk, crowdsourcing is a means of gathering data from users via "human intelligence tasks" (HITs) which are typically surveys for subjects, or "turkers" to answer. Turkers are paid a nominal fee, in this case between 12 and 20 cents per relevance judgment. If the data from the turker is deemed of sufficient quality, the owner of the HIT approves the payment. The quality of the data can be determined automatically by the system whereby after a set period of time, say an hour, then the data will be approved whereby the turker will be paid. This process can also be done manually before and after approval; thus increasing the quality of data collected. In this experiment, the data were manually approved.

The advantages of crowdsourcing is that data can be collected quickly, on a fairly large scale and at a reasonable price. The disadvantage is the extra effort needed in order to safeguard the quality of the data. As Mechanical Turk is internet based, there is little control over who the turkers are, where they are, and indeed, whether they are even human. For example, "bots," i.e., software programs mimicking humans are known to take part and more or less randomly contribute data to an experiment. As a consequence, the quality of crowdsourced data can vary greatly. To combat this, we purposefully inserted questions in the HITS to collect qualitative data—a technique often used in crowdsoured experiments.

Furthermore, as an additional factor to ensure quality data, both "masters" as well as "normal" turkers were used. Masters have "demonstrated excellence" in performing crowdsourced experiments over an extended period with a required HIT Approval Rate of above 95% over at least one thousand HITs. In contrast to the "masters," nothing much is known of regarding the performance of "normal" turkers. The experiment was timed to primarily source U.S. based turkers, who are thus likely to be proficient in English, however no tests were conducted to verify English proficiency.

2.2. MATERIALS

The materials comprised queries and information in the form of document snippets.

Five queries were developed for this study, each of which is based around an information need, for example, see **Figure 3**. The query description comprises the name of a query topic, a short description and an accompanying narrative. The narrative is intended to frame the subject's perception of relevance. There is a possibility that the turker's background may interfere with the narrative around the query. For example, if the turker is a fan of technology, then there is significant likelihood that they will be biased toward specific information or brands of technology. The

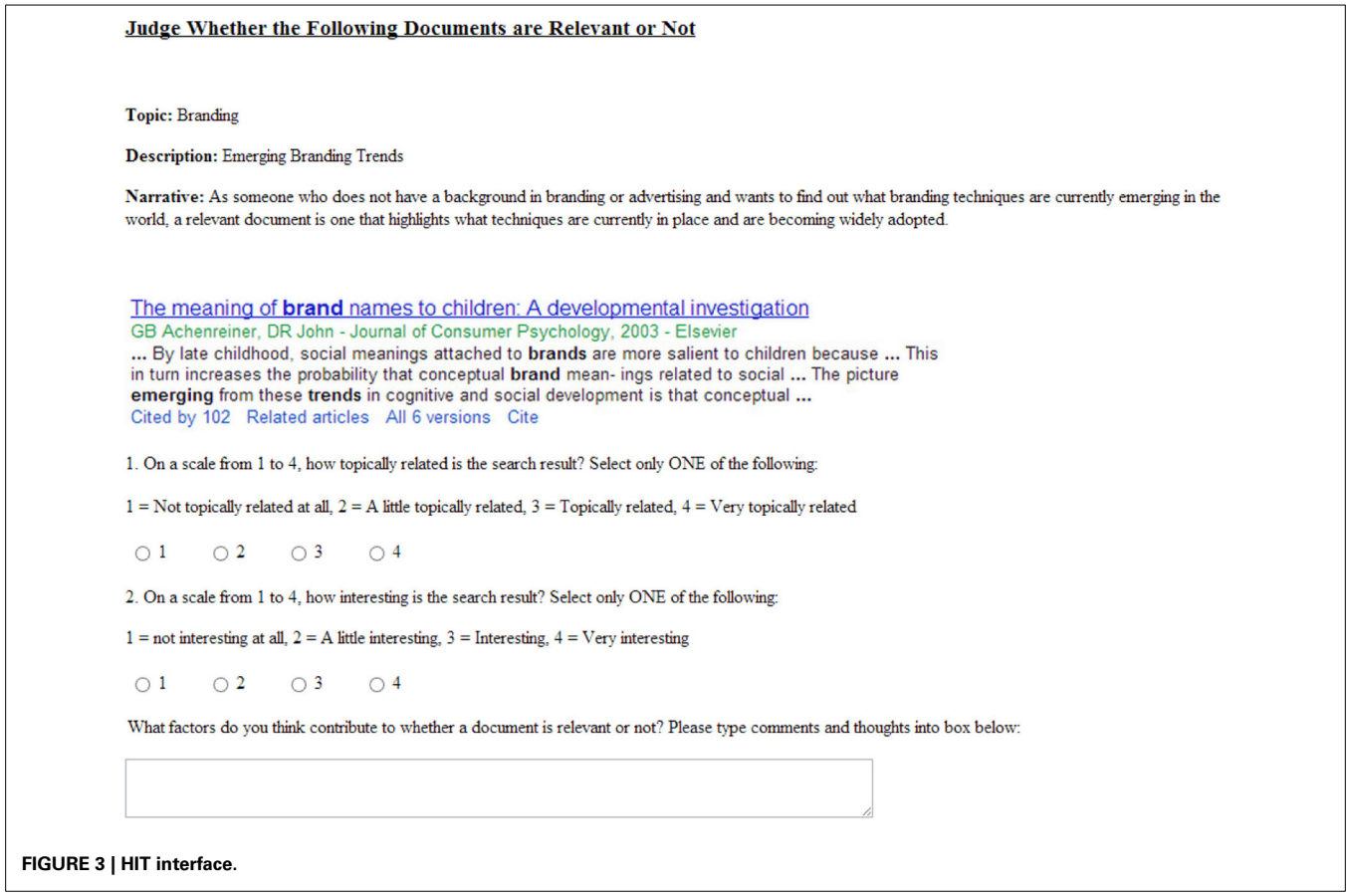


FIGURE 3 | HIT interface.

experimenters viewed that bias is intrinsic to search and therefore did not to try to compensate for it (White, 2013). In addition, the background of the turker may hinder their ability to sufficiently engage with the narrative. However there was evidence via the qualitative feedback questions that turkers were able to role-play in a satisfactory way, particularly the "masters." For example, "..a little hard to determine what this is talking about and if I were a beginner I would have no clue." or "...makes [the] document highly relevant, since the focus is for emerging technologies in 2013." Finally, the narrative structure of the queries was adopted from long running Text Retrieval Conference Series run yearly by the U.S. National Institute of Standards and Technology[1]. Each query was designed to collect judgments pertaining to two specific dimensions of relevance chosen by the authors. **Table 1** details the titles of the queries and the dimensions of relevance which were studied.

The relevance dimensions studied are further detailed in **Table 2**. "Topicality" has been chosen as a primary dimension to be examined across all queries because this dimension has been consistently identified in previous studies as a primary factor in relevance judgments (e.g., Barry and Schamber, 1998; Borlund, 2003; Chu, 2012). In addition, search engine algorithms are based on queries and finding a match in regards to keywords as a matter or correlating topically related material.

[1]http://trec.nist.gov/

Table 1 | Queries and relevance dimensions.

Query Title	Dimension 1	Dimension 2
Treatment for arthritis	Topicality	Understandability
Emerging branding trends	Topicality	Interest
Emerging technology	Topicality	Credibility
Causes of global warming	Topicality	Believability
Influence of media on the Vietnam war	Topicality	Sentimentality

Secondary dimensions depend on the query. Once the queries had been established, the authors designated likely secondary dimensions. Through pilot studies, the choice for the secondary dimension was refined when other factors began to creep into turker's comments. For example, during initial stages of the pilot, one of the first HITs published was the "Emerging Technology" query involving the dimensions of "topicality" and "understandability." Very quickly, it was realized that "credibility" was a factor that was constantly brought up by turkers in qualitative feedback. This was possibly also due to the advancement and ubiquity of technology thus rendering "understandability" as not an issue. Other secondary factors were chosen in a similar fashion while some were heavily dependent on the query topic

Table 2 | Definition of relevance dimensions.

Relevance dimension	Definition
Topicality	How topically related is the information in the snippet to the query
Credibility	How credible is the information in the snippet
Understandability	How easy is it to understand the information in the snippet
Believability	How believable is the information in the snippet
Interest	How novel/entertaining/interesting is the information in the snippet
Sentimentality	The degree of affective response to the information in the snippet

at hand. For example, the topic of global warming is one involving fixed dichotomous positions e.g., people either believe that this is occuring or they don't. Therefore, "believability" seemed likely to be a prominent relevance dimension in this case.

Secondary dimensions that were chosen for study are listed in the column labeled "Dimension 2" of **Table 1**. "Understandability" was chosen as snippets can sometimes be full of technical jargon, acronyms or specialized terms that can be challenging for the average person to comprehend. The dimension of "Believability" stems from a subject's personal beliefs and biases in relation to the information. A recent study showed that users were subject to their own biases as well as biases inherent in the search engine (White, 2013). "Interest" is the dimension of relevance pertaining to how novel or entertaining the information is. "Sentimentality" is a dimension which pertains to emotional responses to information. Sentiment analysis is a very active area of research in relation to internet-based technologies and applications, for example, data mining techniques to identify positive or negative sentiments or opinions in product reviews.

Corresponding to each query was a query-biased summary of a document, which we will refer to as a document "snippet." (See **Figure 2**). Document snippets were used as these are an increasingly prevalent form of information on which decisions of relevance are made in relation to modern information environments. The document snippets used in this study were sourced from the Google search engine.

Snippets were selected based on the likelihood that decisions regarding the two chosen dimensions of relevance were likely to involve some uncertainty. This is because we hypothesize that incompatibility between these dimensions is more likely to occur when such uncertainty was present. Unfortunately, there is no theory to predict which dimensions may be incompatible so a crowsourced pilot study was conducted. This study involved 10 snippets per query with between 8 and 10 master turkers making judgments in each order condition. In order to verify that uncertainty was present a four point rating scale was used to collect decisions. For each query, the snippet for which the $q-$ test was closest to zero was selected as being most likely to be subject to incompatibility. None of the subjects in the pilot took part in the experiment presented here. This could easily be verified as each turker has a unique identifier.

2.3. PROCEDURE

The experiment (i.e., the HIT) consisted of five elements which were presented in sequence. Each element was based around a query, and a subject was required to process all five elements.

Each element comprises the query description followed by a document snippet, two judgments and finally the input of qualitative data. **Figure 3** depicts one such element. In each judgment a subject is asked to rate a dimension of relevance on a four point scale. It was assumed that a subject can make judgments on dimensions within a given query topic independently of other query topics.

A single factor design was employed where the order of the judgments was manipulated. For example, in one condition a given dimension, e.g., "topicality" is rated first (the "non-comparative" context for the decision on topicality), followed by a rating of a the "understability" dimension. In the second condition, the order of the ratings is reversed e.g., the rating on "topicality" is second after the "understandability" dimension is rated (the "comparative" context for the decision on topicality). As each turker has a unique identifier, those turkers who attempted both conditions were removed from the data.

Subsequent to the judgments, subjects were asked to comment on factors that influenced their judgments. This aspect served for both quality control as well as a source of qualitative data to better undertsand the factors involved when turkers make judgments. By doing so, we discarded the data from any turker where the answers were blank, superfluous, e.g., "this is very good and gainful," or didn't make sense, e.g., "The sway there marketed with different topics." In the event that qualitative data were borderline acceptable such as "don't know," or "not sure" (both of which could be supplied by a bot), the time taken to complete the HIT was also taken into consideration: If the time spent was less than 50 s for the HIT, the data were also discarded as we deemed a minimum of 10 s per query as being required to meaningfully read the query topic, rate two dimensions and supply qualitative feedback.

Finally, the Mechanical Turk interface does not afford the ability to time a turker per query, so the time taken to make judgments in relation to a given query could not be collected for analysis.

3. RESULTS

A total of fifty "normal" turkers submitted data for the condition where the "topicality" dimension was presented first (non-comparative context for topicality), of which eighteen were discarded. Conversely, thirty-six "normal" turkers submitted data for the comparative context of "topicality," of which four were discarded. This left $n = 32$ subjects in each condition. Despite repeated attempts to recruit "master" turkers, we failed to secure numbers sufficient for reliable statistical analysis. Therefore, their rating data are not reported but some qualitative responses were retained for illustrative purposes.

The results are presented in yes/no contingency tables in order for the $q-$test to be applied. This was achieved by mapping the four point graded relevance judgments to yes/no decisions in the following way: A grade of 3 or 4 was translated to a "yes," whereas a grade of 1 or 2 was translated to a "no." For example, consider **Figure 3**. Using the proposed mapping, a topical judgment of "4 = Very topically related" and "3 = Topically related" translate into a decision of "yes." After the yes/no mapping, contingency tables can be constructed for each decision and these are presented in **Figure 4** for the "normal" turkers. Some of the queries have data with less than 32 subjects as for these queries a turker rated one dimension, without rating the other. In such cases, the data for that query were omitted.

In order to apply the $q-$test, the presence of an order effect must first be established. An order effect is determined by comparing the agreement rates obtained in a non-comparative vs. a comparative context. An order effect occurs when the proportion of subjects who decided "yes" differs significantly in the comparative vs. non-comparative contexts. A two-tailed $\chi-$square test of equality of proportions between populations was carried out ($\alpha = 0.05$) and those queries exhibiting an order effect are bolded in **Table 3**.

Based on the contingency tables presented in **Figure 4**, the $q-$ test values for the "normal" turkers were computed using equation (6) and presented in **Table 3**.

4. DISCUSSION

For the query topics where there is an order effect, the quantum model based on incompatible decision perspectives predicts $q = 0$ (Wang and Busemeyer, 2013). **Table 3** accords with this prediction for the queries "Treatment for Arthritis" and "Causes of Global Warming." However, there are two other queries exhibiting an order effect but for which $q \neq 0$. In these cases, the prediction of the quantum model may not actualize due to the quite small sample sizes in both conditions, or that the quantum model is not a valid explanation for these queries. More experimentation with larger sample sizes is needed to resolve this distinction.

Four out of five queries displayed an order effect ($\alpha = 0.05$). The presence of an order effect means that the subjects' decision cannot be validly modeled by a joint probability distribution spanning binary variables corresponding to the underlying dimensions of relevance. For example, consider **Figures 4A,B**. In the non-comparative context for a decision on topicality, the

A

T \ U	y	n
y	0.4063	0.2813
n	0.1563	0.1563

Arthritis: Topicality-Understandability ($n = 32$)

B

U \ T	y	n
y	0.1936	0.0323
n	0.3871	0.3871

Arthritis: Understandability-Topicality ($n = 31$)

C

T \ I	y	n
y	0.2903	0.1613
n	0.2581	0.2903

Branding: Topicality-Interest ($n = 31$)

D

I \ T	y	n
y	0.2581	0.4194
n	0.0968	0.2259

Branding: Interest-Topicality ($n = 31$)

E

T \ C	y	n
y	0.3226	0.0978
n	0.2258	0.3548

Emerging Technology: Topicality-Credibility ($n = 31$)

F

C \ T	y	n
y	0.4063	0.3125
n	0.0938	0.1875

Emerging Technology: Credibility-Topicality ($n = 32$)

G

T \ B	y	n
y	0.6875	0.0313
n	0.1875	0.0938

Global Warming: Topicality-Believability ($n = 32$)

H

B \ T	y	n
y	0.5667	0.1667
n	0.0667	0.2

Global Warming: Believability-Topicality ($n = 30$)

I

T \ S	y	n
y	0.2581	0.4194
n	0.129	0.1936

Vietnam: Topicality-Sentimentality ($n = 31$)

J

S \ T	y	n
y	0.2813	0.0938
n	0.375	0.25

Vietnam: Sentimentality-Topicality ($n = 32$)

FIGURE 4 | Yes/no contingency tables from "normal" turkers. The left hand side represents the condition where topicalilty is decided first (Non-comparative context for a decision on topicality). The right hand side represents the condition where topicality is decided second (Comparative context for a decision on topicality).

Table 3 | Summary table of *q*−test values.

Query title	Dimension 1	Dimension 2	*q*
Treatment for arthritis	Topicality	Understandability	0.02†
Emerging branding trends	Topicality	Interest	0.10
Emerging technology	Topicality	Credibility	0.08
Causes of global warming	Topicality	Believability	0.01†
Influence of media of the Vietnam war	Topicality	Sentimentality	0.08

Queries with order effect (α =0.05) are bolded. Queries where q−test holds (α =0.05) are flagged by †.

marginal probability that the document is topical is summed across understandability:

$$p(T = y) = p(T = y, U = y) + p(T = y, U = n) \quad (7)$$

$$= 0.4063 + 0.2813 \quad (8)$$

$$= 0.6876 \quad (9)$$

Note that this probability is significantly different ($\alpha = 0.05$) when understandability provides the comparative context for deciding topicality: $p(T = y) = 0.1936 + 0.3871 = 0.5807$. It is this difference which identifies an order effect *but* as the marginal probability is not constant, it is not possible to construct a single joint probability distribution $p(T, U)$ to model the relevance decisions. As a consequence, a common modeling approach is ruled out. This approach assumes $p(T, U)$ exists whereby the decision in the non-comparative context around topicality is modeled by the marginal probability $p(T)$ and the decision in the comparative context is modeled by conditioning the distribution based on how understanding was first decided, i.e., $p(T|U = y)$ or $p(T|U = n)$.

In summary, order effects were detected between dimensions of relevance for the majority of queries and some evidence that a quantum model based on incompatible decision perspectives is a valid explanation. However, this evidence is not yet strong. Experiments with larger sample sizes and a larger collection of queries and snippets are required to determine the prevalence of incompatible perspectives in perceptions of document relevance. It should be mentioned, however, that this study differentiates itself from many previous studies in that a much larger number of subjects were involved. For example, nine subjects provided relevance judgments in Chu (2012).

According to Cooper (1971) the concept of relevance comprises both "logical relevance" and "utility." Logical relevance is defined as "whether or not a piece of information is on a subject which has some topical bearing on the information need" and utility has to do with "the ultimate usefulness of the piece of information." It seems that perceptions of utility or usefulness of a particular snippet involves cognitive processing of a variety of factors including those dimensions examined in this study. It became apparent from the qualitative feedback that relevance is a multifaceted, dynamic decision process. For example, in the "Global Warming" query, "reputation," "credibility" and "scientific" were used to describe factors that the turkers themselves ranked highly compared to "believability" which was the chosen secondary dimension. This could suggest that the dimensions of "credibility" and "believability" mentioned as being distinct in previous studies are in fact hardly distinguishable during some relevance decisions. Not only were there more than a few factors at play, but the dimensions of "topicality" and "understandability" were featured in qualitative feedback across all queries. Furthermore, comments mentioning multiple (i.e., greater than two) factors were reasonably common. For example, one turker elegantly wrote "whether it (the search result) is on topic, credible, and goes into sufficient detail." Interestingly, many of these comments noted "topicality" in ways that suggested that even though a snippet was topically related, this did not necessarily translate to the snippet being deemed relevant. This was a shift from the pilot study where turkers would state very clearly in their comments that topicality was nearly always the first factor they considered and if a snippet was topically related, then they would judge it to be relevant. The shift may have been due to the final design in which turkers processed five different queries which exposed them to a broader spectrum of relevance dimensions than was the case in the pilot study. Such qualitative feedback calls the experimental design into question, namely, is it methodologically sound to focus the subjects' attention on two dimensions when more are at play? In addition, were these extra dimensions coming into play because the subject was learning about relevance as they proceeded through the queries? The experimental design did not control for such a learning effect as it was assumed that each query topic could be judged independently of the others. An alternate design would allow subjects to select the two dimensions they deem most prominent and then rate them, or only allow subjects to rate a singe query topic.

5. CONCLUSION

This article put forward an experimental framework for examining whether dimensions of relevance interact via an order effect. The data collected from a crowdsourced study suggests that in some decisions regarding dimensions of relevance, this interaction can be explained in terms of a quantum model based on incompatible decision perspectives. Assuming that such interactions are fairly prevalent, what are the consequences? Currently in information processing systems, such as search engines, there is a general lack of effective user models. Should the user be making decisions of relevenance based on incompatible decision perspectives, then a model of the user based on standard probability would not be appropriate. The field of quantum cognition has shown that incompatibility implies that the law of total probability does not hold. Current computational systems are founded on standard probability theory. For example, consider the corpus-based computational model proposed by Lin and He (2009). This model takes the dimensions of both "topicality" and "sentiment" into account.

At the heart of the model is the following factorization: $p(\mathbf{w}, \mathbf{z}, \mathbf{s}) = p(\mathbf{w}|\mathbf{z}, \mathbf{s})p(\mathbf{z}, \mathbf{s})$, where $\mathbf{w}$ is a random variable over a vocabulary of terms extracted from the corpus, $\mathbf{z}$ is a random variable over a set of latent topics, and $\mathbf{s}$ is a random variable

over a set of sentiment labels (e.g., a binary variable describing a positive or negative sentiment). Note at its foundation, the model relies on the joint probability $p(\mathbf{z}, \mathbf{s})$, which describes the joint probability over topics and sentiments. In other words, the model assumes that "topicality" and "sentiment" are *compatible*. Should incompatibility manifest in the user's cognition, such a joint probability is undefined. This opens the door for dissonance between the relevance decisions made by the system as opposed to those made by the user. In short, the presence of incompatible decision perspectives suggests users can better be modeled by a "non-classical" probability theory like that proposed by the field of quantum cognition.

ACKNOWLEDGMENT

The authors would like to thank Jerome Busemeyer for his advice on the application of the $q-$test as well as the helpful comments of the reviewers.

REFERENCES

Aerts, D. (2009). Quantum structure in cognition. *J. Math. Psychol.* 53, 314–348. doi: 10.1016/j.jmp.2009.04.005

Aerts, D., Gabora, L., and Sozzo, S. (2013). Concepts and their dynamics: a quantum-theoretic modeling of human thought. *Top. Cogn. Sci.* 5, 737–772. doi: 10.1111/tops.12042

Atmanspacher, H., and Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. *J. Math. Psychol.* 54, 314–321. doi: 10.1016/j.jmp.2009.12.001

Barry, C. (1994). User-defined relevance criteria: an exploratory study. *J. Am. Soc. Inform. Sci.* 45, 145–159.

Barry, C., and Schamber, L. (1998). Users' criteria for relevance evaluation: a cross-situational comparison. *Inform. Process. Manag.* 34, 219–236.

beim Graben, P., Filk, T., and Atmanspacher, H. (2012). Epistemic entanglement due to non-generating partitions of classical dynamical systems. *Int. J. Theor. Phys.* 52, 723–734. doi: 10.1007/s10773-012-1381-6

Blutner, R., Pothos, E., and Bruza, P. (2013). A quantum probability perspective on borderline vagueness. *Top. Cogn. Sci.* 5, 711–736. doi: 10.1111/tops.12041

Bordley, R. (1998). Violations of compound probability principles: toward a generalized Heisenberg uncertainty principle. *Operat. Res.* 46, 923–926. doi: 10.1287/opre.46.6.923

Borlund, P. (2003). The concept of relevance in IR. *J. Am. Soc. Inform. Sci. Tech.* 54, 913–925. doi: 10.1002/asi.10286

Bruza, P., Busemeyer, J., and Gabora, L. (2009). Introduction to the special issue on quantum cognition. *J. Math. Psychol.* 53, 303–305. doi: 10.1016/j.jmp.2009.06.002

Busemeyer, J., and Bruza, P. (2012). *Quantum Cognition and Decision*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511997716

Busemeyer, J., Pothos, E., Franco, R., and Trueblood, J. (2011). A quantum theoretical explanation for probability judgment errors. *Psychol. Rev.* 118, 193–218. doi: 10.1037/a0022542

Chu, H. (2012). Factors affecting relevance judgment: a report from TREC legal track. *J. Document.* 67, 264–287. doi: 10.1108/00220411111109467

Conte, E. (2012). *Advances in Application of Quantum Mechanics in Neuroscience and Psychology: A Clifford Algebraic Approach*. New York, NY: Nova Science Publishers.

Conte, E., Khrennikov, A., Todarello, O., De Robertis, R., Federici, A., and Zbilut, J. (2011). On the possibility that we think in a quantum mechanical manner: an experimental verification of existing quantum interference effects in cognitive anomaly of conjunction fallacy. *Chaos Complex. Lett.* 4, 123–136.

Conte, E., Todarello, O., Federici, A., Vitiello, F., Lopane, M., Khrennikov, A., et al. (2007). Some remarks on an experiment suggesting quantum-like behaviour of cognitive entities and formulation of an abstract quantum mechanical formalism to describe cognitive entity and dynamics. *Chaos Solit. Fract.* 33, 1076–1088. doi: 10.1016/j.chaos.2005.09.061

Cooper, W. (1971). A definition of relevance for information retrieval. *Infor. Storgae Retrieval* 7, 19–37. doi: 10.1016/0020-0271(71)90024-6

Dzhafarov, R., and Kujala, J. (2012). Selectivity in probabilistic causality: where psychology runs into quantum physics. *J. Math. Psychol.* 56, 54–63. doi: 10.1016/j.jmp.2011.12.003

Gabora, L., and Aerts, D. (2002). Contextualizing concepts using a mathematical generalization of the quantum formalism. *J. Exp. Theor. Artif. Intell.* 14, 327–358. doi: 10.1080/09528130210162253

Gabora, L., Rosch, E., and Aerts, D. (2008). Toward an ecological theory of concepts. *Ecol. Psychol.* 20, 84–116. doi: 10.1080/10407410701766676

Haven, E., and Khrennikov, A. (2013). *Quantum Social Science*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139003261

Khrennikov, A. (2010). *Ubiquitous Quantum Structure: From Psychology to Finance*. Heidelberg; Dordrecht; London; New York, NY: Springer. doi: 10.1007/978-3-642-05101-2

Kim, J., Kazai, G., and Zitouni, I. (2013). "Relevance dimensions in preference-based IR evaluation," in *Proceedings of the 36th Annual ACM Conference of Research and Development in Information Retrieval (SIGIR' 13)* (New York, NY: ACM Press), 913–916.

Lin, C., and He, Y. (2009). "Joint sentiment/topic model for sentiment analysis," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (New York, NY: ACM Press), 375–384.

Mizzaro, S. (1997). Relevance: the whole history. *J. Am. Soc. Inform. Sci.* 48, 810–832.

Pothos, E., and Busemeyer, J. (2009). A quantum probability explanation for violations of "rational" decision theory. *Proc. R. Soc. B* 276, 2171–2178. doi: 10.1098/rspb.2009.0121

Schamber, L., Eisenberg, M., and Nilan, M. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Inform. Process. Manag.* 26, 755–775. doi: 10.1016/0306-4573(90)90050-C

Trueblood, J., and Busemeyer, J. (2011). A quantum probability explanation for order effects on inference. *Cogn. Sci.* 35, 1518–1552. doi: 10.1111/j.1551-6709.2011.01197.x

Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunctive fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi: 10.1037/0033-295X.90.4.293

Wang, J., and Busemeyer, J. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Top. Cogn. Sci.* 5, 689–710. doi: 10.1111/tops.12040

White, R. (2013). "Beliefs and biases in web search," in *Proceedings of the 36th Annual ACM Conference of Research and Development in Information Retrieval (SIGIR' 13)* (New York, NY), 3–12.

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 December 2013; accepted: 30 May 2014; published online: 02 July 2014.
Citation: Bruza P and Chang V (2014) Perceptions of document relevance. Front. Psychol. 5:612. doi: 10.3389/fpsyg.2014.00612
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Bruza and Chang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 24 June 2014
doi: 10.3389/fpsyg.2014.00554

Quantum theory and human perception of the macro-world

Diederik Aerts*

Center Leo Apostel, and Departments of Mathematics and Psychology, Brussels Free University, Brussels, Belgium

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
Maria Luisa Dalla Chiara, University of Florence Dipartimento di Lettere e Filosofia, Italy
Claudio Garola, University of Salento, Italy

****Correspondence:***
Diederik Aerts, Center Leo Apostel, and Departments of Mathematics and Psychology, Brussels Free University, Pleinlaan 2, 1050 Brussels, Belgium
e-mail: diraerts@vub.ac.be

We investigate the question of 'why customary macroscopic entities appear to us humans as they do, i.e., as bounded entities occupying space and persisting through time', starting from our knowledge of quantum theory, how it affects the behavior of such customary macroscopic entities, and how it influences our perception of them. For this purpose, we approach the question from three perspectives. Firstly, we look at the situation from the standard quantum angle, more specifically the de Broglie wavelength analysis of the behavior of macroscopic entities, indicate how a problem with spin and identity arises, and illustrate how both play a fundamental role in well-established experimental quantum-macroscopical phenomena, such as Bose-Einstein condensates. Secondly, we analyze how the question is influenced by our result in axiomatic quantum theory, which proves that standard quantum theory is structurally incapable of describing separated entities. Thirdly, we put forward our new 'conceptual quantum interpretation', including a highly detailed reformulation of the question to confront the new insights and views that arise with the foregoing analysis. At the end of the final section, a nuanced answer is given that can be summarized as follows. The specific and very classical perception of human seeing—light as a geometric theory—and human touching—only ruled by Pauli's exclusion principle—plays a role in our perception of macroscopic entities as ontologically stable entities in space. To ascertain quantum behavior in such macroscopic entities, we will need measuring apparatuses capable of its detection. Future experimental research will have to show if sharp quantum effects—as they occur in smaller entities—appear to be ontological aspects of customary macroscopic entities. It remains a possibility that standard quantum theory is an incomplete theory, and hence incapable of coping ultimately with separated entities, meaning that a more general theory will be needed.

Keywords: human perception, quantum theory, macroscopic entity, separated entities, concepts, objects, quantum effects, quantum axiomatics

1. INTRODUCTION

Why customary macroscopic entities appear to us humans as they do, i.e., as bounded entities occupying space and persisting through time, is a fundamentally puzzling question. It is puzzling because such macroscopic entities are built from microscopic physical entities, which are well described by quantum theory, and, following the quantum description, we know that these microscopical physical entities are 'not at all bounded entities occupying space and persisting in time' (Planck, 1901; Einstein, 1905; de Broglie, 1923; Heisenberg, 1925, 1927; Schrödinger, 1926a,b; Bohr, 1928; von Neumann, 1932; Einstein et al., 1935; Bohm, 1952; Bell, 1964; Jauch, 1968; Piron, 1976). The question of how 'constitutions of microscopic entities that are fundamentally not localized in space-time' build up to the 'customary macroscopic entities' in a way that is compatible with how we perceive their behavior, is not only a theoretical conundrum. Indeed, many experiments have been performed showing that whenever entities on larger scales are pushed in delicate and specific ways to show quantum effects, such as entanglement, non-locality and interference, they reveal 'aspects of' this quantum behavior (Rauch, 1975, 2000; Aspect et al., 1981, 1982; Tittel et al., 1998; Weish et al., 1998; Arndt et al., 1999; Aspelmeyer et al., 2003; Salart et al., 2008; Gerlich et al., 2011; Herbst et al., 2012; Bruno et al., 2013). Such experiments have meanwhile reached the astonishing scales of distances of 143 kilometers in the case of entanglement, and sizes of large macro- and bio-molecules in the case of interference (Gerlich et al., 2011; Herbst et al., 2012).

On the other hand, there is now a level of great detail and consistency in the way the theoretical framework of quantum theory accounts for the core of the weird behavior of micro-entities, and the penetration of aspects of this weird behavior into our everyday macroscopic world. This level of detail reveals the type of consistency which entails that approximate explanatory visions cannot be considered to be serious explanations of the matter. By 'approximate explanatory visions' we mean more concretely the original explanatory vision involving particles and waves (de Broglie, 1923, 1928). In one of its developments, it puts particles and waves in a dual mode with respect to each other—the so-called Copenhagen interpretation of quantum theory (Bohr, 1928), where the question of whether an entity entails particle or wave behavior depends on the measurement being performed upon it—while in another of its developments, it attempts to consider both of them as existing at once—the so-called de Broglie-Bohm interpretation of quantum theory (de Broglie,

1923, 1928; Bohm, 1952), where both particles and waves together and aligned constitute the quantum entity in all of its behavior. Although these wave-particle visions have succeeded in putting forward explanations for some of the quantum behavior, they fail to do so for several other aspects of quantum phenomenology that have now been well established, also experimentally. In what follows, we will first explain how they succeed in accounting for the weird quantum behavior to a considerable extent, and then discuss the aspects of this behavior where they fail in providing an explanation.

2. WAVES, PARTICLES, SPIN, AND IDENTITY

The main explanatory aspect of the wave-particle vision with respect to the question of 'why macroscopic entities behave classically, i.e., as bounded entities occupying space and persisting through time', is already contained in the original formula put forward by de Broglie (1923)

$$\lambda = \frac{h}{p} \qquad h = 6.62 \cdot 10^{-34} J \cdot s \tag{1}$$

where λ is the de Broglie wave length of an entity with momentum p, and h is Planck's constant. The idea is that quantum behavior within a collection of entities, e.g., a gas of particles, only appears when the de Broglie wavelengths of various of these entities can overlap, i.e., they are bigger than the typical distance between the entities. Indeed, only in this case can quantum coherence as an effect inducing the other aspects of quantum behavior manifest itself sufficiently. To give an idea, the de Broglie wavelength of a relativistically moving electron is of the order of magnitude of one nanometer $= 10^{-9}$ meters, which is the same order of magnitude as the size of an atom. This means that the de Broglie waves of electrons inside an atom overlap heavily. However, a car driving down the highway has a de Broglie wavelength of the order of magnitude of 10^{-38} meters, which is extremely small. This means that de Broglie waves of two cars on a highway will never overlap. Why use this criterion of 'overlapping'? The mechanism imagined in the wave-particle vision is the following. Consider particles in a gas that are (almost) at rest, and hence have de Broglie waves with large wavelengths that overlap widely. The waves can then start to vibrate in phase, join together to (more or less) form a single wave. The effect of the behavior of different particles melting together to the behavior of one wave pattern, hence of one particle, is called quantum coherence. Of course, for a real gas, such a situation can only occur at very low temperatures, since heat adds energy and hence momentum to each of the particles, so that their de Broglie wavelengths will become smaller and smaller, to the extent that the waves no longer overlap. It should be noted that the pure effect of becoming smaller is not what makes quantum behavior disappear. It is the non-globally structured way in which the wavelength decreases that destroys the quantum coherence. Indeed, heat is intrinsically a non-structured random way of adding energy, which is why 'it is a process profoundly disturbing the quantum coherence'. The different entities, i.e., particles of the gas, that at low temperatures were united into one macroscopically sized de Broglie quantum wave, start to get disconnected, their de Broglie waves being pushed out of phase as a consequence of the collisions with random packets of heat energy. This means that with rising temperature the gas starts slowly to become a collection of separated particles, behaving classically with respect to each other. Let us remark that a collection of cars on the highway, within this explanatory scheme, is still a collection of quantum entities, but with de Broglie wavelengths that are so small, and heat disturbances so huge, that the different de Broglie waves would never be able to cohere, and hence no quantum effects can be observed.

The wave-particle explanation has an intuitive appeal for a very specific reason, because we can all experience the very similar effects of real wave-like phenomena in our everyday world. For instance, imagine you are in a playground with your children, and you are pushing a swing with one of them on it. We all know from experience how this only works when the frequency of our pushes 'resonates' with the eigenfrequency of the swing-with-child. Now suppose there are two people pushing the swing, one at either side, this will only work if the frequencies of the two pushing adults are coherent, sufficiently similar, i.e., overlap in the time-dimension. This is much more difficult to accomplish in the case of high frequencies—imagine a tiny miniature swing with a very high eigenfrequency being pushed by two persons using their fingertips. The reason is very similar, for with higher frequencies, the effects of the random disturbances that we experience with respect to our attempts to control our movements in finding the coherence with the eigenfrequency of the swing become more prominent. This means that it will become more and more difficult to realize the required coherence as the frequency increases.

The frequencies considered in the above swing example are the analog for time of what wavelengths are for space. But we can easily find an example in space where also our intuition readily lets us understand the wave-particle explanation presented above. Imagine a bath tub filled with water, and two persons on either side moving their hands rhythmically to make waves in the water. If the wavelengths of the water waves are of the order of magnitude of the size of the bath tub, the waves made by one person will interfere with the waves made by the other person. This is actually what will normally happen when water waves are made by hands moving up and down in the water on both sides. However, waves with smaller wavelengths will not have the same effect. Let us consider sound waves in the air, for example. Interference of sound waves is a well-known phenomenon, giving rise to volumes of the sound going up and down, the so-called 'beating sounds'. Two tuning forks whose tones slightly differ and hence produce sound waves with different wavelengths, will produce such a beating effect when sounding together, as a consequence of the interfering sound wave. However, tuning forks are built on purpose from the right material and in the right form to enable them to produce the pure type of eigenfrequency and create very pure, almost plane waves, i.e., with wavelengths that remain the same over large distances. For sound produced by entities not designed for such pure results, interference is a much less obvious phenomenon.

The above wave-particle explanation of 'why macroscopic entities, such as cars, do not show quantum effects, although within the wave-particle vision they too would be quantum entities' may not be incorrect in principle—based as it is on the idea

of coherence at the origin of quantum effects, and de-coherence as a consequence of random disturbances—but it is incomplete. For example, it does not provide a satisfactory explanation for the quantum phenomena linked to 'spin', which is a fundamental property of all microscopic entities. The name 'spin' was given to this quantum property, because in the early days physicists thought it was an expression of angular momentum on the micro-scale. We now know that the property 'spin' is not really the angular momentum of a micro-particle, but rather a genuinely new type of quantum property without any obvious classical equivalent, although it structurally does indeed show significant similarities with angular momentum. Nor does the wave-particle vision provide a satisfactory explanation for the quantum phenomena that are linked with situations of identical entities, and they are numerous. As we will see in the following of this article, most of the spectacular realizations of quantum phenomena on the macroscopic level—superfluidity and supercurrency (London, 1938; Josephson, 1962; Gravroglu and Goudaroulis, 1988)—are related to spin and to identity of entities, and in most cases even to both.

Every micro particle has a property called 'spin', which can essentially be half-integer or integer in value, but is always quantized, i.e., it never takes continuous values. Being always quantized can be understood within the wave-particle vision, spin being analogous with angular momentum. Indeed, consider a micro-particle rotating around itself and also being a wave. For it to be coherent with itself, the wave-pattern will have to repeat itself after a rotation, and this requirement leads to quantization, different possible modes being solutions. This is interesting to note, because a new aspect of the wave-particle vision appears, namely coherence with itself. However, things become more difficult to explain within the wave-particle vision if we point out the so-called 'spin-statistics' relation, formulated at the end of the 1930ies, first by Markus Fierz, and subsequently by Wolfgang Pauli (Fierz, 1939; Pauli, 1940). The relation was eventually proven in the context of relativistic field theory, but the proof remains obscure and still has not provided a satisfactory explanation (Pauli, 1950; Streater and Wightman, 2000; Jabs, 2010).

The relation between 'spin' and 'statistics' in the form of the 'spin-statistics' theorem can be stated as follows: "For a situation of identical integer-spin particles, the wave function describing the state of such particles remains unchanged when the particles are permuted. We call these types of wave functions 'symmetric' and the particles described by it, 'bosons'. For the situation of identical half-integer spin particles, the wave function describing the state of such particles changes sign when the particles are permuted. We call these type of wave functions 'asymmetric' and the particles described by it, 'fermions'." Hence, the spin-statistics theorem states that integer spin particles are bosons, while half-integer spin particles are fermions. What is interesting, is that the spin-statistics theorem implies that half-integer spin particles, hence fermions, are subject to the Pauli exclusion principle—only one fermion can occupy a specific quantum state at a specific time—, this follows directly from the asymmetry of the wave function. Indeed, suppose that we consider two fermions in the same state; in this case, a permutation does not change anything, since they are in the same state, but, because of its asymmetry changes the sign of the wave function. This is only possible for a wave function equal to zero for the fermions in the same state. For integer-spin particles, i.e., bosons, with a symmetric wave function, there is no restriction in occupying the same state.

This difference between fermions and bosons has a dramatic influence on the way both types of particles behave statistically, in other words, when for example they appear in great quantities, in the form of solids, liquids or gasses. Two very different types of statistical behavior have been given the names of 'Fermi-Dirac-statistics' and 'Bose-Einstein-statistics'. The difference in behavior is very fundamental and gives rise to very different types of compound structures. It can, for example, be proven that bosons cannot give rise to stable forms of matter, and as a consequence all matter is formed by fermions, i.e., fermions are the basic building blocks of matter (Dyson and Lenard, 1967; Lenard and Dyson, 1968; Lieb, 1976, 1979; Muthaporn and Manoukian, 2004). Electrons are fermions, which is why only two of them can be in the same lowest-energy state, one with its spin in one direction, and the other one with its spin in the opposite direction. A third electron necessarily needs to be in a higher-level energy state, and so forth for subsequent electrons. This means that the whole range of atoms in the periodic table, giving rise to all the variety in chemistry, mainly finds its origin in the special way in which spin 1/2 quantum particles behave as identical entities, namely fermions.

Bosons are the particles that carry the interaction fields of the forces. We can understand by intuition that fermions, more specifically, electrons, neutrons, and protons, can form building blocks for matter. Indeed, matter takes up space, and this can be imagined to come about because the basic blocks cannot be in the same state. Hence, the spins of these building blocks of matter—electrons, neutrons and protons—entail a type of pressure called 'degenerative pressure' that prevents them from merging into the same state. This 'degenerative pressure' pushes combinations of fermions to become bigger and bigger, where we use the word 'bigger' in its specific meaning of 'taking place within a region of more space, whenever they are forced to take place'. Photons are bosons and have spin equal to 1, hence they are not confronted with 'degenerative pressure', which means that many of them can be in states that are very similar, even equal. So, different photons can in principle be in one and the same state. The realization of a 'laser' is essentially based on this possibility. The word 'laser' stands for 'light amplification by stimulated emission of radiation', and it was Albert Einstein who laid the basis for the quantum-mechanical mechanism of absorption, spontaneous emission and stimulated emission that guides it (Einstein, 1917). What essentially happens in a laser is that an enormous amount of photons is produced, but in such a way that they are in states that are coherent in space as well as in time. In the limit, they are actually all in one and the same state, including the 'wave-aspects' of the state, i.e., the 'phases'. Concretely, photons of a laser beam therefore do not only have the same wavelength and frequency, but are also 'in phase', which means that they have the same phase, and hence are in the same state, which is only possible because photons are bosons.

We have so far considered the fundamental difference in statistical behavior of fermions and bosons by looking at two examples, electrons as fermions, and their statistical behavior within atoms, giving rise to all of the properties of chemistry, and photons as bosons, and their statistical behavior within laser light, giving rise to our first example of a macroscopic quantum system. However, both electrons and photons, as far as we know today, are elementary particles, i.e., they have no known constituents, and to date all attempts to find any such subentities have failed. That is why they are considered to be really elementary. However, the fermion and boson nature of quantum entities is also apparent in composite quantum particles, such as atoms and molecules. And in this respect an additional amazing aspect of quantum physics is revealed, which is that also for such composite particles the relation between spin and statistics remains valid, and 'spin adds up and does so following the mathematics of a vector in a small and finite dimensional vector space, called Hilbert space'.

Let us illustrate the above with two examples. There are two isotopes of the atom Helium, namely Helium-3, with a nucleus consisting of two protons and one neutron, and Helium-4, with a nucleus consisting of two protons and two neutrons. Protons and neutrons are fermions, both with spin equal to 1/2. What are the spins of the Helium-3 and Helium-4 isotopes? Well, spins of compound quantum entities are the vector sums of the spins of their constituents, so that, in case they are aligned, they can be summed or subtracted numerically. This means that Helium-3, consisting of three particles with spin 1/2, will have spin 1/2 or 3/2, but in any case half integer. While Helium-4, consisting of 4 particles with spin 1/2, will have spin 0, 1 or 2, but in any case integer. The spin-statistics relation is also valid for compound quantum entities, which means that Helium-3 is a fermion, while Helium-4 is a boson. And both indeed behave statistically in this way, with Helium-3 being faithful to the Pauli-exclusion principle—no two Helium-3 entities are encountered in the same state—, and Helium-4 allowing to be pushed all into the same state. This is not just theory but can also be realized experimentally. The first Bose-Einstein condensate, which is the name given to the phenomenon where a whole gas of atoms is in such a state that it is one entity, was realized in 1995 by Eric Cornell and Carl Wieman. They made use of an isotope of the atom rubidium, and needed to slow down the motion of the atoms in the gas by cooling it to 1.7×10^{-7} kelvin for the de Broglie waves of different atoms to start overlapping and merging into one quantum wave for the whole gas (Anderson et al., 1995). They received the Nobel Prize in Physics in 2001, together with Wolfgang Ketterle at MIT, for this achievement, which climaxed a 15-year search by physicists worldwide for a realization of such a Bose-Einstein condensate.

Many years before the realization of this genuine macroscopic quantum state of matter called the Bose-Einstein condensate, a phenomenon called superfluidity had been experimentally identified for liquids composed of bosonic atoms, and more specifically for a liquid of Helium-4. Indeed, when Helium-4 is cooled down to below about 2.2 kelvin, it starts behaving weirdly. It passes through narrow tubes seemingly without any friction, and climbs up walls overflowing its container. Although early observations of odd behavior had been recorded, it was only a long time after Heike Kamerlingh Onnes first liquefied helium in 1908 that its superfluidity was fully discovered, in 1938, by Pyotr Kapitsa in Moscow, and independently by John F. Allen and Donald Misener at the University of Toronto (Allen and Misener, 1938; Kapitza, 1938). However, it was to take quite some years before Fritz London put forward the hypothesis—which at the time was still considered highly speculative—that superfluidity was a phenomenon due to Bose-Einstein condensation. Laszlo Tisza worked out a two-fluid model for liquid helium elaborating on London's hypothesis (London, 1938; Tisza, 1938).

A much more complicated phenomenon, superconductivity, was observed as a consequence of the cooling techniques developed by Heike Kamerlingh Onnes, the same as those that allowed him to produce liquid Helium. When he studied the resistance of solid mercury at such low temperatures, he found this resistance to be almost inexistent. In later years, this extreme form of conductivity was to be identified in many other materials at very low temperatures, but remained unexplained, despite major efforts to understand the phenomenon. The resistance being zero is demonstrated by the fact that currents can be sustained in superconducting rings for many years with no measurable reduction, while an induced current in an ordinary metal ring would decay rapidly because of the dissipation through ordinary resistance. An important step toward a deeper understanding was taken when in 1933 Walther Meissner and Robert Ochsenfeld discovered that superconductors expel magnetic fields in an extreme way, a phenomenon which has come to be called the Meissner effect (Meissner and Ochsenfeld, 1933). Several years later, Fritz and Heinz London showed that the Meissner effect is a consequence of the minimization of the electromagnetic free energy carried by a superconducting current, and they developed the first phenomenological theory for superconductivity (London and London, 1935). A more powerful, but still phenomenological theory was developed in 1950 by Ginzburg and Landau (1950). It was not until 1957, however, that a microscopical theory emerged, when John Bardeen, Leon Neil Cooper and John Robert Schrieffer explained superconductivity due to Bose-Einstein condensation, as a consequence of an effect of superfluidity of electron pairs bound in a very specific way, namely such that the pair is a boson (Bardeen et al., 1957). These pairs, now commonly referred to as 'cooper pairs', interact quantum mechanically by means of phonons (Cooper, 1956). What is mind-boggling, is that the electrons in a cooper pair are usually far apart from each other, at distances greater than the average distance between electrons, and remain bound to behave together as a boson by means of an interaction with the crystal lattice of the conductor through phonons, leading to the effect of superconductivity. The proposed mechanism, yielding an explanation for the superconductivity in cold conductors, rests on firm grounds, theoretically as well as experimentally, since for instance the superfluidity of Helium-3, reached only at temperatures much lower than that at which the superfluidity of Helium-4 appears, has now also been explained by cooper-pairing of the atoms of Helium-3 themselves into bosons as pairs, although each of them is a fermion.

In short, the purpose of the above digression was to explain some of the details of the two macroscopic quantum phenomena called superfluidity and superconductivity, both due to their

being a realization of a Bose-Einstein condensate, i.e., to a specific type of entity behaving as a boson in its lowest-energy state, and hence fusing all previously separated entities into one whole. We have now come to the main point to be made for the purpose of the present article. This macroscopic quantum behavior is crucially dependent on the spin of the considered entity. However, the spin is not a property that can be fit well into the wave-particle explanation, it is neither a wave nor a particle, and it is described by a vector in a finite dimensional complex Hilbert space. In the next section, we will analyze some of the founding steps of quantum physics itself to see why it is important to pay attention to the fact that the spin is such a special property.

3. QUANTUM AXIOMATICS AND SEPARATION

Quantum theory arrived in two quite distinct ways. The first was by means of the matrix mechanics of Werner Heisenberg, elaborating further the approach by Max Planck and Niels Bohr with respect to the notion of quantization, and the modeling of the atom (Planck, 1901; Heisenberg, 1925; Bohr, 1928). The second way was as a consequence of the wave mechanics of Erwin Schrödinger, elaborating on the work of Albert Einstein and Louis de Broglie with respect to particle-waves and matter waves (Einstein, 1905; de Broglie, 1923, 1928; Schrödinger, 1926a). Schrödinger and later, more systematically, John von Neumann, showed that matrix mechanics and wave mechanics are equivalent as physical theories (Schrödinger, 1926b; von Neumann, 1932). From a mathematical perspective, it can be proven that the version of quantum mechanics known as matrix mechanics is equivalent to von Neumann's linear algebra and complex Hilbert space quantum mechanics, when the Hilbert space is taken to be l^2, the set of all sequences $(z_1, z_2, \ldots, z_j, \ldots)$ of complex numbers z_j such that the series of the square of their absolute values converges, i.e., $\lim_{n\mapsto\infty} \sum_{j=1}^{n} |z_j|^2 < \infty$. Matrices are then linear functions on such sequences, and indeed, Heisenberg needed 'infinite matrices' for his matrix mechanics. On the other hand, the wave-mechanics version of quantum theory is equivalent to von Neumann's linear algebra and complex Hilbert space, when the Hilbert space is taken to be $L^2(\mathbb{R}^{3n})$, the set of all square integrable complex functions of $3n$ real variables for n quantum particles. These functions are the so-called Schrödinger wave functions.

The fact that quantum theory appeared in two quite different versions which showed to be equivalent contains an important message. It means that we have to be very careful when deriving possible physical images from one or both of these versions, because they are mainly representations of a general theory of linear algebra and complex Hilbert space, which from now on we will call standard quantum theory. The danger of putting forward a physical image that is based only on the specific form of a representation and that hence might not have a profound significance is primarily present for wave mechanics, being developed from the start with such a specific physical image in mind, namely that of a 'wave'. And indeed, throughout the years the notion of wave has been dominant in 'imagining what a quantum entity is'. Matrix mechanics—or more in general, the linear algebra aspects of standard quantum theory, since a matrix or a linear function does not produce a straightforward image—continued to be mainly considered as a mathematical apparatus. If we know the profound mathematical equivalence of both theories, and also the importance of spin, which has no associated 'wave image', since its states are vectors in a finite dimensional Hilbert space, there are strong reasons to doubt the validity of the 'wave image' in our attempt to grasp the physical aspects of a quantum entity. There is indeed the—not to be neglected—chance that the prominence of the wave image in quantum interpretations is only more or less coincidental, because it appears in one specific realization of Hilbert space, namely $L^2(\mathbb{R}^{3n})$. The more so if we remember, as explained in Section 2, that spin is at the origin of manifestly different types of quantum behavior, Bose-Einstein statistics or Fermi-Dirac statistics, even on the level where quantum properties show macroscopically, such as in lasers and Bose-Einstein condensates. So perhaps the only logical conclusion we should allow to be drawn until further relevant information is available is that quantum entities are 'neither particles nor waves', rather than to imagine them as particle-waves. Moreover, in the decades following the early development of quantum theory, various axiomatic and operationally founded quantum formalisms were worked out, all of them more general than the formalism of standard quantum theory of linear algebra and complex Hilbert space (Mackey, 1963; Jauch, 1968; Piron, 1976, 1989, 1990; Aerts, 1982a, 1983a,b; Ludwig, 1983; Foulis, 1999). This means that even more mathematically inspired notions, such as 'the superposition principle', which find its origin in the linearity of standard quantum theory in Hilbert space, should be looked upon with care in case one wants to use them as a foundation for the interpretation of quantum theory, because indeed, these operational axiomatic quantum theories are not a priori linear theories.

In this respect, we specifically want to put forward and analyze a result we obtained ourselves quite some time ago when investigating the situation of 'separated physical entities' within such a generalized axiomatic and operational quantum theory, because of its relevance for the main question considered in the present article. The generalized axiomatic operational quantum formalism in which we performed this investigation on separated physical entities is the one currently referred to as the Geneva-Brussels School on quantum theory (Piron, 1964, 1976, 1989, 1990; Aerts, 1982a, 1983a,b, 1986, 1999a,b, 2009c; Cattaneo and Nistico, 1991, 1993; Aerts et al., 1999; Aerts and Van Steirteghem, 2000; Coeck et al., 2000; Smets, 2003; Engesser et al., 2007; Sassoli de Bianchi, 2010, 2011, 2013).

The Geneva-Brussels School quantum theory is an axiomatic operational generalization of standard quantum theory. It is operational because it attempts to introduce mathematical notions, and also as many axioms as possible, in such a way that they have a clear physical meaning. For the purpose of this article, it is by no means necessary to explain this theory, because I will use it only to formulate the result about separated entities relevant to the subject we are concerned with. Let me first express this result by means of the following simple statement.

Statement A: *Standard quantum theory is incomplete in the sense that it cannot describe the compound entity consisting of two separated subentities.*

To explain the result expressed in statement A in a way that its meaning becomes clear, I will introduce some notions—more specifically the names of axioms—from the Geneva-Brussels School quantum theory, and also sketch some of the history of how I arrived at proving this result. There is no need at all to know what these notions/names mean, because the result about separated entities expressed in statement A can be formulated completely independently of their content. After analyzing its meaning, I will give an example to illustrate this result. The Geneva-Brussels School quantum theory reduces to standard quantum theory if five axioms are satisfied, to wit (1) 'completeness', (2) 'orthocomplementation', (3) 'atomicity', (4) 'weak modularity', and (5) 'the covering law'. It was in fact Constantin Piron who in 1964 proved that these five axioms led to standard quantum theory by means of a now famous representation theorem in axiomatic quantum theory (Piron, 1964, 1976). The motivation to investigate the situation of 'separated entities' goes back to a situation which Ingrid Daubechies and myself analyzed at the end of the 1970ies, namely the description of compound entities within the operational axiomatic Geneva-Brussels School quantum theory. In those days, the well-known tensor product procedure used in standard quantum theory for the description of compound quantum entities had not yet been investigated at the operational axiomatic level. At the time, we had in mind to search for criteria that would give rise to the tensor product procedure in case of standard quantum theory interpreted within the operational axiomatics of the Geneva-Brussels School quantum theory, and the first investigations seemed promising in this respect (Aerts and Daubechies, 1978). Because of the powerful operational aspects of the Geneva-Brussels quantum theory, parallel to the mathematical aspects explored in Aerts and Daubechies (1978), I decided to construct explicitly the model of the most simple of all operational situations, namely the situation of 'two separated physical entities'. A very surprising and also completely unexpected result followed because, when constructing literally by hand the model of the compound of two separated entities, I could prove that this model would never satisfy axioms 4 and 5, called 'weak modularity' and 'the covering law', whenever the two subentities were genuine quantum entities, e.g., described well by standard quantum theory. When I found this result I was working in Geneva on my PhD under guidance of Constantin Piron, and I remember that the whole group in Geneva was in a state of disbelief about it, because it implied, if correct, that a structural shortcoming of standard quantum theory had been identified on its core axiomatic nature, 'its incapacity to model separated entities'. It became the cornerstone of my PhD, which I defended in 1981 (Aerts, 1982a, 1983a,b). Certainly the failure of axiom 5, the covering law, was shocking, since it is an axiom equivalent to the linearity, i.e., the vector space structure, of the set of states of the considered entity. So, I had proven that the set of states of the compound entity of two separated quantum entities could not be linear, hence could not be a vector space. Obviously, if one knows how much standard quantum theory is founded on the linearity of the considered mathematical structure, e.g., the Hilbert space of the set of states, when this linearity is no longer satisfied, all of standard quantum theory breaks down. For example, the superposition principle will no longer be a principle valid for all states.

In the years that followed, I understood that my result was in concordance with findings related to the violation of Bell's inequalities, which was then becoming a focus of attention in the foundation of quantum physics research. My analysis was a constructive one, however, in the sense that I explicitly constructed the model for two separated physical entities, and identified the aspects of that model that made it impossible to be realized within a standard quantum theory. My result did not rely on an argumentation 'ex absurdum', which was, for example, the argumentation contained in the original Einstein Poldolsky Rosen paper (Einstein et al., 1935). As a consequence, I was able to analyze the EPR paradox situation as one that indeed proves standard quantum theory to be not complete, but in the sense that it cannot describe separated quantum entities. This means that the EPR proof contained in Einstein et al. (1935) is correct, but it is a proof 'ex absurdum', consisting in finding a logical contradiction, i.e., 'if quantum theory is complete, then it is not complete'. From this of course follows that 'it is not complete', but this consequence is only the result of 'the hypothesis of completeness leading to a contradiction'. Since my proof of the incompleteness was constructive, I could indicate the origin of this incompleteness, and this was 'not', like EPR inferred from their finding of a contradiction, the necessity of the existence of 'hidden variables', but its failure to model separation. The constructive nature of my analysis of the situation even allowed me to operationally identify the missing elements of reality, and thus indicate the incompleteness operationally and directly (Aerts, 1984). I remember meeting Alain Aspect—the physicist performing the crucial experiments in 1982 about the violation of Bell's inequality with entangled photons (Aspect et al., 1981, 1982)—on several occasions and asking him: "But would you still violate Bell's inequality and identify entanglement in case you made an effort to separate the quantum entities, rather than make every effort to not separate them, as you are doing now?". "Of course not", he answered, "but why would one want to try this?" Naturally, because of my constructive approach to investigating the description of separated physical entities, this question had become relevant and even crucial to me. I had approached the situation from the beginning from the opposite direction than most, if not all, physicists involved in this problem, which is the reason why the relevance of this question remained unnoticed by the others. My analysis showed that separated entities could not be described by standard quantum theory, even if no attempt was made to keep them entangled, while all experiments conducted with respect to EPR started from a different approach, their question being 'how far can we set detectors apart in space, such that entanglement is still registered, while we make every effort to keep this entanglement intact'. They were interested in testing 'how far the quantum effect of entanglement reaches, when it is attempted to keep it intact as much as possible', whereas I had become intrigued by the finding that 'in whatever state of separation two quantum entities are prepared, a standard quantum theoretic description of their compound entity is not possible, as such a description will always introduce entanglement'.

One of the reasons why it is difficult to identify within standard quantum theory itself the result I obtained within the Geneva-Brussels School quantum theory, and the ensuinging inability to model separated quantum entities for standard quantum theory, is that this inability does not appear obviously at the level of the set of states. Indeed, one may wonder, 'Why not just use product states in the tensor product Hilbert space model for the compound entity consisting of separated quantum entities to describe this compound entity consisting of separated quantum entities?'. Although, since the fifth axiom, the convering law, being equivalent to linearity, cannot be satisfied in the case of separated entities—this is what I proved in Aerts (1982a, 1983a,b) –, and hence the set of states 'cannot' be a vector space for the compound entity consisting of two separated quantum entities, there are enough states in the tensor product, namely exactly the set of product states, to cope with each situation of such compound of two separated quantum entities. However, this set of product states in the tensor product does not cope correctly with other fundamental aspects of the situation, even if all entangled states, being the linear superpositions of these product states, are left out of the description. This can be seen straightforwardly only on the level of the observables and/or of the dynamical evolutions. We will show this by looking at a concrete example of the compound entity of two separated entities by focusing on the question, 'What are the possible evolutions that can be described within such a tensor product standard quantum theory description'. Let us remark that 'evolutions', in the case of standard quantum theory, are described by unitary transformations of the Hilbert space—the so-called Schrödinger equation stands for such unitary evolutions. More concretely, if H is the Hamiltonian of the entity, then the Schrödinger equation expressed in unitary evolution form is

$$\psi(t) = e^{i\frac{h}{2\pi}Ht}\psi(0) \tag{2}$$

where $\psi(t)$ is the wave function at time t, as a vector of the Hilbert space of states. There are enough states in the tensor product, namely the set of product states, but there are not enough evolutions, that is where standard quantum theory explicitly fails to describe the compound of two separated quantum entities. To illustrate this, I will first refer to a theorem that can easily be proven within standard quantum theory, which is the following. If one considers the tensor product description of the compound entity consisting of two quantum entities, then a unitary transformation $U(1,\ 2)$ of the compound entity that conserves product states—hence maps product states onto product states – is always of the form $U(1) \otimes U(2)$, the tensor product of a unitary transformation of the first entity with a unitary transformation of the second entity. So, if one attempts to describe separated entities within the tensor product, only tensor products of evolutions of both entities apart keep them separated. Whenever an evolution is not such a tensor product, product states will go to entangled states as a consequence of such an evolution. Next to this theorem, there is a point to be clarified, namely that 'separated' does not mean 'without possible interaction'—entities in the classical world indeed remain separated if they only interact dynamically, because indeed, dynamical interaction does not destroy the product states. In classical physics, most interacting entities are separated but dynamically interacting. This is the meaning of 'separated' that I used in my theorem (Aerts, 1982a, 1983a,b). This means that statement A can be refined as, 'The compound entity of separated quantum entities that interact dynamically cannot be modeled in standard quantum theory'. If we now consider any type of dynamical interaction between two quantum entities, we can see that this interaction will be expressed in a Hamiltonian $H(1,\ 2)$ of the compound entity, which is not a simple sum of two Hamiltonians $H(1)$ and $H(2)$ of each of the subentities apart, whenever the interaction is nontrivial. This means that the unitary transformation generated by this Hamiltonian, i.e., $e^{i\frac{h}{2\pi}H(1,\ 2)t}$, not being a product of two unitary transformations, each of one of the subentities—this would only be the case if $H(1,\ 2)$ is a sum of two Hamiltononians, of which each is the Hamiltonian of one of the subentities, meaning that there is no dynamical interaction between the subentities—will not work within the set of product states. Or, more concretely, it will change any product state right away into an entangled state. Let us make it even more concrete. Suppose two neutrons are placed in faraway spots in completely empty space, which means that only gravitational interaction exists between them. This gravitational interaction expressed in an interaction Hamiltonian will give rise to an evolution of the compound entity of these two neutrons that leads them right away into entangled states. Both neutrons rotating around a common center of mass, as is the case with macroscopic material objects that dynamically interact only through gravity, is hence not possible within a standard quantum theoretic description in Hilbert space. Of course, one possibility is that indeed no such two neutrons in a gravitational Kepler movement exist in our reality, and that the existence of such a Kepler movement of two macroscopic material entities is a specificity of their being macroscopic. In this respect, I want to bring up a subtlety of the theorem that I proved for my PhD thesis, namely that 'only in the case both subentities are quantum entities, technically meaning that at least one superposition state exists for each of the subentities, can the compound of these entities not be a compound of separated entities interacting dynamically'. It is sufficient that one of the two entities is classical—which technically means that no superposition states exist for any of the states—for the investigation I made to allow the description of separated subentities.

This means that from a logical point of view, my finding leaves open the following two possibilities. Statement A: *Standard quantum theory is incomplete, in the sense that it cannot describe the compound entity consisting of two separated subentities*. Statement B: *Such a compound entity does not exist or, in other words, whenever two quantum entities exist, their compound entity is not separated*.

Alain Aspect's experiments (Aspect et al., 1981, 1982) conducted around the same time when I defended my PhD, and also all later experiments aiming to find quantum effect on ever wider macroscopic scales, of which we gave an overview and analysis in this article (Rauch, 1975, 2000; Aspect et al., 1981, 1982; Tittel et al., 1998; Weish et al., 1998; Arndt et al., 1999; Aspelmeyer et al., 2003; Salart et al., 2008; Gerlich et al., 2011; Bruno et al., 2013), suggest that statement B is the correct conclusion to be

drawn. On the other hand, we live surrounded by macroscopic entities such as tables, chairs, cars, etc... that do not show quantum effects of any kind, which would indicate that statement A is to be considered correct. In the next section, we will argue that the situation is more complicated than that, as well as analyze how the experimental attitude of attempting to find quantum effects with all means possible—the root of my question to Alain Aspect in 1981—and adding to this our insights about the very nature of quantum effects themselves, is at the source of a subtle confusion that is not at all understood. This analysis will guide us in proposing our view on the main question of this article, i.e., 'why macroscopic entities present themselves to us the way they do'.

4. QUANTUM AND COGNITION, MEANING AND MATTER

Around the turn of the century, and more intensively so during the first decade of the 21st century, quantum theory, as a formalism, has been used with growing success to model situations in human cognition, so that nowadays 'quantum cognition' is emerging as a flourishing domain of research (Aerts and Aerts, 1995; Aerts et al., 2000, 2011, 2013a,b; Gabora and Aerts, 2002; Aerts and Gabora, 2005a,b; Bruza et al., 2007, 2008, 2009; Aerts, 2009b; Pothos and Busemeyer, 2009, 2013; Khrennikov, 2010; Aerts and Sozzo, 2011; Busemeyer et al., 2011, 2012; Song et al., 2011; Busemeyer and Bruza, 2012; Wang et al., 2013; Sozzo, 2014). Our research group in Brussels at the Center Leo Apostel has played an important role in the initiation (Aerts and Aerts, 1995; Aerts et al., 2000; Gabora and Aerts, 2002; Aerts and Gabora, 2005a,b) and further development (Aerts, 2009b; Aerts and Sozzo, 2011; Aerts et al., 2011, 2013a,b; Sozzo, 2014) of this new domain of research called 'quantum cognition'.

As for my own role in the development of quantum cognition, at least some of the seeds were sown toward the end of the 1970ies when I was reflecting about the result explained in Section 3, i.e., the inability of standard quantum theory to describe separated entities, and also confronting this result with the factual situation of being surrounded by separated entities in our everyday macroscopic world. My first insight was that non-separated entities could be easily realized as well in the macro world, for example, by connecting vessels of water, which even lead to a violation of Bell's inequalities (Aerts, 1982b). When I analyzed this violation of Bell's inequalities by the vessels of water in detail, it became clear that quantum probabilities and their non-Kolmogorovian structure could be explained from the presence of 'hidden measurements' or, in other words, the presence of 'fluctuations in—or a lack of knowledge about—the interaction between the measurement apparatus and the entity to be measured' (Aerts, 1986). Indeed, it was possible to show that such a lack of knowledge about the interaction between the measurement and the entity to be measured was part of the mechanism provoking the violation of Bell's inequalities in the vessels of water situation, and also in subsequent elaborations producing exactly the same numerical violation as the quantum one (Aerts, 1991). Once it was clearly understood how the quantum probability structure of the statistics of collected data arose—by the presence of a lack of knowledge about the interaction between the measurement apparatus and the entity to be measured—this led more or less naturally to the idea that similar situations—characterized by the presence of a similar lack of knowledge—would also appear in typical measurement situations in research in the human sciences, and more specifically in cognitive science. This insight was at the origin of the quantum probability model we worked out for the situation encountered in an opinion poll (Aerts and Aerts, 1995). In the same period I prepared an online lecture together with Liane Gabora, which stimulated me to work out a violation of Bell's inequalities in cognition, along the line of the vessels of water violation, but this time considering the 'change of opinion in a person's mind' as a quantum collapse event (Aerts et al., 2000). It was also during this ongoing collaboration that Liane Gabora suggested looking at the guppy effect, an experimentally tested anomaly in concept combination, and investigating whether quantum theory could deliver a modeling of this effect. It was the start of our in-depth investigation of concepts and their combinations, which yielded not only our SCOP theory (Gabora and Aerts, 2002; Aerts and Gabora, 2005a,b), but also the modeling of the very revealing data of James Hampton on the conjunction and the disjunction of concepts (Hampton, 1988a,b), as well as the development of our Fock space model (Aerts, 2009b), and further analysis and applications (Aerts and Sozzo, 2011; Aerts et al., 2013a,b; Sozzo, 2014).

In parallel with these investigations, I was hatching a new idea but it was still very premature and far too speculative to justify serious investigation. However, it kept popping up, and many times I found myself reflecting about it. The basis of the new idea was very simple, and can be expressed as follows: "If quantum theory is so successful in modeling aspects of cognition, and more specifically, also how the dynamics of concepts and their combinations work, could it not be that quantum particles are not objects, but entities having mainly a conceptual nature?" The additional thought naturally ensuing would be, "And would this perhaps also account for their highly strange behavior?" I have worked on this idea for several years now—albeit in parallel with a large number of other themes of research—and I must admit that my investigations have considerably strengthened my belief that many aspects of it must be true, which made me decide to develop it into a new and complete interpretation of quantum theory (Aerts, 2009a, 2010a,b, 2013). It also dawned upon me that in fact it is even the unique quantum interpretation which also contains an explanation for some of the major unexplained phenomena of quantum physics. Before I will discuss some of these, let me give a more detailed account of this new quantum interpretation.

When we say that the new interpretation assumes that quantum particles are 'conceptual entities' rather than objects, we do not mean this in a vague or merely philosophical way. The idea is that quantum particles are 'not' what they are often imagined to be, namely 'very complex objects flying between pieces of matter, by which they can be absorbed, and then live in bound states inside, and also radiated out again', but they are something much more deeply different still from a classical particle, namely 'conceptual entities mediating between such pieces of matter, these forming a type of memory structure for them'. A fundamental aspect of this new interpretation is therefore that we regard the dynamics on the level of the micro-world, as a dual

type of dynamics, with some of its entities mediating—these are the bosons—, and thus carrying meaning, between other entities that form memory structures—these are the pieces of matter, formed of fermions. The overall dynamics incorporates the co-evolution of these two types of entities, carried by a process of meaning exchange. Let us remark that, according to this new interpretation, 'quantum entities are conceptual with respect to their own memory structures, which are pieces of matter'. This means that they are 'not' concepts interacting directly with the human mind and that the human mind here does not serve as a memory structure for them. Such a direct dynamical interaction with the human mind, in which the human mind serves as a memory structure, exists only for human concepts themselves. The only direct way in which the conceptual nature of quantum entities comes about is through their dynamical interaction with pieces of matter, which act as their memory structures. In other words, the relation of human mind vs. human concepts, and the relation of pieces of matter vs. quantum entities can be said to be analogies taking place in different realms of reality. Of course, since human experiments with these quantum entities necessarily involve the use of measurement apparatuses, which are pieces of matter by definition, indirectly, through the interface of these measurement apparatuses, we, with our human minds, are confronted with the quantum entities behaving as conceptual entities in all our experiments with them. But our confrontation with their conceptual nature is only indirect, because of the unavoidable interfaces in the form of measurement apparatuses. Hence, the success of the quantum formalism as a mathematical formalism, in its description of the microworld, and its modeling of the cognitive dynamics of concepts, would be due to the fact that both realms, the micro-world where bosons mediate between pieces of matter formed of fermions, and the world of human communication, where language is used to mediate between minds, are realms of similar dynamics. For example, this new interpretation allows understanding and explaining the Heisenberg uncertainty principle as being due to the tradeoff between a concept being more abstract or more concrete (see Aerts, 2009a, Section 4.1). Let us be somewhat more specific. In Aerts and Gabora (2005a,b), we introduced the notion of 'state of a human concept', at that moment mainly to apply the mathematical quantum-like formalism that we developed to model human concepts and how they combine. Suppose we consider the human concept *Fruits* and take one of various experimentally measurable observables introduced by psychologists studying concepts, namely typicality. An experiment could then consist in listing different possible properties of the concept *Fruits*, and measuring experimentally the typicalities of these properties. One such property could be *Can be Used to Prepare a Drink*, and its typicality can be measured by asking test subjects to estimate it on a Likert scale, and calculating the average outcome of these estimates. Suppose we now consider the variant *Juicy Fruits* and again measure the typicality of the property *Can be Used to Prepare a Drink*. Obviously, the typicality value will increase. So *Juicy* combined with *Fruits* has changed the value of a measurable observable, such as typicality, and one can easily imagine that the measurable values of other observables will be influenced too. A similar behavior with respect to measurable observables for physical objects is expressed in physics by the notion of 'state', and that is also how we introduced this notion for a concept (Aerts and Gabora, 2005a,b). An exemplar of a concept can then be considered to be also a state of this concept. Indeed *Orange*, as an exemplar of *Fruits*, will obviously increase substantially the measurable observable 'typicality of a property' in the case of the property *Can be Used to Prepare a Drink*. Each concept can then be in states that are more abstract and states that are more concrete. *Orange* is a more concrete state of the concept *Fruits* than *Juicy Fruits*, and both are states that are more concrete than the most abstract state *Fruits* itself. There are two lines that run between 'abstract' and 'concrete' for human concepts. The first line runs from the most abstract, i.e., *Thing*, to the most concrete, i.e., an instantiation of a concept—an instantiation is what psychologists refer to as the realization of a concept in time, and sometimes also in space, if the instantiation is an object. The second line runs from the bare form of the concept, such as *Fruits*, to a qualified form, where the concept appears within a very specific meaning context, e.g., a website on the World-Wide Web. The existence of these two non-coinciding lines for human language is interesting enough, but mainly so for historical reasons, i.e., because of the importance that physical objects in the customary human environment have played in the formation of human language. The most relevant of both lines to the comparison we are making here is the second, where concepts collapsed inside the meaning context of a text, e.g., a website, attaining their most concrete state. Indeed, it is this line running from 'abstract' to 'concrete' that we compare with the states of quantum particles running from 'delocalized' to 'localized'. However, both lines play a role in how the human mind copes with concepts, with their combinations, and with abstract and concrete degrees. For example, the logical connective 'or', put in between two human concepts, e.g., *Fruits* and *Vegetables*, to form the concept *Fruits Or Vegetables*, produces a more abstract state for both concepts, due to the meaning of 'or' in human language. We do not find this abstraction easily represented along the second line, combinations of three concepts, such as *Fruits Or Vegetables*, occur in texts just as combinations of three words. With regard to the above example of the World-Wide Web, this means that text analysis will need to take into account the first line—from abstract to concrete—as well. This is one of the major unsolved problems of semantic space theories and related domains of research, including natural language processing and information retrieval, which is why our approach is of value for the problems encountered in these domains (Aerts and Czachor, 2004; Van Rijsbergen, 2004; Widdows, 2006). In many instances we use the World-Wide Web as the entity playing for the human conceptual realm the role that space-filled-with-pieces-of-matter plays for the micro-physical realm. When we do so, we use the analogy between the two realms by focusing on the second line from 'abstract to concrete' in the human conceptual realm. However, we should bear in mind also to pay attention to the first line from 'abstract to concrete' for being a contributing factor to the meaning carried in texts on the World-Wide Web. So, also in our examples, we are confronted with this difficulty of expressing meaning in language, which is the core difficulty that semantic space theories are confronted with. To be more specific, if a concept from the human conceptual realm, for example the concept *Animal*, is maximally

abstract, it will appear in greatly varying states in many webpages, i.e., it will be strongly delocalized on the World-Wide Web. On the other hand, if we consider a very concrete concept or combination of concepts—the most concrete we can now imagine being the total content of a document on the World-Wide Web—then this 'total' content will be present only in this particular document, i.e., it will be very localized. In other words, in this new interpretation, the delocalization of a quantum entity is interpreted in a similar way, not as a spreading out over space, but as an abstraction of all the parts of space-filled-with-pieces-of-matter where it is not localized. This would also explain why the Heisenberg uncertainty is ontological, and not due to a lack of experimental preciseness. If a quantum particle is a conceptual entity mediating between pieces of matter, it cannot be very abstract and very concrete at once, which means the tradeoff between abstract and concrete is ontological, because of the ontological nature of the quantum entity being conceptual. The new interpretation likewise enables us to understand and explain the weird behavior of quantum entities related to identity. If we consider the concept *Eleven Animals* on its abstract level, all the element animals are identical but ontologically so, because the ontology is conceptual if the entities considered are conceptual. This is exactly how identical quantum entities appear, in theory as well as in experiment. In Section 4.3 of Aerts (2009a) we analyzed how 'identity' behaves for human concepts, more specifically for the concept *Eleven Animals*, its possible states being combinations of *n Cats* and $11 - n$ *Dogs* for $n \in \{0, 1, \ldots, 10, 11\}$, and we showed, by comparing the numbers of webpages on the World-Wide Web and the relative frequency of appearance of the different combinations, that a Bose-Einstein statistics emerges, exactly like it does for bosonic quantum entities. We also identified the presence of entanglement for human concept combinations in Section 2 of Aerts (2009a), notably showing the violation of Bell's inequality, and, using the data from Hampton (1988a,b), in Section 3 of Aerts (2009a), we analyzed how interference of combinations of human concepts appears. See for example Figure 4 of Aerts (2009a) and its analysis for a graphical representation of the interference between the two concepts *Fruits* and *Vegetables*, within the combination *Fruits or Vegetables*.

We have now gathered all the elements that we need to explain 'Why customary macroscopic entities appear to us humans as they do, i.e., as bounded entities occupying space and persisting through time'. We will give a more elaborate answer below but in summary we could say that 'we humans perceive with our senses and mind in a manner unlike that of measuring apparatuses such as those used by physicists in laboratory experiments to detect quantum effects'. In other words, 'the interaction between a human mind, aided by the human eye, and a macroscopic entity, i.e., the entity we identify as a customary macroscopic object, should not be interpreted as belonging to the same category of interactions as those between a customary measurement apparatus used to detect quantum effects and such a macroscopic entity'. They are interactions of a fundamentally different nature. To state it more sharply, for the sake of clarity, we could say that 'the interaction of a human mind, through the human senses of vision and touch—we will analyze below why smell, taste and hearing are different—with a customary macroscopic entity is an interaction 'not' within its own realm of conceptuality', it is in some sense an interaction 'trying to bridge two realms of conceptuality', the first realm being where 'micro-quantum entities interact conceptually with pieces of matter', and the second realm being 'where human minds interact conceptually with memory structures—possibly other human minds, or pieces of text, or the World-Wide Web'. However, in 'seeing or touching a macroscopic customary entity', the human eye, the human fingers and other parts of the body do not interact within one of these conceptual realms. Seeing and touching are in some sense much more primitive types of interaction than those within the two realms mentioned before, namely realm number one, the micro-quantum realm, 'the interaction of bosons with pieces of matter', and realm number two, the human conceptual realm, 'the interaction of words with human memories'. We will not provide a detailed analysis of seeing and touching since this would take us beyond the scope of this article. Instead, we will briefly explain what we mean here.

Seeing takes place by means of light, but mainly by means of a complex interpretation in the visual cortex of the pattern of light falling onto the retina of the eye. Nothing of the quantum nature of light plays any role in this mechanism, on the contrary, the eye has evolved biologically into an organ that can be adequately explained by comparing it to a camera obscura, which is the mechanical environment where the geometrical theory of light fares well, while the visual cortex evolved biologically as well to create a photographic imaging of this pattern on the retina as faithfully as possible. The geometric model for the behavior of light is as far from the quantum behavior of light as we can imagine. Touching is a way of interacting that is profoundly micro-quantum by nature, but only in accordance with one specific quantum rule, namely Pauli's exclusion principle. If we touch a customary macroscopic entity, we try to put our finger, which is also a macroscopic material entity, in the same place as the touched entity. Pauli's exclusion principle forbids this to happen. However, it is essential that both, the customary macroscopic entity and our finger, are composed of fermions, which are the only micro-entities able to form stable pieces of matter (Dyson and Lenard, 1967; Lenard and Dyson, 1968; Lieb, 1976, 1979; Muthaporn and Manoukian, 2004). And the material entities around us and also our finger indeed obey this exclusion principle, for we cannot put them in the same state, being, in this case, in the same place. But Pauli's exclusion principle, although a fundamental rule of quantum theory, is not linked to the typical quantum phenomena, such as interference or entanglement. It is, in some sense, a very classical type of rule persisting in the micro realm, excluding two fermions from being in the same state. This means that our touching sense does not confront us with the quantum nature of macroscopic customary entities either. Let us also note that, if we put forward the question of 'why customary macroscopic entities appear to us humans as they do, i.e., as bounded entities occupying space and persisting through time', we are inclined to think of the two senses of 'seeing' and 'touching', or their prolongations. Indeed, if we were to make a movie of such customary macroscopic entities, the movie would confirm our seeing them, since movie-making is a prolongation of the human sense of seeing, at the same time pushing light into its

geometrically idealized behavior. If we confront such customary macroscopic entities with other such entities, for example by collision, then this is a prolongation of our touching sense, and again Pauli's exclusion principle will determine what happens. What about other human senses, such as smell, taste and hearing? The sentence 'why customary macroscopic entities appear to us humans as they do, i.e., as bounded entities occupying space and persisting through time' would already appear quite differently if we perceived our surrounding reality mainly by smell. To give one example, it would be very easy to create a situation violating Bell's inequalities—much like the situations we proposed in discussing the vessels of water or a connected rod (Aerts, 1982b, 1991)—by considering odors that give rise to correlations in smell. Obviously, we would perceive the world around as much less of a world of clearly separated entities if smell was our main sense. The same is true for taste and hearing. In this sense, it is not a coincidence that what we have called the second realm of conceptuality, the one of human communication, has first emerged through the use of the sense of hearing, namely by means of spoken language. The birth of written language was effectively a major achievement in itself, because the fluidity of spoken language needed to be pushed into the much crisper nature of vision. In this sense, it is not a coincidence either that the invention of the alphabet is seen as a major event in human culture, although even today alphabets are not capable of rendering in a clear way most dialect forms of spoken languages.

As we have seen above, we can explain why humans are not confronted with quantum behavior through the senses of seeing and touching, even though this behavior is profusely apparent on the macro-level—light shining on the skin of our body does react quantum mechanically with our skin, for example, but light entering our eyes behaves along the classical geometric model. This naturally leads to the question of 'What 'are' these customary macroscopic entities, are they quantum or are they not?' This is a question about the ontological status of customary macroscopic entities. Let me go back to some of the quantum phenomena that we described in some detail in the foregoing sections of the present article and attempt to give a nuanced answer to this question. I will also illustrate how, for this question, our new quantum interpretation, and the comparison and analogy of the two realms of conceptual interaction, the micro-realm, and the human realm, can put forward a view that offers an explanation and that is comprehensible. Experiments that aim to detect quantum interference of ever bigger molecules have proved successful (Arndt et al., 1999; Gerlich et al., 2011). The currently most advanced experiments with respect to this quantum phenomenon (Gerlich et al., 2011) make use of organic molecules of up to 430 atoms, and a maximum size of up to 60 angstrom, which is 60×10^{-10} meters, and a de Broglie length of 1 picometer, which is 10^{-12} meters. To get an idea of the relative sizes at play, we could scale them up from angstrom to millimeter. This results in molecules the size of a prune of about 6 centimeter. The de Broglie wavelength, sized up accordingly, would become $\frac{1}{100}$ of a millimeter, which is very small. This means that there are in fact no overlapping de Broglie waves for the molecules in the detected interference. The slits in the grating, hence the separation of the beam into two beamlets, are two orders of magnitude bigger than the size of the molecules. If we scale up the sizes again by the same factor, the two beamlets become separated by 6 m. So, what Gerlich et al. (2011) and his team have done is delocalize a molecule—still according to the scaled-up view—of the size of a prune over a distance of 6 m.

To grasp how spectacular this is, let us restate in more detail what such a delocalization actually is. It means that if we attempted to detect the molecule in spot A, a spot inside one of both beamlets, the probability of finding it in this spot A would be equal to 1/2. The same holds for a spot B in the other beamlet, while A and B are 6 m apart. If we mention only this aspect of delocalization, we can still propose a classical explanation for this, imagining that the molecule just chooses one of the beamlets at the point where the beam splits into two parts, i.e., long before in space and time it reaches one of the spots A or B. But there are other experiments that can be performed to demonstrate that this cannot be the case, and that the molecule is in a state of superposition between 'being in A' and 'being in B' at the moment it passes where spots A and B are located. Some physicists express this by stating that the molecule is in the two places at once, while others say that the molecule is neither in A nor in B, considering the superposition state as a new emergent state, not localized in space, hence not spatial. As long as such experiments were done with very small quantum entities, such as photons, electrons, or neutrons, we could also still imagine the quantum entity as being spread out, like a wave. But in the case of big entities with complicated internal structures, such as the molecules consisting of 430 atoms referred to above, this is no longer possible. Indeed, what is important to note in this respect, is that the internal structure of the molecule is not affected at all by this superposing. Whenever an attempt is made to detect the molecule, it is detected, unaffected, and in its entirety. This means that this superposing effect does not in any way affect the internal structure of the molecule, it is an effect happening on the level of the ontology of the molecule, on the level of 'what the molecule is'.

With smaller quantum entities, delocalization of much greater size has been realized. As early as in the 1970ies, Helmut Rauch delocalized a neutron in a similar double-slit setup, over a distance that, if we scale up the neutron to the size of a prune, would be equal to several thousand kilometers (Rauch, 1975, 2000). What is most relevant, however, and also crucial for the central reflection of this discussion, is that Gerlich et al. (2011) realize a delocalization which is without any doubt big enough—also taken into account the size of the corresponding de Broglie wavelength—to conclude that the same quantum phenomenon is at play here as that observed on so many occasions with small and more typical quantum entities, such as photons, electrons, or neutrons. We should add, however, that 'the detection of the quantum interference effect is only possible with a specific experimental arrangement specially made for the detection of delocalization', namely the whole experimental setup of a double-slit for these sizes of molecules. Does this mean that an adapted experimental setup will enable us to put a chair or a table or any one of our customary macroscopic entities into a state of superposition of two widely separated places? It seems that this is indeed what these experiments indicate. Of course, it might well be that this will not be possible experimentally for many years to come, or indeed, that interference for large entities such as chairs or tables will

remain out of experimental reach (almost) forever. This, however, does not change the fact that 'in principle also chairs, tables, and any customary macroscopic entity are ontologically of the same nature as these huge organic molecules'. Is it possible to comprehend this? The following example serves to illustrate that our new quantum interpretation puts forward a simple and plausible explanation.

In Aerts (2009b), Section 3, we investigated in detail the situation of the two concepts *Fruits* and *Vegetables* and their combination *Fruits or Vegetables*, and showed how data collected in Hampton (1988b) revealed the effect of interference. A graphical representation of the pattern of this interference is shown in Figure 4 of Aerts (2009b). Of course, for two concepts such as *Fruits* and *Vegetables*, there is no problem at all to imagine that the new concept *Fruits or Vegetables* is a state which is neither *Fruits* nor *Vegetables*, but a new state, namely the state *Fruits or Vegetables*. First of all, if we consider *Fruits or Vegetables* as a concept in itself, then both *Fruits* and *Vegetables* are more concrete states of this concept. On the other hand – this was even the single subject of investigation of Hampton (1988b)—typicalities of membership of exemplars of *Fruits* and *Vegetables* change in ways that are not compatible with considering *Fruits or Vegetables* as a category that would allow being presented as a set theoretic union of representations of the categories *Fruits* and *Vegetables* in a set theoretic way, and this impossibility is a well-known fingerprint of the presence of quantum structure. *Tomato* is an exemplar where this effect can be readily and even intuitively understood; indeed, it is an exemplar that fits well in the new category *Fruits or Vegetables*, because it is an entity that people are likely to have doubts about when asked to classify it as an exemplar of either *Fruits* or *Vegetables*. Why is there no problem at all to consider *Fruits or Vegetables* as a new state, and why is there a problem to do this for *Molecule at spot A* or *Molecule at spot B*? The reason is to be found in a profound difference between the notion of 'concept' and the notion of 'object'. More specifically, there is a fundamental difference between the relation that a 'concept' can have with the connective 'or' and the relation that an 'object' can have with the connective 'or'. Indeed, two concepts, such as *Fruits* and *Vegetables*, when connected by 'or', give rise to a concept. However, two objects, when connected by 'or', do not give rise to an object. More concretely, a 'chair at spot *A*' 'or' 'chair at spot *B*' is 'not' an object. A mathematician would say that the set of concepts is closed for the operation of disjunction, while the set of objects is not. We claim that this is the fundamental reason why quantum theory will keep leading to situations that we do not understand, and that we cannot understand, as long as physical entities are believed to be objects. If, as is the case in our new quantum interpretation, quantum entities are considered to be concepts, the problems of understanding the double-slit interference type of situation disappears. Note that our new quantum interpretation, and the experiments proving quantum superposition behavior for macroscopic entities, such as these organic molecules, entail that these macroscopic entities are concepts rather than objects, but concepts of such a type that their 'way of being' closely resembles what we imagine objects to be—we will elaborate on this in the following paragraph. In other words, if we replace the notion of 'physical object' for a quantum entity by the notion of 'conceptual entity', and we interpret the process of 'a quantum entity becoming more localized' as a process of 'this conceptual entity becoming more concrete', we can understand that such a quantum entity as a conceptual entity can be 'localized in spot *A*' 'or' 'localized in spot *B*', and that 'this' is one of its genuine ontological states. This is what the ontology of a superposition state is according to our new quantum interpretation.

The next question that arises is whether our new interpretation enables us to understand why large conceptual entities gradually become more and more like objects. The answer is affirmative, for if we analyze what happens in the human realm with conceptual entities, we can see a rather surprising phenomenon, which is that the behavior of larger entities approaches that of objects. For combinations of human concepts consisting of a small number of concepts there is, at first sight at least, still a symmetry between the use of the connective 'or' and the use of the connective 'and', both being used more or less in the same way. We can intuitively understand this when we look at examples of combinations of two concepts, *Fruits* and *Vegetables*. Combining them to give rise to the new concept *Fruits and Vegetables*, or combining them to give rise to the new concept *Fruits or Vegetables*, takes place on the same footing, the one not being more special than the other. If, however, we consider larger sets of combinations of concepts, the symmetry between the 'or' and 'and' connective is broken, with the dominance of 'and' increasing as the set of combinations of concepts grows in size. Let us remark that, although the 'or' connective is not compatible with the notion of 'object', i.e., object *A* 'or' object *B* is not an object, the 'and' connective is compatible with the notion of object. Indeed, object *A* 'and' object *B* is again an object, namely the object consisting of both objects *A* and *B*. Let us now consider a typical large set of combinations of concepts, for example all those that together make up a story. And let us consider two of such stories, story *A* and story *B*. Then story *A* 'and' story *B* can still be considered to be a story, namely a story consisting of the two stories *A* and *B*. But story *A* 'or' story *B* is not a story. It has no longer the form that we expect a story to have. So here, on the level of the size of concept combinations that we call stories, we can intuitively recognize the breaking of symmetry between 'and' and 'or'. In Aerts (2013) we explicitly investigated this breaking of symmetry between 'and' and 'or' in the texts of documents on the World-Wide Web, and we found the following results. Let us first mention that the experiment we did on the World-Wide Web took place on September 15, 2011, using the Yahoo search engine, so that is the source of our numbers. We found that the asymmetry already appears at the level of combinations of two concepts. Choosing two random concepts, *Table* and *Sun*, and combining them by means of 'and' and 'or', respectively, we found a proportion of 72 to 1, i.e., there were 72 times more documents containing *Table and Sun* than documents containing *Table or Sun*. Larger sets of combinations of concepts made the proportion go up in favor of 'and'. However, when we considered some specific combinations, the proportion shifted in favor of 'or'. Let us give some examples of where this was the case: *The Window or The Door* appeared 2.5 times more often than *The Window and The Door*, *To Laugh or To Cry* appeared 10 times more often than *To Laugh and to Cry*, *Dead or Alive* appeared 100 times more often than *Dead*

and Alive, *Wants Coffee or Tea* appeared 50 times more often than *Wants Coffee and Tea*. How to understand this phenomenon of symmetry breaking? Well, the 'or' will remain abundant in expressions that 'almost form a concept on their own again'. The three expressions *To Laugh or To Cry*, *Dead or Alive* and *Coffee or Tea* are good examples of this. While no new word has been attributed to them, they abound as 'stable combinations' of their constituent concepts *Laugh*, *Cry*, *Dead*, *Alive*, *Coffee* and *Tea*. In addition, the combinations of the 'and' in the three cases is not common, since both constituents are opposites. For *Laugh* and *Cry*, and *Dead* and *Alive*, this oppositeness is clear, but also in the case of *Coffee* and *Tea*, most of the meaningful sentences on the World-Wide Web including these two concepts are likely to refer to situations in which somebody chooses between coffee 'or' tea. Although in both cases, of course, also the 'and' remains meaningful, e.g., in sentences such as, "At the party, trays were carried around with coffee 'and' tea to choose from". The case of *The Window or The Door* is interesting too. Although not quite as strong as in the combinations of *Dead* and *Alive* or *Coffee* and *Tea*, there is a certain connection in meaning in the combination of *Window* and *Door* too. One can intuitively understand that this connection will be stronger in the combination using the connective 'or' (e.g., in the sentence, 'Will he escape through the window or the door?') than in combinations using the connective 'and'.

As we can see, the symmetry breaking between 'or' and 'and' is of a subtle nature. It is not a symmetry breaking that favors either of them in any definite way. However, when the notions of story, memory, pieces of text, etc... in the case of human concepts, and space-filled-with-pieces-of-matter, in the case of micro-quantum-entities, are taken to be a focus of attention, the 'and' becomes dominant with respect to the 'or', when it comes to formation of (i) random new concept combinations, and (ii) ever larger new concept combinations. The 'or' remains dominant for small, abundant stable combinations, and, also for the formation of new concepts. Indeed, the concept *Animal* is a combination that makes use of the 'or'—indeed, it is *Dog or Cat or* . . .. However, we do not encounter it in this large combined form in texts, but as one word *Animal*. So 'abstraction' is an operation that makes use of the 'or'. Does this make the 'or' dominant if it comes to the formation of new concepts that are indicated by one word? Not exactly. For example, the concept *Dog* is not formed out of 'or' combinations, but rather out of 'and' combinations of more abstract types of events that have not even been given names of their own. Here are some descriptions *Running around on fast moving legs and wagging its tail*, *Jumping up against me and quickly disappearing again*, *Chasing cats in the garden*, etc . . . So, *Dog* is formed out of combinations of 'and' of many such short-lasting real-life events. Going back to the realm of micro-quantum entities, we can say that, in our view, the abundance of 'unstable particles' should be interpreted in this way.

Let us extend the analogy to further clarify the state of affairs with respect to macroscopic quantum phenomena. We already mentioned that, when thinking of stories as large collections of combinations of concepts, we have the tendency to allow story *A* 'and' story *B* to be a story again, namely the two stories *A* and *B*—which is completely similar to how we allow object *A* and object *B* to be an object again, namely the two objects *A* and *B*—, and not to allow symmetry for the 'or' in both cases. Indeed, story *A* 'or' story *B* is no longer considered to be a story. This, however, does not mean that we do not encounter specific situations in everyday life where story *A* 'or' story *B* represents what 'is actually happening' in our cognitive reality. Imagine a situation where participants in a quiz are shown a small part of a video, and then asked by the host to choose one from a number of possible continuations. This quiz situation does not make these alternative continuations of the story, now combined by the connective 'or', into one story again, but it does make this 'superposition of stories' what the candidates are confronted with in their cognitive reality. And every other quiz type of situation will confront the participants with similar superpositions of concepts not frequently found in documents on the World-Wide Web. We will now make full use of the explicative potential of our consideration of the analogies of the two realms, viz. human cognition and micro-quantum, and look again at the experiment in Gerlich et al. (2011). We can say that, by producing a beam of large organic molecules that is split into two beamlets when it passes through a double slit, Gerlich et al. (2011) are putting each of the molecules into a quiz situation, with respect to spot *A* and spot *B*, each located in one of the beamlets. However, they do not force the molecules to choose, because they want to measure interference. So the molecules are allowed to stay in superposition, wondering which of the two stories proposed by the host of the quiz, story *A* 'or' story *B*, to choose, if they were forced to do so. This is what would happen in Gerlich et al. (2011)'s experiment in case we attempted to find the molecules in *A* or *B*, which would destroy the interference, as we know from the typical analysis of the double-slit experiment situation. A real human cognition analogy for the whole experiment, with interference, would therefore be as follows: Someone is in superposition because of the choice between two possible stories, story *A* 'or' story *B*, but does not choose, and is not revealed anything about what happened either, and is then confronted with a third choice, between *C* or not *C*, which is the equivalent of the molecule being detected or not being detected. Interference is how the pondering in superposition between *A* 'or' *B* influences the choice between *C* 'or' not *C*. This is what we modeled for Hampton's data (Hampton, 1988b) and the *Fruits or Vegetables* interference in Aerts (2009b), Section 3 and Figure 4.

We have now reviewed all elements to make the loop back to the contents of both Section 2 and Section 3, and we will start with the latter. Saying that statement A—'standard quantum theory is incomplete, in the sense that it cannot describe the compound entity consisting of two separated subentities'—or statement B—'such a compound entity does not exist or, in other words, whenever two quantum entities exist, their compound entity is not separated' —has been shown to be correct and/or false, would be too simple a statement indeed. We can clearly illustrate this by pursuing our analogy between the human conceptual realm and the micro-quantum realm. We will do so by means of a Gedanken experiment that is easy to perform. Consider two rooms and two groups of people, each group having a meeting in one of the rooms. The question we want to consider is a very simple one: "What are the factors that determine

whether members of one group will be able to understand the conversation of the other group, and vice versa?" Two obvious factors will be (a) 'how loud the people speak that participate in the meetings', and (b) 'how well the rooms are isolated from each other'. These are also the main factors to consider if the problem was approached by an architect. Another option would be to test the rooms without a meeting taking place, making artificial noise at a given level of decibels in one room, and measure how loud the noise is in the other room. This goes to show that, for the realm of human cognition, obviously 'separated entities exist'—we just need to provide the walls of the rooms with adequate isolation.—We should add that the two rooms do not even need to share a common wall, indeed, they might even be rooms in different houses, so there is no doubt that the two groups can be separated to the extent that nothing talked about by the one group can be understood by the other group, and vice versa. What we proved in Aerts (1982a, 1983a) is that 'these two well isolated groups, and their cognitive interactions, cannot be modeled in a standard quantum theory using the standard Hilbert space formalism'. The mathematical structure of Hilbert space warrants the creation of states that carry correlations in meaning between the two groups. These are the entangled states. And, coming back to the more detailed situation of also considering the presence of dynamical interaction, quite obviously this type of interaction exists between the objects in both rooms, be it only gravitational interaction.

What about an analogous situation involving micro-quantum entities? We believe that the only statement that can be made now is that 'we do not know' because no experiments have been considered to test which ones of the statements A or B is correct. Quantum entities have the tendency to entangle whenever they are in situations where we would also suppose concepts to entangle. Indeed, here too, the analogy with human communication is enlightening. Humans cannot avoid understanding what other humans say, whenever a number of conditions are fulfilled. One such condition is that the loudness of speech is subject to a minimum level. But this is certainly not the only condition, because it also depends on what is being said, for example, whether the context can be guessed more or less by the listener, or not at all, as well as on quite a number of other elements connected to the meaning of what is said. Another factor is probability. Repeated experiments using the same sentences in one room produce only probabilistic outcomes, particularly if different humans participate in the experiment. What quantum entities do in similar conditions, when for example a conscious attempt is made to shield them off, has not been tested. It is relevant at this stage of our analysis to point out that there is a crucial difference between the above experiment and an experiment that consists in measuring 'the distance by which two humans can be separated from each other in space, such that the one can still understand what the other is saying, if we are allowed to use any available technical means to conserve the meaning of the sentences uttered by the speaker, and to optimize the understanding capacity of the listener'. An example of the latter kind of experiment is that of human's first flight to the Moon in 1968, when Apollo 8 circled around it and its occupants talked with people on Earth, over a distance of 400,000 kilometers. And there is no doubt that much larger distances are possible. Hence, following the above analysis, we can conclude that statement B might well be false, and statement A be true. This would mean that separated quantum entities do exist, and that standard quantum theory fails to model them and is therefore an incomplete theory, albeit not incomplete in the sense that hidden variables need to be added. Rather, the incompleteness can be remedied if we move to a more general quantum-like theory, such as that developed by the Geneva-Brussels School.

Linking up with our analysis in Section 2, we believe that its wave-particle line of reasoning is of value, but only in a relative sense. It would be possible to introduce a notion at least intuitively similar to that of the de Broglie wavelength. To illustrate this, we return to the combination of the concepts *Fruits* and *Vegetables* into *Fruits or Vegetables*. As Hampton's measurements data showed (Hampton, 1988b), and as we analyzed in Aerts (2009a), Section 3, for example Figure 4, there is strong interference. *Fruits or Vegetables* really forms a new state of a concept and many exemplars overextend, which means that they are felt by the participants to fit better in this new concept state than in any of the two component concept states. Hence, if we had to think of an analog of the de Broglie wavelength, it would be natural to consider the ones of *Fruits* and *Vegetables* as very overlapping, like the ones of electrons inside an atom. In the earlier example of the different options *A* and *B* proposed in the quiz as continuations of a video fragment, the connective 'or' between both options functions in such a way that when an analog with the de Broglie wavelength, is made, the wavelength will be very small. This 'is' why we do not consider story *A* 'or' story *B* as a new story—their de Broglie waves hardly overlap. An exception would be if both stories resonate strongly with each other in terms of meaning content. For example, if one story contains clues to understand the other story, or vice versa. So, an intuitive analog of the de Broglie wavelength will depend on many aspects of a piece of text, particularly its meaning content. Whether two concepts and/or two texts have overlapping waves will hence also depend on the degree of resonance between the meanings of the respective concepts and/or texts. The resonance is likely to be strong if there are only two simple concepts. But even in the case of combinations of single concepts, the role played by this aspect is obvious. We would not be able to find a lot of interference for randomly chosen concepts, such as *Table* and *Sun*, combined into *Table or Sun*. Concluding about the de Broglie wavelength type of reasoning, even for material entities, most probably the reasoning needs to be considered as a useful guide but also as an idealization, certainly for larger material entities. So who knows what new interference experiments will reveal with respect to material entities of a much bigger size than the organic molecules tested in Gerlich et al. (2011). The future will have to show.

In the foregoing we analyzed the role of 'size' and, more concretely, how larger pieces of text, such as stories, behave more like objects when compared to smaller pieces of text or single concepts. For the realm of human cognition, we also indicated in which way the meaning content of each of the pieces of text plays a role in their potential for quantum behavior. Amongst the examples of macroscopic quantum behavior within the realm of the micro-quantum world, which we described in Section 2, only

the laser is realized at 'room temperature', i.e., in our customary human environment. At least some of the household appliances in many of today's homes have lasers. The quantum behavior of the other examples, superfluidity, supercurrency, and all realizations of Bose-Einstein condensates, originally only appear at very low temperatures. They have not found their way yet to people's homes, because of the complicated techniques that are required. Magnetic Resonance Machines in hospitals make use of supercurrents to create very strong magnetic fields used in the imaging. This means it is likely that quite a number of us, perhaps without being aware of it, have already been in a machine operated primordially by means of a Bose-Einstein condensate, in the form of a supercurrent. The fact that the laser is an exception to the need for strong cooling to realize a material macroscopic quantum entity, is linked to the special nature of photons and their capacity to escape the disturbances that random packets of heat energy customarily bring to configurations of matter. How do we have to understand this 'disturbance due to heat' throughout the analysis we have developed in the foregoing sections?

Before we reflect about this question, we should mention that quantum experimentalists, definitely wizards of our time, have by now moved their exploits all the way up to room temperature. Recently, scientists created a Bose-Einstein condensate, using a thin non-crystalline polymer film of approximately 35 nanometers thick—for comparison, a sheet of paper is about 100,000 nanometers thick—, in the form of a layer placed between two mirrors and excited with laser light, and the quantum state was realized at room temperature. The bosonic particles are created through interaction of the polymer material and light which bounces back and forth between the two mirrors. The phenomenon only lasts for a few picoseconds—one trillionth of a second—, but long enough to use the bosons to create a source of laser-like light (Plumhof et al., 2014). The realization of this room-temperature Bose-Einstein condensate is the result of an ever deeper quantum physical exploration of condensed matter. The bosons that condensate—appearing all in the same state—are cavity exiton-polaritons, which are quasi-particles arising from the coupling of excitons—i.e., bound states of an electron and an electron hole—and photons. To be able to understand the meaning of the room-temperature Bose-Einstein condensation, we should elaborate on what a 'quasi-particle' is, as it is now commonly used as a notion in condensed matter physics. In principle, matter consists only of combinations of three quantum particles, namely electrons, neutrons, and protons. Quasiparticles are an emergent phenomenon that occurs inside matter as a consequence of the strong interactions that exist between all electrons, neutrons, and protons, in whatever constellation these appear inside matter. Hence, a way to look at it is that a quasiparticle is an idealized substitute for the motions of the real particles inside matter, which are much too complicated to be able to be modeled. In that sense, quasi-particles are not real particles, e.g., they cannot exist outside matter. We already encountered such quasiparticles, namely phonons that play a role in the supercurrency through cooper-pairing of electrons. In this respect, it should be noted that according to some ideas in today's physics community real quantum particles are considered quasiparticles of an aether described by the quantum vacuum (Wilczek, 2008), but independently of these ideas, when we define quantum by means of the characteristic of its behavior, these quasiparticles are quantum. And, if we go a step further and define quantum by means of the nature of the mathematical structure involved in the modeling of the phenomenon, they are quantum too, because they are defined by the mathematical formalism of quantum theory itself. We analyze only one example here because it would take us too far to go into the details of what is happening with respect to quantum structures in solid state physics, where an abundance of quantum effects are identified under well-controlled laboratory conditions (Kasprzak, 2006; Lagoudakis et al., 2008).

Our analysis of the role of temperature in the appearance of quantum effects makes it relevant to mention the findings of quantum effect in biology, for example in the process of photo-synthesis (Engel et al., 2007; Sarovar et al., 2010; Scholes, 2010). The quantum effect identified in biology are 'at room temperature'—or more correctly, at earth crust temperature. Given the above, the question arises, What about the role of temperature? We do believe that the original reasoning related to the de Broglie wavelength, which we put forward in detail in Section 2, namely that temperature, being a measure of the random behavior of energy, is a disturbing factor destroying the potential for quantum coherence, is true to a great extent, but needs to be generalized. To be more concrete, it explains why cars on a highway —and chairs and tables in our living rooms— do not quantum cohere as macroscopic material entities within their natural environment, which is an environment where their intrinsic quantum nature as entities is too much disturbed by random packets of heat energy bombarding them. But, why then do there appear quantum effects of coherence, in biological entities (Engel et al., 2007; Sarovar et al., 2010; Scholes, 2010)—in photo-synthesis, but most probably also in many other biological processes yet to discover—, and in solid state matter entities at room temperature (Plumhof et al., 2014) in controlled laboratory conditions? Could it be that the temperature should not be looked upon as providing an objective scale indicating the situations favorable for the appearance of quantum effect? In fact, if we reflect about the explanation of how temperature is destructive for the presence of quantum coherence, the answer is contained in it. It is because of the disturbing effect of the random bombardment of heat energy packets that quantum coherence disappears. The size of this bombardment depends crucially on the temperature, and hence not on whether an entity is a plant making use of photo-synthesis or whether an entity is the chair or table in our living room, or a car on the highway. However, could it not be that the plant has managed to be less disturbed by this bombardment of random heat packets of energy in the processes that enable it to use photosynthesis, and that this capacity hence could lead to the presence of quantum effects? Of course this is possible, and even plausible, if we take into account the mechanism of biological evolution that has played a fundamental role in what the plant is, and how photo-synthesis works. Does this also explain the appearance of quantum effect in human laboratories at room temperature? Indeed, human culture is also an evolutionary process, albeit not Darwinian. It has not only managed resistance against the random bombardment of heat energy packets, but also evolved to use this heat energy and make it into non-random

energy. Human's energy-harvesting from heat started with the first steam engine, which literally is the transformation of random energy into structured energy. Does this produce quantum structure too? Not always, and not automatically, but this is certainly the case for the energy used in those laboratories that have produced quantum effect at room-temperature. What about the vessels of water and other macroscopic situations we invented to violate Bell's inequalities (Aerts, 1982b, 1991; Aerts et al., 2000), and the identification of quantum structure in cognition (Aerts and Aerts, 1995; Gabora and Aerts, 2002; Aerts and Gabora, 2005a,b; Aerts, 2009b; Aerts and Sozzo, 2011; Sozzo, 2014)? Well, the vessels of water and the other entities violating Bell's inequalities are realized within human culture, so that they can be said to have been specially devised to violate Bell's inequalities, albeit not in explicit laboratory situations. In doing so, they make use of all knowledge available to achieve this. As regards the presence of quantum structure in human cognition, we note that human cognition is a product of human culture, and hence profits from the mechanism of cultural evolution to fight the random destructive effect of bombardments of energy packets of heat.

Does this mechanism of cultural evolution strive specifically toward a presence of quantum structure? In this respect, we cannot but refer to the second law of thermodynamics, which states that, for a closed entity, entropy never decreases. To cool down the atoms in a gas for the realization of a Bose-Einstein condensate, experimentalists need to create an enormous decrease of the entropy of the gas. Of course, this is not in contradiction with the second law of thermodynamics, since the gas is not a closed entity during the experiment. Erwin Schrödinger, one of the founding fathers of quantum theory, wrote a seminal book, entitled 'What is life', in which he puts forward several ideas on the nature of life. One of his ideas was that the order that characterizes life is realized as a decrease of entropy within a non-closed entity, while another one is about the genetic code being guarded within an aperiodic crystal, later to be identified as DNA. The way Schrödinger arrived at the second idea is interesting for the line of reasoning developed in the present article. According to Schrödinger's analysis, the carrier of replicated information for life must have sufficient stability and permanence, and must therefore be solid, a gas or a liquid not being suitable. Solids are crystals, except if they are liquids with a very high viscosity. However, crystals are repetitive structures, hence much less capable of coding a big amount of information, which is why Schrödinger argued that an 'aperiodic crystal' should be the principle element in the process of life. This aperiodic crystal for all life existing on earth turned out to be Deoxyribonucleic acid or DNA. It is a nucleic acid in the form of a double-stranded helix, consisting of two long biopolymers made of simpler units called nucleotides, each of which is composed of a nucleobase of one of the following four types, guanine, adenine, thymine, or cytosine, with the letters G, A, T, and C, are used to indicate the bases. What is interesting for our analysis is that the letters G, A, T, and C, are customarily referred to as elements of an alphabet. But is not the alphabet a human invention characteristic of the written language? Let us note in this respect that the oldest written languages, Chinese and its variants, do not use an alphabet but symbols that directly indicate the meaning carriers themselves, i.e. the words—or parts of words. The origin of the alphabet goes back to Egyptian writing, which had a set of some 24 hieroglyphs to represent syllables that begin with a single consonant of their language. But it would be wrong, at least with respect to the analysis we are making, to connect the mechanism of introducing an alphabet specifically to written language. Indeed, the real challenge to human culture in this respect dates back much further, to the advent and development of language itself, i.e., spoken language. This challenge was to express an enormous amount of meaning by using only a very limited number of basic sounds—the consonants and vowels of spoken language, which are also the items to which later written alphabets correspond—, and making combinations of these basic sounds to create meaning carriers, i.e. words, sentences and longer pieces of language. It is an example where human culture has taken a path similar or better, in prolongation of, life.

And what about quantum structures? Let us say that we can still distinguish two types of their appearance in the practice of scientists involved, a distinction that is also made in the relevant scientific literature. The first type of appearance is when it is identified by scientists as 'climbing out of its natural environment, which is the micro-world, or, in case of the macro-world, a world where the disturbing factor of heat is taken away, hence a very cold world'. We can find examples in how it is currently being encountered in widely separated micro-entities (Tittel et al., 1998; Salart et al., 2008), large organic molecules, (Gerlich et al., 2011), in room-temperature states of solids (Lagoudakis et al., 2008; Plumhof et al., 2014) and in biological entities (Engel et al., 2007; Sarovar et al., 2010; Scholes, 2010). The second type of appearance is when it is identified by scientists by looking at the intrinsic structure of the reliable models of its behavior—for example, whether Bell's inequalities are violated, whether interference and/or entanglement can be identified in the data—, independently of whether there is a suspicion of 'climbing out of its natural environment'. Examples of this are how it is being encountered today in ordinary macroscopic entities (Aerts, 1982b, 1991), cognition (Aerts and Aerts, 1995; Gabora and Aerts, 2002; Aerts and Gabora, 2005a,b; Aerts, 2009b; Pothos and Busemeyer, 2009; Busemeyer et al., 2011; Busemeyer and Bruza, 2012; Sozzo, 2014), economics and biology (Bruza et al., 2007, 2008, 2009; Khrennikov, 2010; Song et al., 2011). Our proposal, following the above analysis, is that both are not different in essence, and hence the need to investigate whether it would be possible to connect the appearance of quantum structure with the presence of organized parts of the world—organized matter, organized life and organized culture—, and by 'organized' we mean 'able to conquer the random influences that destroy quantum coherence'—such as random packets of heat energy—, but this should only be one of the examples in such a broader view (Aerts and Sozzo, 2014). It will of course be necessary to thoroughly investigate the connections with the second law of thermodynamics and evolution theory to work out this view further and in greater depth.

The following brief comment is about the philosophical status of the quantum conceptual interpretation which we have used as an element of the analysis presented here (Aerts, 2009a, 2010a,b, 2013), and about the philosophical status of the analysis itself. It might be thought that this quantum conceptual interpretation

presupposes an idealistic philosophical stance. Let us make clear that this is not what we believe to be true a priori. The aforementioned Geneva-Brussels School quantum theory was conceived, certainly in its original formulation, within a philosophical stance of 'non naive realism'. Indeed, one of its philosophical aims was to prove that a realistic philosophical view is compatible with quantum theory. When I first started to reflect about the idea that 'quantum entities might well be conceptual entities', this was not with an inclination toward idealism as a philosophical stance. There is a subtle but very easy point to be made in this respect, which clearly shows the difference between a realist view on conceptuality and a possible idealist one. Indeed, we can again consider the same example which we have used so many times now to make things clear, namely the situation in the human realm. When two humans talk to each other, they exchange sentences of concepts and their aim, usually, is to transfer meaning. This process is 'really' taking place, within 'ordinary daily reality'. The concepts that are used in such a conversation are 'real'. 'That' is how I have been considering quantum entities to be conceptual entities, namely as 'real entities' of a conceptual nature, engaging in an exchange of 'real meaning' between pieces of matter, functioning as proto-memory entities. In exactly the same way that human conversations as processes of exchange between real memory structures, i.e., human minds, materialized in human brains, but also computer memories, making use of concept combinations in a real language, 'exist'—in the usual sense of the word—, one can imagine that quantum entities are really existing conceptual entities mediating between really existing proto-memory structures which are pieces of baryonic matter. Within such a, what I would like to call in a somewhat challenging way, 'non-naive realist view on conceptuality', there is no difference in principle between the two realms with respect to their 'nature of reality'. Any further philosophical question about the deeper nature of the foundations of one of the two realms can be translated right away into the same philosophical question about the deeper nature of the foundations of the other realm. A platonic type of question of whether concepts exist prior to physical objects—and in our interpretation such physical objects are also conceptual, we will get to this shortly—can equally well be put on the table in both realms, the one of human communication, or the one of micro-physical conceptuality, following from our quantum interpretation. However, it is not necessary at all to make such a philosophical choice between idealism and conceptual realism to understand and explain what we wanted to understand and explain in the first place. Human culture and how it evolved can be fully understood and explained by 'only' supposing the existence of the conceptual entities that 'have come into existence through a historically real human exchange'. Or again, to make the same distinction, but this time focusing on the written conceptual structures, human culture, and its evolution, can be fully explained by considering the books that really have been written, as well as the libraries containing them. Idealism is reasoning about the conversations that 'could have taken place', and the 'books that could have been written'. A realist would say that 'these do not exist', but indeed 'could have existed', but that is a different matter. Of course, the above, showing that a realist philosophical view on conceptuality is possible, does not prove that the world is as such. The more conceptual entities play an important role on a fundamental level, as in the case of what concerns my conceptual quantum interpretation, the more it becomes natural to also wonder about the possibility of an idealist philosophical stance as a foundation. Let us be more specific about the conceptual status regarding pieces of baryonic matter, although this does seem to imply that there is no possibility for objects to play any role as foundational elements, philosophically speaking, this is not true either. Let me again illustrate this by means of a specific possibility, e.g., the strong resistance in unifying gravitation with quantum theory might well indicate that on the level of where gravitation works 'objects' in the traditional sense do exist, and that it is only on the level of where quantum theory works that conceptuality is the rule. I do not want to exclude such a possibility at this stage of research with respect to it. In this sense, to make the above more specific, I would prefer not to have to opt a priori for a specific philosophical stance within this quantum conceptual interpretation, but rather leave it to further research to gather new experimental data, and ways to explain them, to give weight to the different possible philosophical stances. This does not mean that it would not be interesting to already consider these different stances taking explicitly into account this quantum conceptual interpretation as well, and I am planning to write about this in future work.

ACKNOWLEDGMENTS

I thank Sandro Sozzo and Massimiliano Sassoli de Bianchi, two of my close collaborators, for providing me with very valuable and stimulating comments and suggestions after reading the manuscript. I also thank the reviewers for interesting and worthy comments and suggestions. All these interactions have helped me in formulating several parts of the manuscript in a more clear way.

REFERENCES

Aerts, D. (1982a). Description of many physical entities without the paradoxes encountered in quantum mechanics. *Found. Phys.* 12, 1131–1170. doi: 10.1007/BF00729621

Aerts, D. (1982b). Example of a macroscopical situation that violates Bell inequalities. *Lett. Nuovo Cimento* 34, 107–111. doi: 10.1007/BF02817207

Aerts, D. (1983a). "The description of one and many physical systems," in *Foundations of Quantum Mechanics*, ed C. Gruber (Lausanne: AVCP), 63–148.

Aerts, D. (1983b). Classical-theories and non-classical theories as a special case of a more general theory. *J. Math. Phys.* 24, 2441–2453. doi: 10.1063/1.525626

Aerts, D. (1984). The missing elements of reality in the description of quantum mechanics of the EPR paradox situation. *Helv. Phys. Acta* 57, 421–428.

Aerts, D. (1986). A possible explanation for the probabilities of quantum mechanics. *J. Math. Phys.* 27, 202–210. doi: 10.1063/1.527362

Aerts, D. (1991). A mechanistic classical laboratory situation violating the Bell inequalities with $2\sqrt{2}$, exactly 'in the same way' as its violations by the EPR experiments. *Helv. Phys. Acta* 64, 1–23.

Aerts, D. (1999a). Foundations of quantum physics: a general realistic and operational approach. *Int. J. Theor. Phys.* 38, 289–358. doi: 10.1023/A:1026605829007

Aerts, D. (1999b). "Quantum mechanics: structures, axioms and paradoxes," in *Quantum Mechanics and the Nature of Reality*, eds D. Aerts and J. Pykacz (Dordrecht: Springer), 141–205. doi: 10.1007/978-94-017-2834-8-6

Aerts, D. (2009a). Quantum particles as conceptual entities: a possible explanatory framework for quantum theory. *Found. Sci.* 14, 361–411. doi: 10.1007/s10699-009-9166-y

Aerts, D. (2009b). Quantum structure in cognition. *J. Math. Psychol.* 53, 314–348. doi: 10.1016/j.jmp.2009.04.005

Aerts, D. (2009c). "Quantum axiomatics," in *Handbook of Quantum Logic and Quantum Structures*, eds K. Engesser, D. Gabbay, and D. Lehmann (Amsterdam: Elsevier), 79–126.

Aerts, D. (2010a). Interpreting quantum particles as conceptual entities. *Int. J. Theor. Phys.* 49, 2950–2970. doi: 10.1007/s10773-010-0440-0

Aerts, D. (2010b). A potentiality and conceptuality interpretation of quantum physics. *Philosophica* 83, 15–52.

Aerts, D. (2013). La mecànica cuántica y la conceptualidad: Sobre materia, historias, semántica y espacio-tiempo. *Scientiae Studia*, 11, (translated from: Aerts, D. (2012). Quantum theory and conceptuality: Matter, stories, semantics and space-time. arxiv.org/abs/1110.4766), 75–100.

Aerts, D., and Aerts, S. (1995). Applications of quantum statistics in psychological studies of decision processes. *Found. Sci.* 1, 85–97.

Aerts, D., Aerts, S., Broekaert, J., and Gabora, L. (2000). The violation of Bell inequalities in the macroworld. *Found. Phys.* 30, 1387–1414. doi: 10.1023/A:1026449716544

Aerts, D., Broekaert, J., and Gabora, L. (2011). A case for applying an abstracted quantum formalism to cognition. *New Ideas Psychol.* 29, 136–146. doi: 10.1016/j.newideapsych.2010.06.002

Aerts, D., Broekaert, J., Gabora, L., and Sozzo, S. (2013a). Quantum structure and human thought. *Behav. Brain Sci.* 36, 274–276. doi: 10.1017/S0140525X12002841

Aerts, D., Colebunders, E., Van der Voorde, A., and Van Steirteghem, B. (1999). State property systems and closure spaces: a study of categorical equivalence. *Int. J. Theor. Phys.* 38, 359–385. doi: 10.1023/A:1026657913077

Aerts, D., and Czachor, M. (2004). Quantum aspects of semantic analysis and symbolic artificial intelligence. *J. Phys. A Math. Theor.* 37, L123–L132. doi: 10.1088/0305-4470/37/12/L01

Aerts, D., and Daubechies, I. (1978). Physical justification for using the tensor product to describe two quantum systems as one joint system. *Helv. Phys. Acta* 51, 661–675.

Aerts, D., and Gabora, L. (2005a). A theory of concepts and their combinations I. The structure of the sets of contexts and properties. *Kybernetes* 34, 167–191. doi: 10.1108/03684920510575799

Aerts, D., and Gabora, L. (2005b). A theory of concepts and their combinations II. A Hilbert space representation. *Kybernetes* 34, 192–221. doi: 10.1108/03684920510575807

Aerts, D., Gabora, L., and Sozzo, S. (2013b). Concepts and their dynamics: a quantum-theoretic modeling of human thought. *Top. Cogn. Sci.* 5, 737–772. doi: 10.1111/tops.12042

Aerts, D., and Sozzo, S. (2011). Quantum structures in cognition: why and how concepts are entangled. *Lect. Notes Comp. Sci.* 7052, 118–129. doi: 10.1007/978-3-642-24971-6-12

Aerts, D., and Sozzo, S. (2014). *What is Quantum? Unifying its Micro-Physical and Structural Appearance*. Filzbach: Quantum Interaction.

Aerts, D., and Van Steirteghem, B. (2000). Quantum axiomatics and a theorem of M.P. Soler. *Int. J. Theor. Phys.* 39, 497–502. doi: 10.1023/A:1003661015110

Allen, J. F., and Misener, A. D. (1938). Flow of liquid helium II. *Nature* 141:75. doi: 10.1038/141075a0

Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E., and Cornell, E. A. (1995). Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* 269, 198–201. doi: 10.1126/science.269.5221.198

Arndt, M., Nairz, O., Vos-Andreae, J., Keller, C., van der Zouw, G., and Zeilinger, A. (1999). Wave-particle duality of C 60 molecules. *Nature* 401, 680–682. doi: 10.1038/44348

Aspect, A., Grangier, P., and Roger, G. (1981). Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* 47, 460–463. doi: 10.1103/PhysRevLett.47.460

Aspect, A., Grangier, P., and Roger, G. (1982). Experimental realization of Einstein-Podolsky- Rosen-Bohm Gedankenexperiment. A new Violation of Bell's Inequalities. *Phys. Rev. Lett.* 49:91. doi: 10.1103/PhysRevLett.49.91

Aspelmeyer, M., Böhm, H. R., Gyatso, T., Jennewein, T., Kaltenbaek, R., Lindenthal, M., et al. (2003). Long-distance free-space distribution of quantum entanglement. *Science* 301, 621–623. doi: 10.1126/science.1085593

Bardeen, J., Cooper, L. N., and Schrieffer, J. R. (1957). Theory of Superconductivity. *Phys. Rev.* 108, 1175–1205. doi: 10.1103/PhysRev.108.1175

Bell, J. S. (1964). On the einstein-podolsky-rosen paradox. *Physics* 1, 195–200.

Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of hidden variables I&II. *Phys. Rev.* 85, 166–179, 180–193. doi: 10.1103/PhysRev.85.180

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature* 121, 580–590. doi: 10.1038/121580a0

Bruno, N., Martin, A., Sekatski, P., Sangouard, N., Thew, R. T., and Gisin, N. (2013). Displacement of entanglement back and forth between the micro and macro domains. *Nat. Phys.* 9, 545–548. doi: 10.1038/nphys2681

Busemeyer, J. R., and Bruza, P. D. (2012). *Quantum Models of Cognition and Decision*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511997716

Busemeyer, J. R., Dubois, F., Lambert-Mogiliansky, A., and Melucci, M. (2012). *Quantum Interaction. Lecture Notes in Computer Science*, Vol. 7620 (Berlin: Springer). doi: 10.1007/978-3-642-35659-9

Busemeyer, J. R., Pothos, E., Franco, R., and Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment "errors." *Psychol. Rev.* 118, 193–218. doi: 10.1037/a0022542

Bruza, P. D., Lawless, W., van Rijsbergen, C. J., and Sofge, D. (2007). (eds.). *Proceedings of the AAAI Spring Symposium on Quantum Interaction*, March 27–29. Stanford University, Stanford.

Bruza, P. D., Lawless, W., van Rijsbergen, C. J., and Sofge, D. (2008). (Eds.). *Quantum Interaction: Proceedings of the Second Quantum Interaction Symposium*. London: College Publications.

Bruza, P. D., Sofge, D., Lawless, W., Van Rijsbergen, K., and Klusch, M. (2009). *Proceedings of the Third Quantum Interaction Symposium. Lecture Notes in Artificial Intelligence*, Vol. 5494 (Berlin: Springer). doi: 10.1007/978-3-642-00834-4

Cattaneo, C., and Nistico, G. (1991). Axiomatic foundations of quantum physics: critiques and misunderstandings. Piron's question-proposition system. *Int. J. Theor. Phys.* 30, 1293–1336. doi: 10.1007/BF01026176

Cattaneo, C., and Nistico, G. (1993). A model of Piron's preparation-question structures in Ludwig's selection structures. *Int. J. Theor. Phys.* 32, 407–431. doi: 10.1007/BF00673352

Coecke, B., Moore, D., and Wilce, A. (Eds.) (2000). *Current Research in Operational Quantum Logic: Algebras, Categories, Languages*. Dordrecht: Springer. doi: 10.1007/978-94-017-1201-9

Cooper, L. N. (1956). Bound electron pairs in a degenerate Fermi gas. *Phys. Rev.* 104, 1189–1190. doi: 10.1103/PhysRev.104.1189

de Broglie, L. (1923). Ondes et Quanta. *Comptes Rendus*, 177, 507–510.

de Broglie, L. (1928). "La nouvelle dynamique des quanta," in *Proceedings of the Solvay Conference, 1928, Electrons et Photons* (Brussels), 105–132.

Dyson, F. J. and Lenard, A. (1967). Stability of matter. I. *J. Math. Phys.* 8, 423–434. doi: 10.1063/1.1705209

Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik* 322, 132–148. doi: 10.1002/andp.19053220607

Einstein, A. (1917). Zur Quantentheorie der Strahlung. *Phys. Zeitschrift* 18, 121–128.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* 47, 777–780. doi: 10.1103/PhysRev.47.777

Engel, G. S., Calhoun, T. R., Read, E. L., Ahn, T. K., Mancal, T., Cheng, Y. C., et al. (2007). Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature* 446, 782–786. doi: 10.1038/nature05678

Engesser, K., Gabbay, D. M., and Lehmann, D. (Eds.). (2007). *Handbook of Quantum Logic and Quantum Structures: Quantum Structures*. Amsterdam: Elsevier.

Fierz, M. (1939). Über die relativistische Theorie kräftefreier Teilchen mit beliebigem Spin. *Helv. Phys. Acta* 12, 3–37.

Foulis, D. (1999). "A half-century of quantum logic – what have we learned?," in *Quantum Structures and the nature of Reality, The Indigo Book of Einstein Meets Magritte*, eds D. Aerts and J. Pykacz (Dordrecht: Springer).

Gabora, L., and Aerts, D. (2002). Contextualizing concepts using a mathematical generalization of the quantum formalism. *J. Exp. Theor. Artif. Intell.* 14, 327–358. doi: 10.1080/09528130210162253

Gavroglu, K., and Goudaroulis, Y. (1988). Understanding macroscopic quantum phenomena: the history of superfluidity 1941-1955. *Ann. Sci.* 45, 367–385. doi: 10.1080/00033798800200291

Gerlich, S., Eibenberger, S., Tomandl, M., Nimmrichter, S., Hornberger, K., Fagan, P.J., et al. (2011). Quantum interference of large organic molecules. *Nat. Commun.* 2, 263. doi: 10.1038/ncomms1263

Ginzburg, V. L., and Landau, L. D. (1950). On the theory of superconductivity. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 20, 1064.

Hampton, J. A. (1988a). Overextension of conjunctive concepts: evidence for a unitary model for concept typicality and class inclusion. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 12–32. doi: 10.1037/0278-7393.14.1.12

Hampton, J. A. (1988b). Disjunction of natural concepts. *Mem. Cogn.* 16, 579–591. doi: 10.3758/BF03197059

Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift Für Physik* 33, 879–893. doi: 10.1007/BF01328377

Heisenberg, W. (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift Für Physik* 43, 172–198. doi: 10.1007/BF01397280

Herbst, T., Scheidl, T., Fink, M., Handsteiner, J., Wittmann, B., Ursin, R., et al. (2012). Quantum teleportation over 143 kilometres using active feed-forward. *Nature* 489, 269–273. doi: 10.1038/nature11472

Jabs, A. (2010). Connecting spin and statistics in quantum mechanics. *Found. Phys.* 40, 776–792, 793–794. doi: 10.1007/s10701-009-9377-7

Jauch, J. M. (1968). *Foundations of Quantum Mechanics*. Reading: Addison-Wesley Pub. Co.

Josephson, B. D. (1962). Possible new effects in superconductive tunneling. *Phys. Lett.* 1, 251–253. doi: 10.1016/0031-9163(62)91369-0

Kapitza, P. (1938). Viscosity of liquid helium below the l-point. *Nature* 141, 74–75. doi: 10.1038/141074a0

Kasprzak, J, Richard, M., Kundermann, S., Baas, A., Jeambrun, P., Keeling, J. M., et al. (2006). Bose-Einstein condensation of exciton polaritons. *Nature* 443, 409–414. doi: 10.1038/nature05131

Khrennikov, A. Y. (2010). *Ubiquitous Quantum Structure*. Berlin: Springer. doi: 10.1007/978-3-642-05101-2

Lagoudakis, K. G., Wouters, M., Richard, M., Baas, A., Carusotto, I., André, R., et al. (2008). Quantized vortices in an exciton-polariton condensate. *Nat. Phys.* 4, 706–710. doi: 10.1038/nphys1051

Lenard, A., and Dyson, F. J. (1968). Stability of matter. II. *J. Math. Phys.* 9, 698–711. doi: 10.1063/1.1664631

Lieb, E. H. (1976). The stability of matter. *Rev. Mod. Phys.* 48, 553–569. doi: 10.1103/RevModPhys.48.553

Lieb, E. H. (1979). Why matter is stable. *Chin. J. Phys.* 17, 49–62.

London, F. (1938). The λ-phenomenon of liquid Helium and the Bose-Einstein degeneracy. *Nature* 141, 643–644. doi: 10.1038/141643a0

London, F., and London, H. (1935). The electromagnetic equations of the supraconductor. *Proc. R. Soc. Lond. A* 149, 71–88. doi: 10.1098/rspa.1935.0048

Ludwig, G. (1983). *Foundations of Quantum Mechanics I & II*. New York, NY: Springer. doi: 10.1007/978-3-642-86751-4

Mackey, G. W. (1963). *Mathematical Foundations of Quantum Mechanics*. New York, NY: W. A. Benjamin.

Meissner, W., and Ochsenfeld, R. (1933). Ein neuer Effekt bei Eintritt der Supraleitfähigkeit. *Naturwissenschaften* 21, 787–788. doi: 10.1007/BF01504252

Muthaporn, C., and Manoukian, E. B. (2004). Instability of bosonic matter in all dimensions. *Phys. Lett. A* 321, 152–154. doi: 10.1016/j.physleta.2003.12.028

Pauli, W. (1940). The connection between spin and statistics. *Phys. Rev.* 58, 716–722. doi: 10.1103/PhysRev.58.716

Pauli, W. (1950). On the connection between spin and statistics. *Prog. Theor. Phys.* 5, 526–543. doi: 10.1143/ptp/5.4.526

Plumhof, J. D., Stöferle, T., Mai, L., Scherf, U., and Mahrt, R. F. (2014). Room-temperature Bose-Einstein condensation of cavity exciton-polaritons in a polymer. *Nat. Mater.* 13, 247–252. doi: 10.1038/nmat3825

Piron, C. (1964). Axiomatique Quantique. *Helv. Phys. Acta* 37, 439–468.

Piron, C. (1976). *Foundations of Quantum Physics*. Reading, MA: W.A. Benjamin Inc.

Piron, C. (1989). Recent Developments in Quantum Mechanics. *Helv. Phys. Acta* 62, 82–90.

Piron, C. (1990). *Mécanique Quantique: Bases Et Applications.* Lausanne: Presse Polytechnique et Universitaire Romandes.

Planck, M. (1901). Über das Gesetz der Energieverteilung im Normalspektrum. *Ann. Phys.* 309, 553–563. doi: 10.1002/andp.19013090310

Pothos, E. M., and Busemeyer, J. R. (2009). A quantum probability model explanation for violations of 'rational' decision theory. *Proc. R. Soc. B* 276, 2171–2178. doi: 10.1098/rspb.2009.0121

Pothos, E. M., and Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling?. *Behav. Brain Sci.* 36, 255–274. doi: 10.1017/S0140525X12001525

Rauch, H., Zeilinger, A., Badurek, G., Wilfing, A., Bauspiess, W., and Bonse, U. (1975). Verification of coherent spinor rotation of fermions. *Phys. Lett. A* 54, 425–427. doi: 10.1016/0375-9601(75)90798-7

Rauch, H., and Werner, A. (2000). *Neutron Interferometry: Lessons in Experimental Quantum Mechanics*. Oxford: Oxford University Press.

Salart, D., Baas, A., van Houwelingen, J. A. W., Gisin, N., and Zbinden, H. (2008). Spacelike separation in a Bell test assuming gravitationally induced collapses. *Phys. Rev. Lett.* 100:220404. doi: 10.1103/PhysRevLett.100.220404

Sarovar, M., Ishizaki, A., Fleming, G. R., and Whaley, K. B. (2010). Quantum entanglement in photosynthetic light-harvesting complexes. *Nat. Phys.* 6, 462–467. doi: 10.1038/nphys1652

Sassoli de Bianchi, M. (2010). Ephemeral properties and the illusion of microscopic particles. *Found. Sci.* 16, 393–409. doi: 10.1007/s10699-011-9227-x

Sassoli de Bianchi, M. (2011). From permanence to total availability: a quantum conceptual upgrade. *Found. Sci.* 17, 223–244. doi: 10.1007/s10699-011-9233-z

Sassoli de Bianchi, M. (2013). Quantum dice. *Ann. Phys.* 336, 56–75. doi: 10.1016/j.aop.2013.05.018

Scholes, G. D. (2010). Quantum-coherent electronic energy transfer: did nature think of it first? *J. Phys. Chem. Lett.* 1, 2–8. doi: 10.1021/jz900062f

Schrödinger, E. (1926a). Quantizierung als Eigenwertproblem (Erste Mitteilung). *Ann. Phys.* 79, 361–376. doi: 10.1002/andp.19263840404

Schrödinger, E. (1926b). Über das Verhältnis der Heisenberg Born Jordanischen Quantenmechanik zu der meinen. *Ann. Phys.* 79, 734–756. doi: 10.1002/andp.19263840804

Schrödinger, E. (1944). *What is Life?* Cambridge: Cambridge University Press.

Smets, S. (2003). In defense of operational quantum logic. *Logic Logic. Philos.* 11, 191–212. doi: 10.12775/LLP.2003.010

Song, D., Melucci, M., Frommholz, I., Zhang, P., Wang, L., and Arafat, S. (2011). *Quantum Interaction. Lecture Notes in Computer Science*, Vol. 7052. Berlin: Springer. doi: 10.1007/978-3-642-24971-6

Sozzo, S. (2014). A quantum probability model in Fock space for borderline contradictions. *J. Math. Psychol.* 58, 1–12. doi: 10.1016/j.jmp.2013.11.001

Streater, R. F., and Wightman, A. S. (2000). *PCT, Spin and Statistics, and All That*. Princeton, NJ: Princeton University Press.

Tittel, W., Brendel, J., Gisin, B., Herzog, T., Zbinden, H., and Gisin, N. (1998). Experimental demonstration of quantum correlations over more than 10 km. *Phys. Rev. A* 57, 3229–3232. doi: 10.1103/PhysRevA.57.3229

Tisza, L. (1938). Transport Phenomena in Helium II. *Nature* 141, 913. doi: 10.1038/141913a0

Van Rijsbergen, K. (2004). *The Geometry of Information Retrieval*. Cambridge University Press: Cambridge. doi: 10.1017/CBO9780511543333

von Neumann, J. (1932). *Mathematical Foundations of Quantum Mechanics*. Princeton, NJ: Princeton University Press.

Wang, Z., Busemeyer, J.R., Atmanspacher, H., and Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Top. Cogn. Sci.* 5, 672–688. doi: 10.1111/tops.12043

Weihs, G., Jennewein, T., Simon, C., Weinfurter, H., and Zeilinger, A. (1998). Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* 81, 5039–5043. doi: 10.1103/PhysRevLett.81.5039

Widdows, D. (2006). *Geometry and Meaning*. Chicago: University of Chicago Press; CSLI Publications.

Wilczek, F. (2008). *The Lightness of Being: Mass, Ether, and the Unification of Forces*. New York, NY: Basic Books.

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: 05 March 2014; accepted: 19 May 2014; published online: 24 June 2014.
Citation: Aerts D (2014) Quantum theory and human perception of the macro-world. Front. Psychol. ***5****:554. doi: 10.3389/fpsyg.2014.00554*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Aerts. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 29 January 2014
doi: 10.3389/fpsyg.2014.00029

Sameness and the self: philosophical and psychological considerations

Stanley B. Klein*

Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA

Edited by:
Chris Fields, Retired, USA

Reviewed by:
Liliann Manning, Strasbourg University, France
Jason M. Ford, University of Minnesota Duluth, USA

*****Correspondence:***
Stanley B. Klein, Department of Psychological and Brain Sciences, University of California, 551 Ucen Road, Santa Barbara, CA 93106, USA
e-mail: klein@psych.ucsb.edu

In this paper I examine the concept of cross-temporal personal identity (diachronicity). This particular form of identity has vexed theorists for centuries—e.g., how can a person maintain a belief in the sameness of self over time in the face of continual psychological and physical change? I first discuss various forms of the sameness relation and the criteria that justify their application. I then examine philosophical and psychological treatments of personal diachronicity (for example, Locke's psychological connectedness theory; the role of episodic memory) and find each lacking on logical grounds, empirical grounds or both. I conclude that to achieve a successful resolution of the issue of the self as a temporal continuant we need to draw a sharp distinction between the *feeling* of the sameness of one's self and the *evidence* marshaled in support of that feeling.

Keywords: identity of self, memory, personal diachronicity, self, temporal continuity

Many of the constructs we grapple with in the behavioral sciences have a common-sense familiarity that makes them seem both conceptually clear and referentially transparent. However, as often is true of common-sense notions, on careful reflection they are found to be conceptually underspecified and referentially vague (e.g., Russell, 1912/1999).

The deceptively simple notion of "identity" is a case in point. When examined through an analytic lens, it becomes clear that the totality of the topics for which identity is a focal concern—e.g., personal identity, social identity, cultural identity, gender identity, national identity, object identity, numerical identity, occasional identity, contingent identity, indefinite identity, strict identity, loose identity, qualitative identity, multiple identity—is more akin to a complex fractal set (e.g., Mandelbrot, 1983) than a well-formed taxonomy. Prior to engaging questions about identity, therefore, I need to make clear how I will be using the term.

FINDING THE TARGET: WHAT TYPE OF "IDENTITY" ARE WE SEEKING?

In some contexts—primarily philosophical and mathematical—identity takes a quantitatively strict and numerically exhaustive form; it is, and only is, the realization of total property equivalence between X and Y. This is the "numerical identity" of abstract formalism, whose origins in Western thought trace to the writings of Parmenides in the 5th Century BCE (for an illuminating discussion, see Papa-Grimaldi, 2010).

But "identity" can, and often does, mean very different things in the sciences. Physical and social scientists frequently are concerned with more flexible requirements for a numerically imprecise, qualitative form of identity (e.g., specification of a subset of properties necessary and sufficient for an object to be taken as the "same" over time; this sometimes is referred to as "exact similarity," e.g., Garrett, 1998). When questions of identity are asked of things that can take different characteristics at different times, numerical equivalence often gives way to conceptions of identity that admit to degrees and remain applicable in the presence of componential variation.

Many philosophers embrace the challenges that arise when the conditions of identity are relaxed.Numerous "puzzle cases"—amoebic cell division, the gradual replacement of an object's parts, brain transplants, split-brain surgery, body teletransportation, and many more—have received treatment (for reviews see Wiggins, 1980; Parfit, 1984; Brennan, 1988; Noonan, 1989; Oderberg, 1993; Gallios, 1998). But not all are equally accepting.

Hume, for example, felt that by allowing more flexible criteria we inadvertently substitute "similarity" for "identity" (Hume, 1739–1740/1978). Butler (1736/1819) argued that we are wrong to think that an object could gain or lose a part without bringing an end to that object: Any change in an object's constituents would, of logical necessity, bring something new into existence. On these views, we confuse identity with similarity—or, as Butler sees it, the formal notion of strict identity has been substituted by more colloquial notions of approximate or loose identity.

Substitution is not necessarily a problem, however, provided we are clear about what we are doing and our reasons for doing so. In the following section I briefly note some reasons that a change from the strict requirements of numerical identity to a more malleable conception is warranted when questions of identity are tackled by scientists. To avoid confusion, I will adopt the term "sameness" when discussing this type of less-than-perfect identity: sameness allows for less rigid, more qualitative criteria than does the quantitatively exacting demands of numerical identity (which is a type of sameness; see below).

DEGREES OF SAMENESS

In the sciences, objects of interest (whether concrete or abstract) often are held to admit to "identity" despite alterations in their properties and predicates. When entertaining the possibility of identity in the face of change, words such as "exact similarity" and "sameness" seem better suited to convey the type of identity under consideration (e.g., Williams, 1973; Noonan, 1989; Gallios,

1998; Garrett, 1998). Accordingly, rather than "identity"—which implies a binary opposition conditionalized on the presence or absence of complete property equivalence—I will use the term "sameness." This more flexible notion allows for a spectrum of possibilities—ranging from sameness in its strict, numerical form to sameness despite (sometimes considerable) componential variation. It thus is better positioned to capture the diversity of quantitative as well as qualitative interests that characterize the numerically imprecise identities often of interest in the sciences.

In its most analytically rigid form, "sameness" entails a quantitative equivalence between X and Y. This, I suggest, typically is what we have in mind when we consider the term "identity" absent qualifying contextualization (e.g., personal identity, ethnic identity, gender identity, and so forth). Numerical sameness is a property everything has to itself and to nothing else. Formally, it is expressed as "X is the same as Y if and only if every property or characteristic true of X is true of Y as well." Historically, this commonly is referred to as the "identity of indiscernibles," and in modern form traces to the work of Leibnitz (e.g., Williams, 2002). Despite its inherent circularity—i.e., numerical sameness necessarily is true if and only if what is "true of X" is taken to include "being identical with X"—it remains the foundational expression of the concept of quantitative sameness (e.g., Brennan, 1988; Williams, 1990; Oderberg, 1993; Gallios, 1998). Interest in this strict version of sameness is found primarily in philosophical treatment and mathematical analysis, and will not be discussed herein.

Satisfaction of the criteria for numerical sameness seldom is in play when questions are directed toward issues of concern in the sciences[1]. Numerical sameness is an equivalence relation that must be true by virtue of the tautology it entails. Entities satisfying the requirements for numerical sameness would result in a very narrowly circumscribed set (albeit a numerically large one, since everything is quantitatively identical to itself at a given point in time). The interests of social and physical scientists more often are trained on sameness relations of entities that undergo changes wrought by the passage of time. A stone, for example, can endure erosion or supplementation (e.g., by mineral seepage), yet still be judged the same stone; a person can be considered the same person despite alterations in physical characteristics and mental states. In these domains of inquiry, equivalence, construed as numerically exhaustive, has little theoretical or empirical traction.

Accordingly, less restrictive notions are needed to accommodate the type of things toward which questions of sameness can be posed despite property variance (e.g., Brennan, 1988; Oderberg, 1993). Questions of the sameness of objects (e.g., "is that the same car I saw yesterday?"), propositions (e.g., "on closer analysis, the two theories seem to be the same"), mental states (e.g., "I think we have the same idea") and more complex cases (e.g., the self: "Am I the same person I was 10 years earlier?") allow for the possibility that X and Y are, in some sense, the same despite not satisfying the strict requirements for numerical equivalence. An important consequence of this relaxation in criteria is that it draws greater attention to the thing being evaluated, broadening the scope of analysis to include consideration not only of the sameness relation, but also of the nature of the relata placed in relation. This change in accent, as we will see, takes a particular significance when the object of inquiry is one's self (e.g., Shoemaker, 1963; Wiggins, 1971; Rorty, 1976; Hirsch, 1982; Baillie, 1993; Garrett, 1998; Baker, 2000; Lund, 2004; Perry, 2008; Sani, 2008)[2].

SAMENESS AND THE SELF: THE PROBLEM OF PERSONAL DIACHRONICITY

As Hume's and Butler's insights suggest, when addressing questions of identity in the sciences we often loosen the requirements for numerical sameness. Our concern is how people judge or perceive the qualitative sameness of an entity despite changes with time. When the person is taken as the object of a sameness judgment, the requirements of quantitative equivalence would, of definitional necessity, preclude affirmation for any observation falling outside the narrow boarders of instantaneity: the continual change associated with the psycho-physical existence would make personal diachronicity (i.e., the sameness of the person over time) a logical impossibility (unless one subscribed to a view in which change is an illusion, and the reality behind the illusion is in a state of stasis; for discussion see Barbour, 2000 and Papa-Grimaldi, 2010).

Quantitative sameness clearly is *not* what we have in mind when the sameness of persons is in question; rather, we are interested in criteria that can be used to justify a belief that Person X is the same at time T_1 and at time T_2. Under these circumstances, conditions satisfying the tautological certainty of numerical sameness give way to the search for criteria capable of allowing for the possibility personal sameness despite alteration in properties or predicates. As we will see in the section titled Types of Self and Types of Personal Diachronicity: Evidence and Certainty, criterial emendations are particularly complex when the object of a sameness judgment also is the one making the judgment—i.e., the sameness of one's self.

Accordingly, questions of strict, numerical identity are not (and cannot be) the concern theorists interested in personal diachronicity (adoption of such criteria would result in an empty set). Rather, our interest is in the criteria we rely on to attribute spatio-temporal *continuity* to persons in general and the self in particular. And, of logical and empirical necessity, these criteria must entail the flexibility necessary to ascertain sameness despite inevitable transformations in a person's physical and mental constituents.

In short, despite the use of the word "same" to categorize our theoretical and empirical interests in the sameness of self (e.g., personal identity), we are concerned not with sameness in its

[1]This is not to suggest that questions pertaining to less demanding notions of identity are ignored by philosophers. This is far from the case (e.g., Wiggins, 1971; Brennan, 1988; Noonan, 1989; Gallios, 1998). For example, questions such as absolute vs. relative identity, the Ship of Theseus paradox and the identity of a clay sculpture and the unformed lump out of which it was fashioned are some of the less exacting, boundary issues of identity debated by philosophers.

[2]In this paper, I sometimes will use the term "person" in place of the term "self". This philosophically debatable move is one not everyone will be comfortable with (e.g., Locke, 1689–1700/1975; Wilkes, 1988). My (occasional) substitution of terms entails nothing beyond expositional convenience. While I recognize the conceptual issues it raises, there should be little question of the meaning I intend.

strict, Liebnizian sense, but rather with a more qualitative question of personal sameness or diachronicity. Indeed, when taken as numerical equivalence, the question of personal sameness has no meaning (save for the possibility of a Parmenidean vision of reality as static perfection; e.g., Papa-Grimaldi, 2010).

Personal diachronicity (which henceforth will be restricted in application to the "self") is unique among topics amenable to considerations of sameness. In addition to judgments made by the self-as-subject of the self-as-object (see the next section), sameness pertains also to judgments by the self-as-subject of the self-as-subject. These self-reflexive[3] acts are limited to sentient beings and likely apply with reasonable assurance only to homo sapiens (e.g., Snodgrass and Thompson, 1997; Terrace and Metcalfe, 2005). However, before pursuing the conditions that must be satisfied to justify a judgment of personal diachronicity, we need to make explicit what it is we take to be the target of the sameness judgment—the self.

THE PROBLEM OF THE SELF

As those who study the self-have discovered, answers to the question "What *is* the self?" are elusive at best (for reviews see Johnstone, 1970; Gergen, 1971; Lewis, 1982; Vierkant, 2003; Klein, 2012). Indeed, some are of the opinion that the question is based on the illusion that there is an elusive self to be found (e.g., Albahari, 2006; Metzinger, 2009; for discussion see Siderits et al., 2011). Of course, a problem with this perspective is that an illusion is an experience and an experience requires an experiencer (e.g., Strawson, 2011a; Klein, 2014a). As Meixner (2008) observes, "The fictionalization of subjects of experience is incoherent, since it involves the incoherent idea that I, for example, am an illusion of myself" (p. 162). Kant (1998) goes further, arguing that the self of subjective awareness (his transcendental ego) must accompany experience" [related views can be found in James (1890), Lund (2005)].

Despite ontological concerns, psychology has found work for the "self" in an abundance of subject-hyphen-predicate relations (e.g., self-comparison, self-concept, self-esteem, self-handicapping, self-image, self-perception, self-regulation, self-reference, etc.). However, the focus of investigation rests firmly on the predicate, to the detriment of an appreciation of what exactly is the object of this diverse set of predicates—i.e., the self being verified, conceptualized, esteemed, deceived, verified, regulated, and handicapped (for review see Klein, 2012, 2014a).

This is not to say that psychology has failed to propose models of the self: formalizations have been on display for more than 100 years [e.g., James, 1890; Greenwald, 1981; Neisser, 1988; Kihlstrom and Klein, 1994; Conway, 2005; for recent reviews see Leary and Tangney (2012) and Sedikides and Spencer (2007)]. Yet, most of these offerings target the self in a particular context, rather than the self *per se*. We thus find models of cultural selves, social selves, cognitive selves, synaptic selves, autobiographical selves, social selves, narrative selves, etc. (cf., Leary and Tangney, 2003, 2012). But consideration of what the self *is* that serves as the bedrock of these cultural, social, cognitive, synaptic, and narrative instantiations, typically is under-specified (e.g., Klein and Gangi, 2010; Klein, 2014a).

THE TWO SELVES: THE NEURAL SELF OF SCIENCE AND THE SUBJECTIVE SELF OF FIRST-PERSON PHENOMENOLOGY

One reason for the difficulties we face when attempting to describe what we mean by the word "self" is that there is not a single self to be described (e.g., Stern, 1985; Neisser, 1988; Klein, 2001, 2004, 2012, 2014a; Legrand and Ruby, 2009). Rather, two distinct (but normally interacting) aspects of the self are conjoined in almost every discussion of the topic, although these aspects seldom are separated. As reviewed at length in Klein (2012, 2014a), the self meaningfully can be partitioned into the neurally instantiated systems of self-knowledge and the self of first-person subjectivity (e.g., James, 1890; Zahavi, 2005; Legrand and Ruby, 2009; Strawson, 2009; Klein, 2012, 2014a).

It is beyond the scope of this paper to go into detail about the material and subjective aspects of self [extensive discussion can be found in Klein (2012, 2014a)][4]. Briefly, they cannot be deduced from, or reduced to, a single, underlying principle, structure, process, substance or system (e.g., Kant, 1998; Zahavi, 2005; Klein, 2012, 2014a). One—the neuro-cognitive systems of the psychophysical self (consisting of such things as personal memory, body image, emotions)—is materially (primarily, but not exclusively, neural) instantiated and therefore capable of being apprehended and treated as an *object* of scientific inquiry.

The other—the self of first-person subjectivity—is the *subject* having the experience, rather than the *object* of that experience. This aspect of self cannot be directly known by acts of perception or introspection (e.g., Earle, 1972; Kant, 1998; Zahavi, 2003, 2005; Lund, 2005; Klein, 2012; Swinburne, 2013). Rather, our appreciation of the self of first-person subjectivity is a matter of acquaintance or feeling, something that cannot (easily) be conveyed via descriptive analysis (e.g., Nagel, 1974; Kant, 1998; Zahavi, 2005; Klein, 2012, 2014a).

Despite differences in their epistemological (and possibly ontological; e.g., Klein, 2014a) status, under normal circumstances these two aspects of self-interact, and this interaction is a prerequisite for our experience of self. Indeed, it is *only* via their interaction that a particular form of consciousness—self-awareness—becomes possible [these assertions are treated extensively in Klein (2012, 2014a); see also Gallagher and Zahavi (2008)]. In this regard, I follow Fitche's dictum (e.g., Neuhouser, 1990) that there can be no subject without an object or object without a subject.

Considerable progress has been made describing the cognitive and neurological bases of the material aspects of self (recent treatments can be found in Conway, 2005; Klein and Gangi, 2010;

[3]Questions pertaining to how a subject takes itself qua subject, to be one and the same, open the door to complex issues of self-reflexivity and the philosophical puzzles they engender (e.g., Falk, 1995; Bolander et al., 2006; Strawson, 2009). Their treatment is beyond the scope of this paper.

[4]In Klein (2012, 2014a) I use the terms "epistemological self" and "ontological self" to describe the "material self" and the "self of first-person subjectivity," respectively. My reasons for this unconventional usage are complex (Klein, 2014a) and need not concern us here. For the clarity that comes with conceptual familiarity, the latter terminology (i.e., the material and subjective aspects of self) are adopted in the present text.

Klein and Lax, 2010; Renoult et al., 2012; Martinelli et al., 2013; Prebble et al., 2013). This is because the material, neuro-cognitive bases of self-knowledge can be (and have been) objectified, and thus amenable to scientific analysis.

The subjective aspect of self, by contrast, is too poorly understood to bear the definitional weight required when placed in relation to predicates (e.g., regulation, image, complexity, handicapping, verification, etc.) or contexts (e.g., synaptic, cultural, narrative, etc.). Moreover, as discussed below, treating the subjective self as an object has the unfortunate consequence of stripping it of its core feature—its subjectivity (for discussions see Zahavi, 2005; Ganeri, 2012; Klein, 2012, 2014a).

Researchers often fail to appreciate that the self of first-person subjectivity is *not* the object of their experimental inquiries (e.g., Klein, 2012; Klein and Nelson, 2014). Nor could it be. Objectivity is based on the assumption that an event or object exists independent of any individual's awareness of it (e.g., Earle, 1955; Nagel, 1974; Rescher, 1997; Martin, 2008); it is something other than self. When objectivity is the stance adopted by the self to study itself, the self must, of logical necessity, be directed toward what is not self—i.e., to some "other" that serves as the self's object (e.g., Husserl, 1964; Earle, 1972; Lund, 2005; Zahavi, 2005; Klein, 2014a). Thus, to study myself as an object, I must transform myself into an "other," that is, into a "not-self,"

Accordingly, the subjective self is not, and cannot, be an object for itself and still maintain its subjectivity. Considered by first-person subjectivity, the subjective aspect of self becomes an object in the manner all objects (both mental and physical) must, of necessity, become when apprehended (e.g., Husserl, 1964; Zahavi, 2003; Klein, 2012). In the process, the subjective aspect of the self of first-person experience is lost from view. Paradoxically, the subjective aspect of self can achieve objectivity only at the cost of forfeiting its essence as a subjective center (e.g., Kant, 1998; Zahavi, 2005; Klein, 2012, 2014a).

TYPES OF SELF AND TYPES OF PERSONAL DIACHRONICITY: EVIDENCE AND FEELING

Personal diachronicity concerns our belief that we have an identity that originated in our past and will follow us into our future. Although most treatments take this to be a question of how we *know* (I am using "knowledge" in its non-technical, colloquial sense, rather than its philosophical sense as true, justified, belief) that we are the same over time, a second, equally important aspect of diachronicity often is overlooked—i.e., on what do we base our *feeling* that we are continuous in both temporal directions from the present?

Different criteria come into play depending on whether the self-posses "itself to itself" as an object or as a subject. When treated as the object of subjectivity, criteria that enable us to know that we are the same despite componential change are relevant. I refer to these knowledge-based criteria as *evidential* sameness.

When the self as subject takes its own subjectivity as the basis for sameness, by contrast, the criteria for sameness are felt. In contradistinction to evidential criteria—i.e., a consideration of facts relevant to a diachronicity judgment and the inferences such considerations permit—one's *feeling* of sameness derives from one's pre-reflective feeling that despite change in the object of awareness, the subjective "I" by which the object is apprehended remains unchanged (the potentially ageless nature of the subjective self is addressed in the section titled The Timelessness of the Subjective Self). The feeling of the sameness of the subjective self is a-theoretic—it is feeling devoid of reason and directly apprehended (for discussion, see Earle, 1955; Kant, 1998; Zahavi, 2005; Gallagher and Zahavi, 2008; Klein, 2012, 2014a).

Thus, when questions of sameness are addressed to the self, the answers we seek depend in important ways on the aspect of self being judged. To anticipate my conclusions, logically viable and empirically justifiable arguments for the continuity of the material aspects of self—both its physical properties and psychological features clearly tied by experimental evidence to neural activity (e.g., memory, perception, and so on) are hard to come by: there simply are no unambiguous *evidential* criteria (at least, given the resources currently on hand) capable of underwriting a belief in the diachronicity of the material aspects of self.

In contrast, one's *sense* of personal diachronicity is sustainable when the aspect of self under consideration is its subjectivity. But this comes at a cost—our criteria for sameness, being felt rather than known, are not amenable (in any obvious way) to quantitative, evidential analysis (for discussion, see the section titled The Problem of the Self). And this may seem too high a price to those entrenched in a materialist world view (for discussion, see Papa-Grimaldi, 1998; Meixner, 2005; Koons and Bealer, 2010; Nagel, 2012; Klein, 2014a).

PHILOSOPHICAL TREATMENTS OF PERSONAL DIACHRONICITY: EVIDENTIAL SAMENESS AND THE MATERIAL SELF

Let's begin by examining the evidential criteria commonly used to address the sameness of the material self. Questions of the sameness of the material aspects of self can ordered roughly with regard to their scope or inclusiveness. At the most general level, questions of the sameness can be posed to the self qua physical body: What is the relation between bodily continuity and personal diachronicity?

While Bodily Criteria have been subject to extensive philosophical analysis and debate (e.g., Williams, 1973; Parfit, 1984; Olson, 1997, 2007; Baker, 2000), on examination it becomes apparent that not all parts of the body carry equal evidential weight. One organ—the brain—seems particularly germane to evidence-based treatments of personal diachronicity.

However, even at this more nuanced level, cracks in the criterial base begin to appear. Ultimately, we find that the criteria for continuity of the material aspects of self, if they are to have any possibility of evidential warrant, must focus on its psychological, rather than its physical properties (e.g., Parfit, 1984). I refer to these as Informational Criteria. One psychological property in particular—the continuity of personal memory—traditionally has been taken by psychologists and philosophers alike as the *most likely* informational candidate for grounding judgments of personal diachronicity in an evidential nexus (for reviews see Perry, 2008; Sani, 2008). In this paper, I restrict analysis largely to this aspect of the material self; I only briefly mention a few of the less well-studied informational candidates - e.g., empathic access).

THE BODILY CRITERION

The most general level at which evidence for personal diachronicity might be found comes from analysis of the conditions required to bestow spatio-temporal continuity on the body. As is the case for all physical objects, rapidity of change plays a critical role for judgments of diachronicity (e.g., Spencer Brown, 1957; Campbell, 2004). This is made salient by consideration of Plutach's famous paradox of the "Ship of Theseus" (for discussions see Wiggins, 1980; Brennan, 1988; Noonan, 1989; Oderberg, 1993). In one of its several adaptations (the one most relevant to personal diachronicity), the question posed is whether a ship which has had some or all its planks replaced remains the same ship?

Variations in the replacement schedule play a critical role in the answers one is likely to intuit (e.g., Campbell, 2004). Gradual replacement of the ship's planks (e.g., one at a time, at a leisurely pace), generally support the inference that the ship remains the same. If change is too rapid, however, one's certainty of the ship's continuity is challenged. Yet it is important to bear in mind that all that varies between scenarios is the rapidity of change, not change itself. Differences in sameness judgments appear to trade, to some degree, on temporal considerations.

Judgments of the sameness of objects also pivot on amount of change. Most of us are willing to grant sameness to a ship that has one, or a few, planks replaced. But judgment is less secure when the ship undergoes substantial (or complete) physical alteration, even when the change is gradual. Some have proposed quantitative boundaries beyond which confidence in sameness drops precipitously (for example, Parfit, 1984, suggests componential replacement exceeding 50% has serious negative consequences for sameness judgments). But these numerical constraints are based more on reasonable intuition than on logical analysis or experimental demonstration.

When the sameness of the material self is called into question, a similar set of issues arise. We constantly are adding to and subtracting from our body—e.g., as we age we grow taller, gain and lose pounds, change cells, molecules and atoms. The degree of bodily change can be extraordinary: By some accounts all the atoms in our body are replaced over a 10 year span. The mental properties of the material self-change as well—e.g., we gain and lose knowledge, add and lose memories, acquire new skills, modify goals, and so on.

If change (whether physical or mental) happens slowly, most of us assume we are the same person today we were 1 min, 1 h, or one decade earlier (e.g., James, 1890; Hirsch, 1982; Brennan, 1988; Campbell, 2004)[5]. But is this belief justified (e.g., Wiggins, 1971, 1980; Oderberg, 1993)? If, as per impossible, the "me" of age 60 were to meet the "me" of age 10, most of the evidential bases for spatio-temporal continuity clearly would be lost. The old "me" would bear neither a physical resemblance to the young me, nor would we share many experiences, beliefs, goals, memories and other mental features. In short, these temporally separated, gradually altered selves would have little in common—save a largely intact genetic code. Should we meet, we likely would meet as strangers (although the older "me" might "know better"). In what would our sameness consist?

THE BRAIN CRITERION

As many philosophers have observed, not all aspects of the body are equally positioned to underwrite personal diachronicity (e.g., Shoemaker, 1963; Williams, 1970; Wiggins, 1971; Noonan, 1983; Baker, 2000; Olson, 2007). One part of the body in particular—the brain—seems disproportionately relevant to questions of the sameness of the material self. Perhaps by focusing on a more restricted range of bodily parts, some of the problems associated with the Bodily Criterion can be avoided.

The role of the brain in determination of personal diachronicity is placed in sharp relief by a thought experiment, popularized by Shoemaker (1963) and subsequently elaborated on by Parfit (1984). In the original scenario, Mr. Brown has his brain transplanted into Mr. Robinson's body. Let's call the resulting individual—consisting in Robinson's body and Brown's brain—Mr. Brownson. Assuming the operation was successful, "what is the identity of Mr. Brownson?"

When philosophers (and non-philosophers alike) are asked to reflect on this scenario, the common intuition is that Brownson is the same as the original Mr. Brown (e.g., Noonan, 1989). This suggests that a broad Bodily Criterion must give way more circumscribed view in which certain body parts count more than do others in determinations of personal sameness. What appears required for continuity of the self is not the body, taken en toto, but rather one of its parts—the brain.

THE INFORMATIONAL CRITERION

However, even the Brian Criterion may be too gross a characterization of what matters for personal diachronicity (e.g., Proust, 2003). The brain, after all, simply is the part of the body that happens to host memory, personality, mood, thought and a number of other psychological faculties and functions. Perhaps body-based criteria for the re-identification of the material self, even those restricted to the brain, are not the best place to search for evidentially-based criteria for self-continuity.

The argument can (and has been) made that what serves as the criterion of sameness is not the persistence of the physical brain, but rather the continuity of the personally-relevant *information* contained within that body part. Although this information contingently is located in the brain, the continuity of the information, not of the organ in which it is housed, is what really matters (e.g., Williams, 1973; Parfit, 1984; Brennan, 1988; Noonan, 1989; Gallios, 1998).

Consider, as an example, the case of information transfer popularized by Parfit (1984). Imagine there is a machine capable of extracting all the information in Person X's brain and transferring it to the brain of Person Y, and vice versa. Under this "science-fiction" scenario, who would be Person X and who would be Person Y? As Williams (1970) and many others (e.g., Wiggins, 1971; Noonan, 1989; Baillie, 1993; Garrett, 1998) see it, the answer is clear—where knowledge goes identity follows.

[5]Rapid and substantial change, by contrast, can lead to serious doubts about personal continuity. The classic case of Phineas Gage, who suffered profound changes in personality closely following brain injury, led to the well-known observation by his attending physician that Gage was "no longer Gage." (e.g., O'Driscoll and Leach, 1998).

The most famous version of the Information Criterion is contained in a passage from Locke: "Personal identity—that is, the sameness of a rational being—consists in consciousness alone, and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person." (Locke, 1689 Bk. II, Ch. 27, Sec. 9). Although, as we will see in the section titled The Need to Take Seriously The Self of First-Person Subjectivity in Accounts of Personal Diachronicity, there is some question about exactly what Locke had in mind here (e.g., Strawson, 2011b), the passage usually is taken to involve a person remembering self-referential action or thought (e.g., Shoemaker and Swinburne, 1984; Noonan, 1989), that is what cognitive psychologists call episodic memory (e.g., Tulving, 1983).

Building on this reading, a prominent interpretation of Locke's view goes as follows: A person at one time, P_2 at T_2, is the same person at an earlier time, P_1 at T_1, if and only if P_2 can remember having done and experienced various things performed by P_1 (e.g., Shoemaker, 1963; Greenwood, 1967; Noonan, 1989; Schechtman, 1990; Proust, 2003). Thus, it is the transitivity of episodic memory that establishes the continuity of self.

Similar views are common in psychology (for reviews see Fivush and Haden, 2003; Sani, 2008). An especially clear exposition is offered by two prominent neuroscientists: "We are not who we are simply because we think. We are who we are because we can remember what we have thought about.... Memory is the glue that binds our mental life, the scaffolding that holds our personal history and that makes it possible to grow and change throughout life. When memory is lost, as in Alzheimer's disease, we lose the ability to recreate our past, and as a result, we lose our connection with ourselves and with others." (Squire and Kandel, 1999, p. ix). On this analysis, what makes a person the same across time are relations of memory: it is by memory of past action that the self attains a sense of continuity.

Because Locke's memory-based account (and by memory he typically is taken to mean episodic memory) has received the bulk of attention from philosophers and psychologists, I focus on this aspect of the informational criterion in what follows. However, the reader should be made aware that this is not the only candidate for an informational criterion capable of supporting our belief in personal diachronicity (I briefly mention a few others, though my treatment rests firmly on the evidential offerings of memory).

Unfortunately, as his critics were quick to note, Locke's account seems to entail a vicious form of circularity (Butler, 1736/1819; Reid, 1813/1969). For a mental state to count as my memory of a past action, it has to be the case that I was the one who performed the past action. If it wasn't me who performed the action, then my apparent recollection is simply a mistake, not a memory. Butler states the problem bluntly: "one should really think it self-evident, consciousness of personal identity presupposes, and therefore cannot constitute, personal identity" (p. 290). If memory presupposes sameness of self, then trying to give an account of identity in terms of memory seems hopeless.

Although the circularity objection is a serious problem for any simple version of Lockean theory (e.g., Williams, 1973; Brennan, 1985; Noonan, 1989; Proust, 2003), many still favor a memory-based account of personal diachronicity (as opposed to, say, a bodily account; see Olson, 2007). Accordingly, a number of emendations have been proposed to rein in the tautology (e.g., Schechtman, 1990; Hamilton, 1995; Collins, 1997; Slors, 2001; Klein and Nichols, 2012; for review, see Bernecker, 2010).

I have discussed the circularity objection at length elsewhere, and presented evidence that episodic recollection and the self are contingently, not logically, intertwined (Klein and Nichols, 2012). Treatment of these issues would take us far beyond the scope of the present paper. Instead, I focus on the other well-known criticism of Locke's memory criterion—i.e., that it cannot work due to "gaps" that necessarily occur in our memorial record. This issue of episodic transitivity has exercised theorists from the earliest days of the debate.

Hume (1739–1740/1978), conceptualizing the problem in terms of numerical sameness, asks how could there be a quantitatively strict sameness across time in light of the fact that a person's psychology constantly is changing? Reid (1813/1969) also takes issue with Locke's memory criteria, arguing that even less numerically exacting accounts present seemingly insurmountable difficulties (although he famously rejects the memory theory of personal identity, Reid does acknowledge that memory seems to provide "irresistible" evidence that I am the very person who did the action; 1813/1969).

Suppose, Reid observes, a military officer had been flogged for robbing an orchard when he was a boy at school, had bravely vanquished an enemy during battle, and had been made a general later in life. Further, suppose that when he won his military campaign, he could remember having been flogged at school and that when made a general he was remembered his military victory but no longer remembered his flogging.

As Reid sees it, if a person at time t_n remembers an event that occurred at time t_1, then the person at time t_n is identical with the person who was witness to or the agent responsible for the event at time t_1. Thus, if the brave officer who defeated the enemy remembers being beaten at school, then the officer is identical with the boy who was beaten. By similar logic, if the general remembers defeating the enemy in battle, then the general is identical with the brave officer. If the general is identical with the brave officer, and the officer is identical with the boy, then, by the logic of transitivity, the general is identical with the boy.

However, since the sameness of memory is a necessary condition for sameness of self, if a person at time t_n does not remember an event that occurred at time t_1, then the person at time t_n cannot be the same as any person who was witness to or agent of the event at time t_1. Thus, if the general cannot remember being beaten at school, he cannot be the same as the boy who was beaten. Locke's memory account thus suffers from a set of mutually incompatible theses—i.e., the general is both the same as and different from the boy.

Williams (1973) has identified another obstacle facing an evidential account of personal sameness based on memory criteria. He invites us to imagine a situation in which the memory claims of Person X are continuous with those of deceased Person Y. That is, Person X's memory claims map unanimously with the life-history of Person Y. Does this mean that Person X is Person Y? And if so, does this mean that a person can be alive and dead at the same time? It is clear that memory-based evidence (and episodic

recollection in particular) suffers from problems that render the utility of the Informational Criterion less than optimal.

Some have attempted to circumvent these problems by proposing that information other than memory might provide the evidential basis for judgments of personal diachroncity. Schechtman (2001), for example, suggests we shift emphasis from an exclusive reliance on memorial criteria to what she calls "empathic access"—i.e., one's psychological make-up, broadly construed to include desires, feelings, goals, values, beliefs, memory, etc. (Schechtman is not alone in this regard, though others do not adopt her terminology of "empathic access"). Others have argued that relaxing the requirement of immediate access to a temporally continuous succession of remembered events might avoid the problem of "gapy" memorial records (e.g., Brennan, 1985). On this account, it is sufficient that we show enough coherence in our recollections to merit the assignment of sameness to a person.

But, with regard to the former approach, potential gaps and issues of transitivity still remain in play even when mental states other than those strictly taken as memory are recruited as evidential criteria (for discussion, see Klein, 2014b). And the relaxation argument is shown to be inadequate in light of circumstances in which individuals maintain a sense of personal continuity despite the *complete* loss of episodic memory (as we will see in the next section). In addition, it is unclear just what constitutes "enough" coherence.

PSYCHOLOGICAL TREATMENTS OF PERSONAL DIACHRONICITY: EVIDENTIAL SAMENESS AND THE MATERIAL SELF

Most philosophical treatments of personal diachronicity, as we have seen, rely on "thought experiments" to identify the evidential bases of sameness judgments. Arguments resulting from this "mental empiricism" are believed viable if they can be shown to be internally consistent and logically coherent.

Recently, however, philosophers have begun to question the utility of thought experiments unconstrained by scientific empiricism (e.g., Wilkes, 1988; Focquaert, 2003). Coherence and consistency may allow us to judge the logical warrant of a criterion, but conceivability should not be confused with empirical possibility. Perhaps if logical considerations were supplemented with empirical evidence, there still might be hope for a memory-based approach to personal sameness.

Psychologists apparently think so: Many accept (often uncritically) the idea that memory—in particular, its episodic component —is the basis of personal diachronicity (e.g., Rubin, 1986; Conway, 2005; Markowitsch and Staniliou, 2011; Bluck and Liao, 2013; for reviews see Fivush and Haden, 2003 and Sani, 2008). Neurological case studies appear especially suited to shedding light on this issue (e.g., Rathbone et al., 2009; Illman et al., 2011; Duval et al., 2012; Picard et al., 2013; Klein, 2014b). Specifically, cases of neurological impairment offer the possibility of observing dissociations between a belief in one's temporal continuity and the neurological mechanisms posited to support that belief. In this way, one can examine the extent to which belief in the sameness of self contingently depends on the availability of neurally instantiated informational criteria.

When examined critically, however, the evidence is not encouraging. As I show below, episodic memory cannot, by itself, do the work needed to underpin one's belief in one's sameness over time. While recollection may be *useful* in response to personally or socially motivated requests for evidential support, case studies have shown that episodic memory can be lost (even completely) without any obvious consequences for one's sense of diachronicity (for reviews see Klein and Gangi, 2010; Craver, 2012; Klein, 2012, 2014b). In short, empirical evidence (as well as logical considerations; e.g., the issues of non-transitivity identified by Reid) make clear that while episodic memory may be sufficient for one's sense of personal continuity, it is not necessary[6].

SEMANTIC MEMORY AND PERSONAL DIACHRONICITY

Before abandoning a memorial criterion, however, it is important to keep in mind that the self is represented in systems other than episodic memory (for reviews, see Klein, 2004; Gillihan and Farah, 2005; Klein and Gangi, 2010; Klein and Lax, 2010; Renoult et al., 2012; Martinelli et al., 2013). Within semantic memory, for example, there are (at least) two different subsystems devoted to autobiographical knowledge (for review and discussion, see Klein and Lax, 2010). One contains factual self -knowledge (e.g., "I am 61" and "I live in Goleta"). The other is the repository of knowledge of one's personality traits (e.g., "I am intelligent" and "I am not punctual").

There now exists an extensive data-base showing that patients suffering episodic amnesia still can retain access personal facts and trait characteristics (for evidence and reviews see Tulving et al., 1988; Tulving, 1993; Klein et al., 1996; Rathbone et al., 2009; Klein and Gangi, 2010; Klein and Lax, 2010; Martinelli et al., 2013). It is possible, some have suggested, that one's sense of personal identity can be maintained by semantic forms of self-knowledge (factual and trait) in the presence of episodic amnesia. Consistent with this position, evidence suggests that one's sense of personal sameness is not lost despite (sometimes pervasive) episodic memory impairment (e.g., Rathbone et al., 2009; Haslam et al., 2010; Illman et al., 2011; Duval et al., 2012; Klein, 2014b).

A DISSOCIATION BETWEEN FACTUAL SELF-KNOWLEDGE AND TRAIT SELF-KNOWLEDGE

These cases and others like them [reviewed in Klein and Lax (2010)] demonstrate a dissociation between episodic and semantic forms of self-knowledge[7]. But can semantic knowledge of one's traits dissociate from other types of semantic knowledge (both self- and non-self-referential)? Further testing suggests that it can.

[6]To argue that episodic memory is neither necessary nor sufficient, one would need to produce a case in which a person has episodic memory but no semantic memory, and that under these circumstances a sense of personal diachronicity was absent. Such a case, however, is not found in the annals of neuroscience (and I am not sure that a situation in which a person has intact episodic memory accompanied by complete absence of semantic memory is—on definitional, linguistic or phylogenetic grounds—possible).

[7]The relation between semantic trait self-knowledge and episodic recollections of trait-relevant behavior is a complicated affair. Suffice it to say that a substantial body of research shows that the respective roles of these two systems of memory in the creation of trait self-knowledge depend on a large number of factors (for a recent review see Klein et al., 2008).

Consider the case of Patient D.B., a 79-year old man who became profoundly amnesic as a result of anoxia following cardiac arrest. One particularly noxious consequence of his anoxia was that it rendered him incapable of episodically recollecting a *single* thing he ever had done or experienced.

To test his semantic trait self-knowledge, we asked D.B. on two separate occasions to judge a list of personality traits for self-descriptiveness. We also asked his 49-year-old daughter (with whom he lives) to rate him on the same traits. Our findings revealed that D.B.'s trait ratings were both reliable and consistent with the way he is perceived by others (for analyses and discussion see Klein et al., 2002c). Moreover, his access to trait self-knowledge was indistinguishable from age-matched, neurological healthy controls. He thus maintained accurate and reliable knowledge of his personality despite lacking access to specific actions and experiences on which that trait knowledge was based. A similar picture is presented by patient K.C. Tulving (1993). Despite suffering a complete loss of episodic memory, K.C.'s ability to access trait self-knowledge remained intact.

Although D.B. knew which traits described him, he had considerable difficulty accessing semantic-based factual self-knowledge. For example, he no longer could recall the names of any friends from his childhood or even the year of his birth. He also showed spotty knowledge of facts in the public domain. For instance, although he was able to accurately recount a number of details about certain historical events, his knowledge of other historical facts was seriously compromised (e.g., he claimed that America was discovered by the British in 1812).

Taken together, these findings evidence dissociations *within* semantic memory. On the one hand, D.B.'s general semantic knowledge and factual self-knowledge was impaired; on the other hand, his semantic trait self-knowledge was spared and, at least with respect to the measures used, indistinguishable from that of control participants.

Moreover, his ability to retrieve trait self-knowledge was not due simply to the sparing of the systems responsible for maintaining a data-base of trait knowledge (whether about self or other). For example, D.B. was unable to produce accurate knowledge of his daughter's traits (e.g., Klein et al., 2002c). Similar selectivity favoring trait self-knowledge also has been found to characterize autistic memory function (e.g., Klein et al., 1999, 2004).

These findings suggest that the resilience of trait self-knowledge is not a general property of semantic trait-knowledge. Rather, it appears specific to trait generalizations about the self. Indeed, my colleagues and I have yet to find a population (e.g., amnesia, autism, ADHD, Alzheimer's Dementia, Prosopagnosia, Schizophrenia) that cannot reliably and accurately report knowledge of their own traits despite (often considerable) disruption of other neurological and cognitive function (for reviews see Klein and Lax, 2010; Klein et al., 2013).

In contrast to the conclusions just voiced, work reported in a volume edited by Prigatano and Schacter (1991) suggests that people suffering deficits following neural injury sometimes do not recognize the extent to which particular trait-based characterizations apply to them. In addition, evidence is presented that patient and family members may give different answers to questions about traits that describe the patient.

However, as discussed at some length in Klein et al. (2013), the question of the stability of one's beliefs about his or her personality traits does not trade on agreement between one's views and those of others. People—whether brain damaged or fully intact—often disagree with others about which traits best describe them (e.g., Klein et al., 2002a). The question is whether a person's beliefs about his or her dispositions remains stable (even if at odds with the beliefs of others) over time, not the assumed accuracy of those beliefs. And with respect to the former concern, the evidence is that our beliefs about our dispositions remain remarkably stable even in the presence of considerable neurological damage and cognitive chaos.

Returning to the question of personal diachronicity, a review of the evidence suggests that that individuals suffering loss of both episodic *and* factual semantic knowledge still have a sense of temporal self-extension (Klein, 2012, 2014b). Perhaps, then, the remarkable stability of semantic trait self-knowledge provides the bedrock from which one's sense of personal diachronicity springs.

In summary, with respect to the Evidential Criterion, long-term memory does not seem necessary for one's feeling of personal identity across time (for a similar conclusion, see Craver, 2012). The fact that patients like D.B. lack access to episodic memory and show impairments of factual semantic personal memory yet maintain a sense of personal diachronicity (possibly influenced to some degree and in some, as yet, unspecified manner by the stability of semantic trait self-knowledge) suggests one does not need either episodic memory or factual semantic self-knowledge to experience the sameness of self [a similar conclusion, based on philosophical considerations, is found in Strawson (2005)].

PERSONAL DIACHRONICITY AND THE SUBJECTIVE SELF

"If we would have true knowledge of anything, we must quit the body."

(Phaedo, quoted in Russell, 1949, p. 159).

Thus, far I have examined some of the evidential criteria by which we might make judgments of personal sameness over time. These criteria apply in their most straight-forward manner to those aspects of the self that fall under the heading "material"—i.e., the psycho-physical features of self-amenable to objectification. Unfortunately, as we have seen, with the possible exception of trait self-knowledge, the utility of this evidence for underwriting our sense of personal continuity is at best questionable.

There is, however, another aspect of self—its first-person subjectivity—that has received little attention as a possible basis of personal diachronicity. Is there any reason to suspect this aspect of self may serve as the foundation of our feeling of diachronicity? I believe there is, and my reasons for so believing, as well as the empiricism on which they are based, are the focus of the next several sections of this paper.

THE NEED TO TAKE SERIOUSLY THE SELF OF FIRST-PERSON SUBJECTIVITY IN ACCOUNTS OF PERSONAL DIACHRONICITY

First-person subjectivity is a universal aspect of our experience of self; one that, despite well-known difficulties situating it in a materialist framework (e.g., Klein, 2014a), is a phenomenological reality that cannot be ignored if one is to fully appreciate what

it means to be a self (e.g., James, 1890; Kant, 1998; Lund, 2005; Zahavi, 2005; Dainton, 2008; Legrand and Ruby, 2009; Strawson, 2009; Klein, 2012, 2014a). Equations and measurements can be useful when they are related to experience; but experience comes first (e.g., Gallagher and Zahavi, 2008; Klein, 2014a).

It is undeniable that many, if not all, of the great achievements in modern science were made possible by the exclusion of "subjectivity" from the world around us. However, a comprehensive appreciation of reality must include that aspect of reality that makes its understanding possible—i.e., the subjectivity of self that provides us with the ability to be aware of the world of which it is a part. To do otherwise is to exclude by stipulation that aspect of nature that makes nature knowable to itself. As Ricard and Thuan (2001) observe, "If we define the terrain field of science as what can be physically studied, measured, and calculated, then right from the start we leave out everything that is experienced in the first person, and all immaterial phenomena. If we forget this limitation, then we soon start affirming that the universe is everything that can be objectified in the third person, and only what is material." (p. 241).

It thus seems prudent to consider the possibility that our sense of self-sameness derives from feelings obtained from and apprehended by the conscious aspect of self. In this regard, it is interesting to note that Strawson (2011b) has made a strong case for taking Locke at his word—to wit, when Locke posits the continuity of consciousness as the foundation of diachronic personal identity, he means just that: Continuity derives from the felt invariance of subjectivity (i.e., the subjective aspect of self), *not* from evidential (e.g., memorial) sources which subjectivity takes as its objects (which are in a continual state of change). On awakening each morning, I immediately am aware of my self, that "I" exist. My feeling of self as a psychological continuant is not something I need to deduce or reconstruct to justify my feeling of continuity [8]. As Heidegger observes, "I am always somehow acquainted with myself" (1993, p. 251). Locke is more blunt: "consciousness *alone* makes self" (Locke, 1689 Bk. II, Ch. 27, Sec. 9; emphasis added).

While non-memory impaired individuals can recollect material with self-referential content—and often do so for legal, personal, or, more typically, social reasons—such recollections do not appear to be *required* for one's *feeling* of personal continuity. During most waking moments, I simply am I, an enduring, conscious presence given directly and pre-reflectively to awareness absent any analytic reckoning (e.g., Neuhouser, 1990; Kant, 1998; Klein, 2014a).

THE CONTINUITY OF THE SUBJECTIVE SELF: EVIDENCE-BASED DIACHRONICITY

In the section titled Sameness and the Self: The Problem of Personal Diachronicity I made the observation that the self of first-person subjectivity entails a feeling, and that this feeling does not vary over time. In that sense, it always is present as an "experiential given" underpinning our feeling of sameness (cf., James, 1890).

However, this is not to imply that this feeling serves as a *comparative* basis (i.e., with past feelings of sameness) thereby supporting a conclusion of temporal continuity. To do so would be to conflate the modes of operation of two ontologically distinct aspects of the self—the neuro-psychological (e.g., memory-based comparisons) with the subjective, non-evaluative aspect of self (my reasons for positing ontological separability—but causal relatedness—between these two aspects of the self are given in Klein (2014a). I cannot repeat them here, as to do so would greatly exceed the limits on word count for manuscripts of this type. Accordingly, the interested—or confused—reader is referred to arguments presented in detail in the above reference. I apologize in advance for any lack of clarity within the present text).

Moreover, to construe the felt invariance of the subjective self as a basis for comparative judgments of personal diachronicity would conflate evidential with felt sameness. These two modes of experiencing sameness, I am arguing, need to be kept both conceptually and functionally distinct.

However, for some this will seem to beg the question of why or how felt invariance translates into a directly given, conceptually unanalyzed sense of being a temporal continuant—a feeling that, under most circumstances, we take as default—i.e., it is an unreflected core aspect of our experiential being—and thus does not require (and is not subjected) to critical analysis.

Two considerations merit mention. First, there is no reason why felt sameness cannot be taken as an object of subjectivity and consequently evaluated. Indeed, I suspect it often is when motivation (either internally or externally mandated) argues in favor of considerations of evidential support for personal diachronicity. Second, however, I also am arguing that evidential criteria typically are not part of our experience of continuity. Rather, what underwrites are feeling of being a personal continuant is just that—the pre-reflectively given, conceptually unexamined feeling that I am I (e.g., Zahavi, 1999; Strawson, 2005). In this sense, personal diachronicity is not even (typically) a belief (though it can become so under circumstances calling for evidential warrant); rather it is a background presumption that is as much a part of our phenomenology as is the feeling that "I am alive" (e.g., we simply take it as an un-reflected given absent any analysis—though reasons can be provided when necessary).

With these considerations in mind, let's turn again, the case of patient D.B., whose uninterrupted access to personal memory was severely restricted. Might his intact subjectivity provide a basis for his sense of diachronicity? The answer depends on the criteria we use to investigate one's sense of personal diachronicity and the manner in which "sense of personal diachronicity" is conceptualized.

Seen in terms of *evidential* criteria, episodic memory loss renders patients such as D.B. and K.C. (e.g., Tulving, 1993; Klein et al., 2002c) unable to access information about their life history—i.e., their lived past as well as imagined future (for a recent review, see Klein, 2013a). Patient K.C., who suffered a total loss of episodic memory due to a motorcycle accident, describes his personal future as content-free and informationally vacant (it is important to note that individuals with

[8]For example, it should require time to reconstruct a coherent, "sufficiently" unbroken self-narrative; thus, evidential sources of diachronicity could not easily provide the immediate sense that I am the continuing existent I take it most people refer to when they claim to experience sameness of self over time.

intact episodic memory have no problem imagining content-rich, future-oriented personal scenarios; for a recent review see Klein, 2013a)[9].

> E.T.: (Endel Tulving): "Let's try the question again about the future. What will you be doing tomorrow?"
> K.C.: smiles faintly (following a 15-s pause) and responds: "I don't know."
> E.T.: "Do you remember the question?"
> K.C.: "About what I will be doing tomorrow?"
> E.T.: "Yes. How would you describe your state of mind when you try to think about it?"
> K.C., after a 5-s pause, replies: "Blank, I guess."

(Tulving, 1985; Tulving, p. 4: Note, in the original, patient K.C. was referred to as N.N.)

D.B. shows similar difficulties. When asked to provide information about personal events in his past, he is at a complete loss. In addition, he shows a conspicuous inability to project himself into an imagined future (Klein et al., 2002b).

These limitations in evidence-based mental time travel—and the difficulties they present for the sense of personal diachronicity construed as an evidence-based ability to subjectively navigate personal time—are not due to patients' difficulties comprehending the meaning of temporal concepts. Unpublished data (Klein, 2000; Craver, 2013) make plain that both K.C. and D.B. have a firm grasp of the concepts of past, present and future.

In response to the question "What is the future?" K.C. replies "Events that haven't happened yet," while the question "What is the past?" is answered "Events that have already happened". Asked "Can you change the past?" K.C. emphatically states "No!" When queried "Can you change the future, and if so, how?" he observes "Yes. By doing different things." To the question "Can something that happens in the future change what has happened in the past?" K.C. again responds with an emphatic "No," while the query "If an event is in the future will it always stay in the future?" elicits the response "No. Because time moves on."

Patient D.B. also presents a nuanced understanding of temporality. In response to the question "What is the future?" he answers "Things that haven't happened yet, but someday will." He describes the past as "Things that happened before…but are not happening now." Asked "Can you change the past?" D.B. says: "Don't think so, unless you had a time machine or something. Don't think so…not really. Maybe in science fiction (laughs)." To the question "Can the past influence the present?" he replies "Sure. All the time…that's the way things work."

In short, when the sense of personal diachronicity is conceptualized in terms of evidential criteria, individuals lacking total access to episodic memory (as well as suffering impairments of semantic personal knowledge) show a profound inability to engage in personally-relevant temporal extension. They are unable to (a) provide evidence-based knowledge of their personal past or (b) generate content-based personal future scenarios.

[9]Future-oriented mental time travel is well-known to depend on memory (for recent reviews see Szpunar, 2010; Schacter, 2012; Klein, 2013a).

THE CONTINUITY OF THE SUBJECTIVE SELF: FELT DIACHRONICITY

When personal diachronicity is considered in terms of *felt* rather than evidence-based criteria, however, a markedly different picture of personal continuity emerges. As we have seen, when he was asked to recall his past or to describe his possible future, D.B.'s interlocutors were met either with uncomfortable silence or expressed bewilderment.Gaping holes in his corpus of self-knowledge —brought to his attention by explicit requests—caused D.B. confusion, concern and fear; i.e., the type of reactions one would expect from a mentally coherent individual unable to fully comprehend the evidential vacuum experienced by his subjective self (Klein, 2012, 2014b).

This is a critical point, but one easily missed: When requested to provide evidence in support of his sense of personal diachronicity, D.B. expresses agitated concern: "I should, shouldn't I?" he wonders aloud. But he can't. In response to my query "Do you feel as though you are the same person you were before your heart attack?" D.B. replies: "If you mean, am I the same person…well not really. I have these head issues you know…can't seem to remember like I use to. But if you mean have I, D.B. (for confidentiality, this is not the name he actually used), lived a long life…well, of course. And I hope to keep at it." In short, D.B. is troubled when made aware (either by personal concerns or the requests of others) of the unavailability of evidence that, under normal circumstances, would be available to inform his sense of self as a temporal continuant.

This clearly is *not* a person lacking a sense of temporal persistence (although he is unable martial evidence in support of that sense). He is concerned about the fate that has befallen his (apparently intact) feeling of himself as an enduring entity. What he lacks is the ability to supplement this feeling with evidential offerings from his material self. Interestingly, the absence of an ability to recollect a personal past or imagine a personal future does not appear either to trouble or to capture the attention of his subjective sense of self *unless* the situation makes his deficits the object of his awareness.

A similar appreciation of the continuity of self in the presence of evidential deficit is found with patient H.M. Replying to the question "How do you feel about yourself?" he observes "I feel I have failed more than the average person…I feel like a complete failure as a person…I am disappointed in myself." (Hilts, 1995, p. 153). Like D.B., H.M. may not be able to offer evidential support for his feeling of continuity, but he clearly feels himself to be a temporal continuant, one whose past acts have failed to meet his current expectations. Apparently, something more than *evidential* criteria is at work in underwriting one's sense of diachronicity.

A particularly compelling example of intact sense of personal diachronicity in the presence of severe impairment to the evidential bases for that felt sameness comes from the case of Zasetsky (Luria, 1972). Zasetsky was a Russian soldier, who, as the result of battle, was left aphasic, perceptually and proprioceptively disoriented and hemianopic. He also became densely amnesic, with severe impairment (both antrograde and retrograde) of episodic as well as semantic memory function.

As a result of deficits in proprioception and kinesthetic feedback, Zasetsky had trouble feeling and locating parts of his own body. His perception of the external world suffered as well.

External objects either were nonexistent or appeared as fragmented, flickering background entities.

Having lost most of his personal memory, his ability to recall his past and plan for his future was virtually non-existent (for a recent discussion of the relation between memory and mental time travel, see Klein, 2013a). He professed to have no clear idea of his preferences, beliefs, values, or goals. In short, Zasetsky was unable to access most of his sources of epistemic self-knowledge.

Despite the great challenges presented, Zasetsky struggled to piece together the evidential fragments that remained from his material self. Under the patient tutelage of Luria and others, he slowly and painfully regained some rudimentary ability to read, write and perform basic bodily functions. As a consequence, he was able to provide Luria with a record of his thoughts and feelings about the changes to the self brought about by his difficulty providing first-person subjectivity with content from the now largely dysfunctional aspects of his material self.

But- and this is the key point—despite monumental loss of access to material bases of self, Zasetsky maintained a feeling of personal sameness. *He* was painfully aware of *his* deficits and greatly troubled by their effects on *his* ability to place *himself* physically, temporally and spatially. *He* complained often about the confusion engendered by impairments of perceptual, kinesthetic and proprioceptive feedback; *he* was disorientated by *his* loss of preferences and difficulties imagining *his* future or recalling *his* past.

Thus, at no time was his subjective self-awareness lost (save, perhaps, periods of dreamless sleep): The "I" always was there—troubled, bewildered, angered, and confused by its loss of access to sources of self-knowledge, yet determined to salvage whatever it could of a life left in cognitive and perceptual shambles. In the end, it was this subjectively felt determination to improve his situation that led Zasetsky to undertake the arduous rehabilitative program that enabled the subjective self to regain partial contact with the external world and aspects of the material self. He doggedly maintained hope for a life better than the one that had befallen him in battle. And "hope" is word whose meaning unambiguously implies a sense of self as a personal continuant.

In short, there is strong empirical support for the proposition that a person, absent most of what we would place under the heading of "material self" still can retain a clear feeling of his or her sameness and temporal continuity. What is particularly noteworthy in Zaztesky's case are his concerted efforts to distance himself from what he had become and recapture a semblance of normality.

THE TIMELESSNESS OF THE SUBJECTIVE SELF

One fascinating, but often overlooked, aspect of the experience of patients with temporally graded amnesia (e.g., the law of "first in, last out"; Ribot, 1882) is that the subjective aspect of self typically is not confused by, or troubled over, its inability to recollect events and experiences covered by memory loss (unless, of course, the self is confronted with evidence of the incongruity between the passage of time and current self-beliefs. Absent such confrontation, the patient appears relatively content to see him or herself as being of the age at which personal memories remains available; for review and discussion, see Klein, 2012).

Consider, for example, the case of patient, J.G. (Sacks, 1985). As the result of Korsakoff syndrome, J.G. was unable to recollect any personal happenings postdating 1948. Despite passage of nearly 30 years since the onset of his amnesia, testing revealed that J.G. believes he still is a young man, and that the year still is 1948 (it was 1975). Consistent with his beliefs, on being shown his face in a mirror (i.e., that of a much older man) J.G. is stunned and confused. Fortunately, due to his amnesia, after a few moments of distraction and J.G. once again is relaxed and comfortably situated in 1948.

The remarkable case of patient B. (Storring, 1936) brings the relation between memory, personal temporality and the subjective self into strong relief. As a result of a gas poisoning accident, patient B. was rendered incapable of remembering anything occurring post-injury for more than roughly one second! At the time of testing (the mid-1930's) he knew nothing of the life he had lived post-poisoning or of his marriage of the past 5 years. Like J.G., he is perplexed every time he sees himself in a mirror – 10 years earlier he looked much different. While psycho-physcial aspects of his self have changed with time, the subjective aspect of self-shows no comparable evidence of change: For B.'s first-person subjectivity, it is, and always will be May 1926.

There are many aspects of this case that merit extensive discussion. For my purposes, however, the relevant features pertain to what it can tell us about B.'s self of first-person subjectivity, a self whose knowledge of the aging process has been decoupled from changes to the material self-brought about by the passage of time. The subjective self, no longer having access to these changes, does not show a parallel aging of its own. B. has become a man of the eternal present.

However, as Storring (1936) notes at length, B. is *not* a man of the moment: "B. gives meaning to the situation before his senses. And it is this context that reaches from one second to the next that creates the flowing transition. A sensible, reasonable task is harmoniously carried to its completion, regardless of how long it takes, because . . . the rational whole is known in the situation as a goal which is then fulfilled" (Storring, 1936, pp. 75–76). This is a person, Storring concludes, with a second-long consciousness that nevertheless has a clear sense of personal continuity. The subjective self, anchored in the past as a result of disruption of sensory and cognitive processes, nevertheless, remains a constant, experiencing, feeling, thinking center of subjectivity unperturbed by the passage of time.

The take-away message is that seldom, if ever, do we find a patient who claims to experience himself as much older than his or her intact recollections would suggest; rather, we find the reverse—the patient resides in the past (provided he or she has access to some personal recollections) and is troubled only when a discrepancy between content provided by the material self (or one's senses) fails to match current beliefs (Klein, 2012). The self of first-person subjectivity thus seems outside of the aging process, accepting whatever the material self has to offer vis a vis evidence of temporal placement.

PERSONAL DIACHRONICITY AND THE SENSE OF SELF

At this point in the discussion, a reasonable question concerns the extent to which personal diachronicity is a "phenomenological

given." That is, to what degree are considerations of self-continuity the result of social requests, moral obligations and conceptual curiosity, as opposed to a basic attitude we spontaneously adopt toward our everyday experience of self?

If one accepts the proposition that first-person subjectivity is a pre-reflectively given and ageless aspect of the subjective self, then diachronicity *per se* may not actually be a live issue for one's everyday sense of self. It may only become so when a person attempts to provide *evidential* support for personal continuance (e.g., from memories supplied by the material aspect of self). On this view, diachronicity is not so much felt as it is logically constructed from informational content (which, as we have seen, does not easily translate into unambiguous criteria for personal continuity).

Personal diachronicity thus may come into play largely at the evidential level. Since, the self of first-person subjectivity is not experienced as changing with time, considerations of personal diachronicity typically are not a part of our sense of self. They become a part when the self is taken as an object of reflection; and this, in turn, occurs when we are called on—either by personal concerns or external contingencies—to directly address issues pertaining to the self as a temporal continuant.

An obvious objection would be that "the reason that the self of subjectivity in not experienced as changing is because the self of subjectivity is not experienced at all." While such an objection has some force, I believe there are two different responses that can partly address this concern. First, first-person subjectivity is perhaps the single most salient aspect of everyday experience. Of course, it is always (per Brentano) conflated with intentional objects (although advocates of "pure consciousness" argue that subjective states absent intentional objects can, with extensive training, be attained; e.g., Forman, 1990). But, as many have argued (for recent reviews, see Legrand and Ruby, 2009; Klein, 2012, 2014a), our acquaintance with the self of first-person subjectivity is a necessary postulate to capture what we mean by a sense of oneself. Second, the evidence presented in this paper of patient's suffering varying degrees of cognitive impairment, yet still maintaining a coherent sense of diachronicity are consistent with the notion that what underwrites this feeling is the constancy of subjectivity and not the flickering or non-existent objects taken by that subjectivity. While these arguments are consistent with the assumption that the felt invariance of subjectivity underlies the sense of sameness, it must be acknowledged that neither provides a conclusive refutation of the objection raised. Accordingly, it remains a live possibility.

In summary, we typically do not *feel* ourselves to be different over time. When we do, it most often is the result of our continuity being called to question by self or other. Moreover, to the extent that memory, in particular, and evolutionary considerations, in general, play a part in our sense of temporal continuity, it is the "now and the next", not the past, to which our pre-reflective sentiments gravitate (for discussion of the future-orientation of memory, see Klein, 2013b). As Strawson (personal communication) puts it, the temporality of subjective self consists in "and now and now and now."[10]

We should not draw from these observations the conclusion that the subjective aspect of self necessarily is immortal or transcendental. It very well may be incapable of existing apart from the body (e.g., Olson, 2007). It may be an emergent property of the material self (e.g., Hasker, 1999). But this emergence—if indeed it is emergence—is something we clearly do not know how to deal with within the context of current theory and research in science and philosophy.

We are a long way from beginning to answer questions about the self of first-person subjectivity. Yet, in my opinion, answers to these questions are fundamental for a psychology that takes as its goal the full appreciation of human experience.

SOME FINAL THOUGHTS

In this paper, I have argued that our non-analytic, pre-reflective feeling of the self may be the primary determinant of our intuition of self as temporally extended. Explicit considerations of personal diachronicity come into play primarily when contingencies make it necessary to contemplate (and provide evidence in support of) personal continuity. In their absence, one's sense of self as a temporal continuant is more one of unreflected acceptance than explicit formulation. Sameness is the self's default mode – a felt identity uninformed by evidence. And, in virtue of the unchanging nature of subjectivity, diachronicity becomes a concern of the self only when considerations of personal temporality are selected by environmental demand, personal concern or philosophical query to serve as the objects of subjectivity[11].

Plato maintained that true knowledge could only be sensed by the soul. Aristotle, in contrast, believed knowledge is derived from evidence provided by the body. The tension between sense and evidence has been a source of academic, social, political and religious debate (often acrimonious) for more than two millennia, with the emphasis shifting as a function of cultural as well as intellectual imperatives (e.g., Koestler, 1989).

In this paper I have made my case for the non-evidential basis of one's sense of sameness over time. However, this should not be seen as a call to reaffirm the Platonic distaste of understanding by reliance on the "grossness of bodily senses." Rather, it is an appeal to broaden our criteria for understanding beyond the reductionist materialism that characterizes much of Western thinking, and to embrace the possibility that there are aspects of reality that may not (easily) submit to such highly circumscribed treatment (e.g., Meixner, 2008; Papa-Grimaldi, 2010; Nagel, 2012; Klein, 2014a).

[10]The Earl of Shaftesbury (1698) captures in a few well-chosen sentences much of what I have been struggling to say: "The metaphysicians…affirm that if memory be taken away, the self is lost. [But] what matter for memory? What have I to do with that part? If, whilst I am, I am as I should be, what do I care more? And thus let me lose self every hour, and be 20 successive selfs, or new selfs,' tis all one to me: so [long as] I lose not my opinion [i.e., my overall outlook, my character, my moral identity]. If I carry that with me'tis I; all is well…—The now; the now. Mind this: in this is all." (cited in Strawson, 2008, p. 198, parenthetical comments added).

[11]Ricoeur's (1994) distinction between sameness (memete) and selfhood (ipseite) is particularly germane to our discussion of personal diachronicity. Ricoeur argues that selfhood is maintained despite changes in the evidential criteria for sameness (i.e., character dispositions and other marks that permit the reindentification of an individual as being the same over time). This view, based exclusively on philosophical considerations, offers strong support for our empirical observations concerning the insufficiency of evidential criteria for continuity of self.

Some phenomena, if they are to be saved, can be saved only if we allow that knowledge by acquaintance (e.g., Russell, 1912/1999) sometimes may be the metaphysically propitious stance. Personal diachronicity very well may be a case in point.

An appreciation of "reality" in its fullness likely requires we strike a balance between the different approaches to knowledge championed by Plato and Aristotle. Only by affecting a rapprochement between these "seemingly" conflicting metaphysical commitments will a sufficiently inclusive understanding of reality be a potentially realizable objective.

REFERENCES

Albahari, M. (2006). *Analytical Buddhism: The Two-Tiered Illusion of Self*. Houndsmills: Palgrave Macmillan. doi: 10.1057/9780230800540

Baillie, J. (1993). *Problems in Personal Identity*. New York, NY: Paragon House.

Baker, L. R. (2000). *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139173124

Barbour, J. (2000). *The End of Time*. New York, NY: Oxford University Press.

Bernecker, S. (2010). *Memory: A Philosophical Study*. Oxford: Oxford University Press.

Bluck, S., and Liao, H.-W. (2013). I was therefore I am: creating self-continuity through remembering our personal past. *Int. J. Reminisc. Life Rev.* 1, 7–12.

Bolander, T., Hendricks, V. F., and Pedersen, S. A. (2006). *Self-Reference*. Stanford, CA: CSLI Publications.

Brennan, A. (1985). Amnesia and psychological continuity. *Can. J. Philos.* 11, 195–209.

Brennan, A. (1988). *Conditions of Identity: A Study in Identity and Survival*. Oxford: Oxford University Press.

Butler, J. (1736/1819). "Of personal identity," in *The Analogy of Religion* ed J. Angus (London: Allman and Sawers), 211–215.

Campbell, S. (2004). Rapid psychological change. *Analysis* 64, 256–264. doi: 10.1093/analys/64.3.256

Collins, A. W. (1997). Personal identity and the coherence of q-memory. *Philos. Q.* 47, 73–80. doi: 10.1111/1467-9213.00047

Conway, M. A. (2005). Memory and the self. *J. Memory Lang.* 53, 594–628. doi: 10.1016/j.jml.2005.08.005

Craver, C. F. (2012). A preliminary case for amnesic selves: toward a clinical moral psychology. *Soc. Cogn.* 30, 449–473. doi: 10.1521/soco.2012.30.4.449

Craver, C. F. (2013). *Temporal Concepts in Patients Suffering Complete Episodic Amnesia*. Talk presented to the department of Philosophy, UCSB.

Dainton, B. (2008). *The Phenomenal Self*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199288847.001.0001

Duval, C., Desgranges, B., de La Sayette. V., Belliard, S., Eustache, F., and Piolino, P. (2012). What happens to personal identity when semantic knowledge degrades? A study of the self and autobiographical memory in semantic dementia. *Neuropsychologia* 50, 254–265. doi: 10.1016/j.neuropsychologia.2011.11.019

Earle, W. (1955). *Objectivity: An Essay on Phenomenological Ontology*. New York, NY: The Noonday Press.

Earle, W. E. (1972). *The Autobiographical Consciousness*. Chicago, IL: Quadrangle Books.

Falk, A. (1995). Consciousness and self-reference. *Erkenntnis* 43, 151–180. doi: 10.1007/BF01128194

Fivush, R., and Haden, C. A. (2003). *Autobiographical Memory and the Construction of a Narrative Self*. Mahwah, NJ: Lawrence Erlbaum Associates.

Focquaert, F. (2003). Personal identity and its boundaries: philosophical thought experiments. *Philosophica* 72, 131–152.

Forman, R. K. C. (Ed.) (1990). *The Problem of Pure Consciousness*. Oxford, UK: Oxford University Press.

Gallagher, S., and Zahavi, D. (2008). *The Phenomenological Mind*. New York, NY. Routledge.

Gallios, A. (1998). *Occasions of Identity*. Oxford: Oxford University Press.

Ganeri, J. (2012). *The Self: Naturalism, Consciousness, and the First-Person Stance*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199652365.001.0001

Garrett, B. (1998). *Personal Identity and Self-Consciousness*. London: Routledge. doi: 10.4324/9780203015667

Gergen, K. J. (1971). *The Concept of Self*. New York, NY: Holt, Rinehart, and Winston, Inc.

Gillihan, S. J., and Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychol. Bull.* 131, 76–97. doi: 10.1037/0033-2909.131.1.76

Greenwald, A. G. (1981). "Self and memory," in *The Psychology of Learning and Motivation*, Vol. 15, ed G. H. Bower (New York, NY: Academic Press), 201–236.

Greenwood, T. (1967). Personal identity and memory. *Philos. Q.* 17, 334–344. doi: 10.2307/2217455

Hamilton, A. (1995). A new look at personal identity. *Philos. Q.* 45, 332–349. doi: 10.2307/2219654

Hasker, W. (1999). *The Emergent Self*. Ithaca, NY: Cornell University Press.

Haslam, C., Jetten, J., Haslam, S. A., Pugliese, C., and Tonks, J. (2010). "I remember therefore I am, and I am therefore I remember": exploring the contributions of episodic and semantic self-knowledge to strength of identity. *Br. J. Psychol.* 102. 184–203. doi: 10.1348/000712610X508091

Heidegger, M. (1993). *Grundprobleme Der Phanomenologie (1919–1920). Gesamtausgaba Band 58*. Frankfurt: Vittorio Klostermann.

Hilts, P. J. (1995). *Memory's Ghost: The Strange Case of Mr. M and the nature of memory.* New York, NY: Simon and Schuster.

Hirsch, E. (1982). *The Concept of Identity*. New York, NY; Oxford University Press.

Hume, D. (1739–1740/1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Husserl, E. (1964). *The Phenomenology of Internal Time-Consciousness*. Bloomington, IN: Indiana University Press.

Illman, N. A., Rathbone, C. J., Kemp, S., and Moulin, C. J. A. (2011). Autobiographical memory and the self in a case of transient epileptic amnesia. *Epilepsy Behav.* 21, 36–41. doi: 10.1016/j.yebeh.2011.02.022

James, W. (1890). *Principles of Psychology,* Vol. 1. New York, NY: Henry Holt and Company. doi: 10.1037/11059-000

Johnstone, H. W. (1970). *The Problem of the Self*. University Park, PA: The Pennsylvania State University Press.

Kant, I. (1998). *The Cambridge Edition of the Works of Immanuel Kant; Critique of Pure Reason*. New York, NY: Cambridge University Press.

Kihlstrom, J. F., and Klein, S. B. (1994). "The self as a knowledge system," in *Handbook of Social Cognition.* Vol. 1. *Basic Processes* eds R. S. Wyer and T. K. Srull. (Hillsdale, NJ: Erlbaum), 153–208.

Klein, S. B. (2000). *Transcripts of patient D.B.* Unpublished raw data, UCSB.

Klein, S. B. (2001). "A self to remember: a cognitive neuropsychological perspective on how self creates memory and memory creates self," in *Individual Self, Relational Self, and Collective Self,* eds C. Sedikides and M. B. Brewer (Philadelphia, PA: Psychology Press), 25–46.

Klein, S. B. (2004). "The cognitive neuroscience of knowing one's self," in *The Cognitive Neurosciences III,* ed M. A. Gazzaniga (Cambridge, MA: MIT Press), 1007–1089.

Klein, S. B. (2012). The self and its brain. *Soc. Cogn.* 30, 474–516. doi: 10.1521/soco.2012.30.4.474

Klein, S. B. (2013a). The complex act of projecting oneself into the future. *WIREs Cogn. Sci.* 4, 63–79. doi: 10.1002/wcs.1210

Klein, S. B. (2013b). The temporal orientation of memory: it's time for a change of direction. *J. Res. Appl. Memory Cogn.* 2, 222–234. doi: 10.1016/j.jarmac.2013.08.001

Klein, S. B. (2014a). *The Two Selves: Their Metaphysical Commitments and Functional Independence***.** New York, NY: Oxford University Press.

Klein, S. B. (2014b). The sense of diachronic personal identity. *Phenomenol. Cogn. Sci.* 12, 791–811.

Klein, S. B., Altinyazar, V., and Metz, M. A. (2013). Facets of self in schizophrenia: the reliability and accuracy of trait self-knowledge. *Clin. Psychol. Sci.* 1, 276–289. doi: 10.1177/2167702612474263

Klein, S. B., Chan, R. L., and Loftus, J. (1999). Independence of episodic and semantic self-knowledge: the case from autism. *Soc. Cogn.* 17, 413–436. doi: 10.1521/soco.1999.17.4.413

Klein, S. B., Cosmides, L., Murray, E. R., and Tooby, J. (2004). On the Acquisition of knowledge about personality traits: does learning about the self engage different mechanisms than learning about others? *Soc. Cogn.* 22, 367–390. doi: 10.1521/soco.22.4.367.38295

Klein, S. B., Cosmides, L., Tooby, J., and Chance, S. (2002a). Decisions and the evolution of memory: multiple systems, multiple functions. *Psychol. Rev.* 109, 306–329. doi: 10.1037/0033-295X.109.2.306

Klein, S. B., Loftus, J., and Kihlstrom, J. F. (2002b). Memory and temporal experience: the effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Soc. Cogn.* 20, 353–379. doi: 10.1521/soco.20.5.353.21125

Klein, S. B., Rozendal, K., and Cosmides, L. (2002c). A social-cognitive neuroscience analysis of the self. *Soc. Cogn.* 20, 105–135. doi: 10.1521/soco.20.2.105.20991

Klein, S. B., and Gangi, C. E. (2010). The multiplicity of self: neuropsychological evidence and its implications for the self as a construct in psychological research. *Ann. N.Y. Acad. Sci.* 1191, 1–15. doi: 10.1111/j.1749-6632.2010.05441.x

Klein, S. B., and Lax, M. L. (2010). The unanticipated resilience of trait self-knowledge in the face of neural damage. *Memory* 18, 918–948. doi: 10.1080/09658211.2010.524651

Klein, S. B., Loftus, J., and Kihlstrom, J. F. (1996). Self-knowledge of an amnesic patient: toward a neuropsychology of personality and social psychology. *J. Exp. Psychol. Gen.* 125, 250–260. doi: 10.1037/0096-3445.125.3.250

Klein, S. B., and Nelson, C. R. (2014). "The effects of self-reference on memory: a conceptual and methodological review of inferences warranted by the self-reference effect," in *The Sage Handbook of Applied Memory* eds T. J. Perfect and D. S. Lindsay (Los Angeles, CA: Sage Publications Ltd), 256–272.

Klein, S. B., and Nichols, S. (2012). Memory and the sense of personal identity. *Mind* 121, 677–702. doi: 10.1093/mind/fzs080

Klein, S. B., Robertson, T. E., Gangi, C. E., and Loftus, J. (2008). The functional independence of trait self-knowledge: commentary on Sakaki (2007). *Memory* 16, 556–565. doi: 10.1080/09658210802010489

Koestler, A. (1989). *The Sleepwalkers: A History of Man's Changing Vision of the Universe*. New York, NY: Arkana.

Koons, R. C., and Bealer, G. (2010). *The Waning of Materialism*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199556182.001.0001

Leary, M. R., and Tangney, J. P. (2003). *Handbook of Self and Identity*. New York, NY: The Guilford Press.

Leary, M. R., and Tangney, J. P. (2012). *Handbook of Self and Identity,* 2nd Edn. New York, NY: The Guilford Press.

Legrand, D., and Ruby, P. (2009). What is self specific? A theoretical investigation and a critical review of neuroimaging results. *Psychol. Rev.* 116, 252–282. doi: 10.1037/a0014172

Lewis, H. D. (1982). *The Elusive Self*. Philadelphia, PA: Westminster Press.

Locke, J. (1689–1700/1975). *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.

Lund, D. H. (2004). *Perception, Mind and Personal Identity: A Critique of Materialism*. New York, NY: University Press of America, Inc.

Lund, D. H. (2005). *The Conscious Self*. Amherst, NY: Humanity Books.

Luria, A. R. (1972). *The Man with a Shattered World*. Cambridge, MA: Harvard University Press.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. New York, NY: W. H. Freeman.

Markowitsch, H. J., and Staniliou, A. (2011). Memory, autonoetic consciousness, and the self. *Conscious. Cogn.* 20, 16–39. doi: 10.1016/j.concog.2010.09.005

Martin, C. B. (2008). *The Mind in Nature*. Oxford: Oxford University Press.

Martinelli, P., Sperduti, M., and Piolino, P. (2013). Neural substrates of the self-memory system: new insights from a meta-analysis. *Hum. Brain Mapp.* 34, 1501–1747. doi: 10.1002/hbm.22008

Meixner, U. (2005). Physicalism, dualism and intellectual honesty. *Dual. Rev.* 1, 1–20.

Meixner, U. (2008). "The reductio of reductive and non-reductive materialism – and a new start," in *Psycho-Physical Dualism: An Interdisciplinary Approach,* eds A. Antonietti, A. Corradini, A., and E. J. Lowe (Boulder, CO: Rowman and Littlefield Publishers, Inc), 143–166.

Metzinger, T. (2009). *The Ego Tunnel: The Science of Mind and the Myth of the Self*. New York, NY: Basic Books.

Nagel, T. (1974). What is it like to be a bat? *Philos. Rev.* 83, 435–450. doi: 10.2307/2183914

Nagel, T. (2012). *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly Wrong*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199919758.001.0001

Neisser, U. (1988). Five kinds of self–knowledge. *Philos. Psychol.* 1, 35–39. doi: 10.1080/09515088808572924

Neuhouser, F. (1990). *Fichte's Theory of Subjectivity*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511624827

Noonan, H. (1983). Personal identity and bodily continuity. *Analysis* 43, 98–104.

Noonan, H. (1989). *Personal Identity,* 2nd Edn. New York, NY: Routledge.

Oderberg, D. S. (1993). *The Metaphysics of Identity*. New York, NY: St Martin's Press, Inc. doi: 10.1057/9780230377387

O'Driscoll, K., and Leach, J. P. (1998). "No longer Gage": an iron bar through the head. Early observations of personality change after injury to the prefrontal cortex. *Br. Med. J.* 19, 1673–1674. doi: 10.1136/bmj.317.7174.1673a

Olson, E. T. (1997). *The Human Animal: Personal Identity without Psychology*. New York, NY: Oxford University Press.

Olson, E. T. (2007). *What are We? A study in Personal Ontology.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195176421.001.0001

Papa-Grimaldi, A. (1998). *Time and Reality*. Aldershot: Ashgate.

Papa-Grimaldi, A. (2010). *Orphans of the One or the Deception of the Immanence*. Bern: Peter Lang.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Perry, J. (2008). *Personal Identity*. Berkeley, CA: University of California Press.

Picard, L., Mayor-Dubois, C., Maeder, P., Kalenzaga, F., Abram, M., Duval, C., et al. (2013). Functional independence within the self-memory system: new insights from two cases of developmental amnesia. *Cortex* 49, 1463–1481. doi: 10.1016/j.cortex.2012.10.003

Prebble, S., Addis, D. R., and Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychol. Bull.* 139, 815–840. doi: 10.1037/a0030146

Prigatano, G. P., and Schacter, D. L. (1991). *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues*. New York, NY: Oxford University Press.

Proust, J. (2003). Thinking of oneself as the same. *Conscious. Cogn.* 12, 495–509. doi: 10.1016/S1053-8100(03)00077-1

Rathbone, C. J., Moulin, C. J. A., and Conway, M. A. (2009). Autobiographical memory and amnesia: using conceptual knowledge to ground the self. *Neurocase* 15, 405–418. doi: 10.1080/13554790902849164

Reid, T. (1813/1969). *Essays on the Intellectual Powers of Man*. Cambridge, MA: MIT Press.

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch. M., and Levine, B. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends Cogn. Sci.* 16, 550–558. doi: 10.1016/j.tics.2012.09.003

Rescher, N. (1997). *Objectivity: The Obligations of Impersonal Reason*. Notre Dame, IN: University of Notre Dame Press.

Ribot, T. H. (1882). *Diseases of Memory*. New York, NY: D. Appleton and Company.

Ricard, M., and Thuan, T. X. (2001). *The Quantum and the Lotus*. New York, NY: Three Rivers Press.

Ricoeur, P. (1994). *Oneself as Another*. Chicago, IL: University of Chicago Press.

Rorty, A. O. (1976). *The Identities of Persons*. Berkeley, CA: University of California Press.

Rubin, D. C. (1986). *Autobiographical Memory*. Cambridge: Cambridge University Press.

Russell, B. (1912/1999). *The Problems of Philosophy*. Mineola, NY: Dover Publications.

Russell, B. (1949). *History of Western Philosophy, and its Connection with Political and Social Circumstance from the Earliest Times to the Present Day*. London: George Allen and Unwin, Ltd.

Sacks, O. (1985). *The Man Who Mistook his Wife for a Hat*. New York, NY: Harper Collins.

Sani, F. (2008). *Self Continuity: Individual and Collective Perspectives*. New York, NY: Psychology Press.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *Am. Psychol.* 67, 603–613. doi: 10.1037/a0029869

Schechtman, M. (1990). Personhood and personal identity. *J. Philos.* 87, 71–92. doi: 10.2307/2026882

Schechtman, M. (2001). Empathic access: the missing ingredient in personal identity. *Philos. Explor.* 2, 95–111. doi: 10.1080/10002001058538710

Sedikides, C., and Spencer, S. J. (2007). *The Self*. New York, NY: Psychology Press.

Shoemaker, S. (1963). *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press.

Shoemaker, S., and Swinburne, R. (1984). *Personal Identity*. Oxford: Basil Blackwell, LTD.

Siderits, M., Thompson, E., and Zahavi, D. (2011). *Self, no Self: Perspectives from Analytical Phenomenological and Indian Traditions*. Oxford: Oxford University Press.

Slors, M. (2001). *The Diachronic Mind: An Essay on Personal Identity, Psychological Continuity and the Mind-body Problem*. Boston, MA: Kluwer Academic Publishers. doi: 10.1007/978-94-017-3276-5

Snodgrass, J. G., and Thompson, R. L. (1997). *The Self Across Psychology: Self-Recognition, Self-Awareness, and the Self-Concept*. Annals of the New York Academy of Sciences, Vol. 818. New York, NY: New York Academy of Sciences.

Spencer Brown, G. (1957). *Probability and Scientific Inference*. London: Longmans, Green and Co.

Squire, L. R., and Kandel, E. R. (1999). *Memory: From Mind to Molecules*. New York, NY: Freeman.

Stern, D. (1985). *The Interpersonal World of the Infant*. New York, NY: Basic Books.

Storring, G. E. (1936). Memory loss by gas poisoning: a man without memory of time. *Archiv fur die gesamte Psychologie* 95, 436–511.

Strawson, G. (2005). *The Self?* Oxford: Blackwell Publishing.

Strawson, G. (2008). *Real Materialism: and other Essay*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199267422.001.0001

Strawson, G. (2009). *Selves: An Essay in Revisionary Metaphysics*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198250067.001.0001

Strawson, G. (2011a). *The Evident Connection*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199608508.001.0001

Strawson, G. (2011b). *Locke on Personal Identity: Consciousness and Concernment*. Princeton, NJ: Princeton University Press.

Swinburne, R. (2013). *Mind, Brain, and Free Will*. Oxford: Oxford University Press.

Szpunar, K. K. (2010). On subjective time. *Cortex* 47, 409–411. doi: 10.1016/j.cortex.2010.07.008

Terrace, H. S., and Metcalfe, J. (2005). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195161564.001.0001

Tulving, E. (1983). *Elements of Episodic Memory*. New York, NY: Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12. doi: 10.1037/h0080017

Tulving, E. (1993). "Self-knowledge of an amnesic individual is represented abstractly," in *Advances in Social Cognition,* Vol. 5 eds T. K. Srull and R. S. Wyer (Hillsdale, NJ: Erlbaum), 147–156.

Tulving, E., Schacter, D. L., McLachlan, D. R., and Moscovitch, M. (1988). Priming of semantic autobiographical knowledge: a case study of retrograde amnesia. *Brain Cogn.* 8, 3–20. doi: 10.1016/0278-2626(88)90035-8

Vierkant, T. (2003). *Is the Self Real?* London: Transaction Publishers.

Wiggins, D. (1980). *Sameness and Substance*. Oxford: Blackwell.

Wiggins, S. (1971). *Identity and Spatio-Temporal Continuity*. Oxford: Basil Blackwell.

Wilkes, K. V. (1988). *Real People: Personal Identity without thought Experiments*. Oxford: Oxford University Press.

Williams, B. (1970). The self and the future. *Philos. Rev.* 79, 161–180. doi: 10.2307/2183946

Williams, B. (1973). *Problems of the Self*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511621253

Williams, C. F. J. (1990). *What is Identity?* Oxford: Oxford University Press.

Williams, T. (2002). "Vagueness, identity and Leibniz' law," in *Individuals, Essence and Identity: Themes of Analytic Metaphysics,* eds P. Giaretta, A. Bottani, and M. Carrara (Dordrecht: Kluwer), 273–304. doi: 10.1007/978-94-017-1866-0_14

Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston, IL: Northwestern University Press.

Zahavi, D. (2003). *Husserl's Phenomenology*. Stanford, CA: Stanford University Press.

Zahavi, D. (2005). *Subjectivity of Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: The MIT Press.

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: 03 December 2013; accepted: 10 January 2014; published online: 29 January 2014.

*Citation: Klein SB (2014) Sameness and the self: philosophical and psychological considerations. Front. Psychol. **5**:29. doi: 10.3389/fpsyg.2014.00029*

This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.

Copyright © 2014 Klein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 17 June 2014
doi: 10.3389/fpsyg.2014.00577

Objects of consciousness

Donald D. Hoffman[1]* and Chetan Prakash[2]

[1] *Department of Cognitive Sciences, University of California, Irvine, CA, USA*
[2] *Department of Mathematics, California State University, San Bernardino, CA, USA*

Edited by:
Chris Fields, New Mexico State University, USA (retired)

Reviewed by:
John Serences, University of California San Diego, USA
David Marcus Appleby, University of Sydney, Australia

****Correspondence:***
Donald D. Hoffman, Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA
e-mail: ddhoff@uci.edu

Current models of visual perception typically assume that human vision estimates true properties of physical objects, properties that exist even if unperceived. However, recent studies of perceptual evolution, using evolutionary games and genetic algorithms, reveal that natural selection often drives true perceptions to extinction when they compete with perceptions tuned to fitness rather than truth: Perception guides adaptive behavior; it does not estimate a preexisting physical truth. Moreover, shifting from evolutionary biology to quantum physics, there is reason to disbelieve in preexisting physical truths: Certain interpretations of quantum theory deny that dynamical properties of physical objects have definite values when unobserved. In some of these interpretations the observer is fundamental, and wave functions are compendia of subjective probabilities, not preexisting elements of physical reality. These two considerations, from evolutionary biology and quantum physics, suggest that current models of object perception require fundamental reformulation. Here we begin such a reformulation, starting with a formal model of consciousness that we call a "conscious agent." We develop the dynamics of interacting conscious agents, and study how the perception of objects and space-time can emerge from such dynamics. We show that one particular object, the quantum free particle, has a wave function that is identical in form to the harmonic functions that characterize the asymptotic dynamics of conscious agents; particles are vibrations not of strings but of interacting conscious agents. This allows us to reinterpret physical properties such as position, momentum, and energy as properties of interacting conscious agents, rather than as preexisting physical truths. We sketch how this approach might extend to the perception of relativistic quantum objects, and to classical objects of macroscopic scale.

Keywords: consciousness, quantum theory, Markov chains, combination problem, geometric algebra

INTRODUCTION

The human mind is predisposed to believe that physical objects, when unperceived, still exist with definite shapes and locations in space. The psychologist Piaget proposed that children start to develop this belief in "object permanence" around 9 months of age, and have it firmly entrenched just 9 months later (Piaget, 1954). Further studies suggest that object permanence starts as early as 3 months of age (Bower, 1974; Baillargeon and DeVos, 1991).

Belief in object permanence remains firmly entrenched into adulthood, even in the brightest of minds. Abraham Pais said of Einstein, "We often discussed his notions on objective reality. I recall that on one walk Einstein suddenly stopped, turned to me and asked whether I really believed that the moon exists only when I look at it" (Pais, 1979). Einstein was troubled by interpretations of quantum theory that entail that the moon does not exist when unperceived.

Belief in object permanence underlies physicalist theories of the mind-body problem. When Gerald Edelman claimed, for instance, that "There is now a vast amount of empirical evidence to support the idea that consciousness emerges from the organization and operation of the brain" he assumed that the brain exists when unperceived (Edelman, 2004). When Francis Crick asserted the "astonishing hypothesis" that "You're nothing but a pack of neurons" he assumed that neurons exist when unperceived (Crick, 1994).

Object permanence underlies the standard account of evolution by natural selection. As James memorably put it, "The point which as evolutionists we are bound to hold fast to is that all the new forms of being that make their appearance are really nothing more than results of the redistribution of the original and unchanging materials. The self-same atoms which, chaotically dispersed, made the nebula, now, jammed and temporarily caught in peculiar positions, form our brains" (James, 1890). Evolutionary theory, in the standard account, assumes that atoms, and the replicating molecules that they form, exist when unperceived.

Object permanence underlies computational models of the visual perception of objects. David Marr, for instance, claimed "We ... very definitely do compute explicit properties of the real visible surfaces out there, and one interesting aspect of the evolution of visual systems is the gradual movement toward the difficult task of representing progressively more objective aspects of the visual world" (Marr, 1982). For Marr, objects and their

surfaces exist when unperceived, and human vision has evolved to describe their objective properties.

Bayesian theories of vision assume object permanence. They model object perception as a process of statistical estimation of object properties, such as surface shape and reflectance, that exist when unperceived. As Alan Yuille and Heinrich Bülthoff put it, "We define vision as perceptual inference, the estimation of scene properties from an image or sequence of images . . . " (Yuille and Bülthoff, 1996).

There is a long and interesting history of debate about which properties of objects exist when unperceived. Shape, size, and position usually make the list. Others, such as taste and color, often do not. Democritus, a contemporary of Socrates, famously claimed, "by convention sweet and by convention bitter, by convention hot, by convention cold, by convention color; but in reality atoms and void" (Taylor, 1999).

Locke proposed that "primary qualities" of objects, such as "bulk, figure, or motion" exist when unperceived, but that "secondary properties" of objects, such as "colors and smells" do not. He then claimed that ". . . the ideas of primary qualities of bodies are resemblances of them, and their patterns do really exist in the bodies themselves, but the ideas produced in us by these secondary qualities have no resemblance of them at all" (Locke, 1690).

Philosophical and scientific debate continues to this day on whether properties such as color exist when unperceived (Byrne and Hilbert, 2003; Hoffman, 2006). But object permanence, certainly regarding shape and position, is so deeply assumed by the scientific literature in the fields of psychophysics and computational perception that it is rarely discussed.

It is also assumed in the scientific study of consciousness and the mind-body problem. Here the widely acknowledged failure to create a plausible theory forces reflection on basic assumptions, including object permanence. But few researchers in fact give it up. To the contrary, the accepted view is that aspects of neural dynamics—from quantum-gravity induced collapses of wavefunctions at microtubules (Hameroff, 1998) to informational properties of re-entrant thalamo-cortical loops (Tononi, 2004)—cause, or give rise to, or are identical to, consciousness. As Colin McGinn puts it, "we know that brains are the *de facto* causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so" (McGinn, 1989).

EVOLUTION AND PERCEPTION

The human mind is predisposed from early childhood to assume object permanence, to assume that objects have shapes and positions in space even when the objects and space are unperceived. It is reasonable to ask whether this assumption is a genuine insight into the nature of objective reality, or simply a habit that is perhaps useful but not necessarily insightful.

We can look to evolution for an answer. If we assume that our perceptual and cognitive capacities have been shaped, at least in part, by natural selection, then we can use formal models of evolution, such as evolutionary game theory (Lieberman et al., 2005; Nowak, 2006) and genetic algorithms (Mitchell, 1998), to explore if, and under what circumstances, natural selection favors perceptual representations that are genuine insights into the true nature of the objective world.

Evaluating object permanence on evolutionary grounds might seem quixotic, or at least unfair, given that we just noted that evolutionary theory, as it's standardly described, assumes object permanence (e.g., of DNA and the physical bodies of organisms). How then could one possibly use evolutionary theory to test what it assumes to be true?

However, Richard Dawkins and others have observed that the core of evolution by natural selection is an abstract algorithm with three key components: variation, selection, and retention (Dennett, 1995; Blackmore, 1999). This abstract algorithm constitutes a "universal Darwinism" that need not assume object permanence and can be profitably applied in many contexts beyond biological evolution. Thus, it is possible, without begging the question, to use formal models of evolution by natural selection to explore whether object permanence is an insight or not.

Jerry Fodor has criticized the theory of natural selection itself, arguing, for instance, that it impales itself with an intensional fallacy, viz., inferring from the premise that "evolution is a process in which creatures with adaptive traits are selected" to the conclusion that "evolution is a process in which creatures are selected for their adaptive traits" (Fodor and Piattelli-Palmarini, 2010). However, Fodor's critique seems wide of the mark (Futuyma, 2010) and the evidence for evolution by natural selection is overwhelming (Coyne, 2009; Dawkins, 2009).

What, then, do we find when we explore the evolution of perception using evolutionary games and genetic algorithms? The standard answer, at least among vision scientists, is that we should find that natural selection favors veridical perceptions, i.e., perceptions that accurately represent objective properties of the external world that exist when unperceived. Steven Palmer, for instance, in a standard graduate-level textbook, states that "Evolutionarily speaking, visual perception is useful only if it is reasonably accurate . . . Indeed, vision is useful precisely because it is so accurate. By and large, *what you see is what you get.* When this is true, we have what is called **veridical perception** . . . perception that is consistent with the actual state of affairs in the environment. This is almost always the case with vision . . . " (Palmer, 1999).

The argument, roughly, is that those of our predecessors whose perceptions were more veridical had a competitive advantage over those whose perceptions were less veridical. Thus, the genes that coded for more veridical perceptions were more likely to propagate to the next generation. We are, with good probability, the offspring of those who, in each succeeding generation, perceived more truly, and thus we can be confident that our own perceptions are, in the normal case, veridical.

The conclusion that natural selection favors veridical perceptions is central to current Bayesian models of perception, in which perceptual systems use Bayesian inference to estimate true properties of the objective world, properties such as shape, position, motion, and reflectance (Knill and Richards, 1996; Geisler and Diehl, 2003). Objects exist and have these properties when unperceived, and the function of perception is to accurately estimate pre-existing properties.

However, when we actually study the evolution of perception using Monte Carlo simulations of evolutionary games and genetic algorithms, we find that natural selection does not, in general, favor perceptions that are true reports of objective properties of the environment. Instead, it generally favors perceptual strategies that are tuned to fitness (Mark et al., 2010; Hoffman et al., 2013; Marion, 2013; Mark, 2013).

Why? Several principles emerge from the simulations. First, there is no free information. For every bit of information one obtains about the external world, one must pay a price in energy, e.g., in calories expended to obtain, process and retain that information. And for every calorie expended in perception, one must go out and kill something and eat it to get that calorie. So natural selection tends to favor perceptual systems that, *ceteris paribus*, use fewer calories. One way to use fewer calories is to see less truth, especially truth that is not informative about fitness.

Second, for every bit of information one obtains about the external world, one must pay a price in time. More information requires, in general, more time to obtain and process. But in the real world where predators are on the prowl and prey must be wary, the race is often to the swift. It is the slower gazelle that becomes lunch for the swifter cheetah. So natural selection tends to favor perceptual systems that, *ceteris paribus*, take less time. One way to take less time is, again, to see less truth, especially truth that is not informative about fitness.

Third, in a world where organisms are adapted to niches and require homeostatic mechanisms, the fitness functions guiding their evolution are generally not monotonic functions of structures or quantities in the world. Too much salt or too little can be devastating; something in between is just right for fitness. The same goldilocks principle can hold for water, altitude, humidity, and so on. In these cases, perceptions that are tuned to fitness are *ipso facto* not tuned to the true structure of the world, because the two are not monotonically related; knowing the truth is not just irrelevant, it can be inimical, to fitness.

Fourth, in the generic case where noise and uncertainty are endemic to the perceptual process, a strategy that estimates a true state of the world and then uses the utility associated to that state to govern its decisions must throw away valuable information about utility. It will in general be driven to extinction by a strategy that does not estimate the true state of the world, and instead uses all the information about utility (Marion, 2013).

Fifth, more complex perceptual systems are more difficult to evolve. Monte Carlo simulations of genetic algorithms show that there is a combinatorial explosion in the complexity of the search required to evolve more complex perceptual systems. This combinatorial explosion itself is a selection pressure toward simpler perceptual systems.

In short, natural selection does not favor perceptual systems that see the truth in whole or in part. Instead, it favors perceptions that are fast, cheap, and tailored to guide behaviors needed to survive and reproduce. Perception is not about truth, it's about having kids. Genes coding for perceptual systems that increase the probability of having kids are *ipso facto* the genes that are more likely to code for perceptual systems in the next generation.

THE INTERFACE THEORY OF PERCEPTION

Natural selection favors perceptions that are useful though not true. This might seem counterintuitive, even to experts in perception. Palmer, for instance, in the quote above, makes the plausible claim that "vision is useful precisely because it is so accurate" (Palmer, 1999). Geisler and Diehl agree, taking it as obvious that "In general, (perceptual) estimates that are nearer the truth have greater utility than those that are wide of the mark" (Geisler and Diehl, 2002). Feldman also takes it as obvious that "it is clearly desirable (say from an evolutionary point of view) for an organism to achieve veridical percepts of the world" (Feldman, 2013). Knill and Richards concur that vision "... involves the evolution of an organism's visual system to match the structure of the world ... " (Knill and Richards, 1996).

This assumption that perceptions are useful to the extent that they are true is *prima facie* plausible, and it comports well with the assumption of object permanence. For if our perceptions report to us a three-dimensional world containing objects with specific shapes and positions, and if these perceptual reports have been shaped by evolution to be true, then we can be confident that those objects really do, in the normal case, exist and have their positions and shapes even when unperceived.

So we find it plausible that perceptions are useful only if true, and we find it deeply counterintuitive to think otherwise. But studies with evolutionary games and genetic algorithms flatly contradict this deeply held assumption. Clearly our intuitions need a little help here. How can we try to understand perceptions that are useful but not true?

Fortunately, developments in computer technology have provided a convenient and helpful metaphor: the desktop of a windows interface (Hoffman, 1998, 2009, 2011, 2012, 2013; Mausfeld, 2002; Koenderink, 2011a; Hoffman and Singh, 2012; Singh and Hoffman, 2013). Suppose you are editing a text file and that the icon for that file is a blue rectangle sitting in the lower left corner of the desktop. If you click on that icon you can open the file and revise its text. If you drag that icon to the trash, you can delete the file. If you drag it to the icon for an external hard drive, you can create a backup of the file. So the icon is quite useful.

But is it *true*? Well, the only visible properties of the icon are its position, shape, and color. Do these properties of the icon resemble the true properties of the file? Clearly not. The file is not blue or rectangular, and it's probably not in the lower left corner of the computer. Indeed, files don't have a color or shape, and needn't have a well-defined position (e.g., the bits of the file could be spread widely over memory). So to even ask if the properties of the icon are true is to make a category error, and to completely misunderstand the purpose of the interface. One can reasonably ask whether the icon is usefully related to the file, but not whether it truly resembles the file.

Indeed, a critical function of the interface is to *hide* the truth. Most computer users don't want to see the complexity of the integrated circuits, voltages, and magnetic fields that are busy behind the scenes when they edit a file. If they had to deal with that complexity, they might never finish their work on the file. So the interface is designed to allow the user to interact effectively with the computer while remaining largely ignorant of its true architecture.

Ignorant, also, of its true causal structure. When a user drags a file icon to an icon of an external drive, it looks obvious that the movement of the file icon to the drive icon *causes* the file to be copied. But this is just a useful fiction. The movement of the file icon causes nothing in the computer. It simply serves to guide the user's operation of a mouse, triggering a complex chain of causal events inside the computer, completely hidden from the user. Forcing the user to see the true causal chain would be an impediment, not a help.

Turning now to apply the interface metaphor to human perception, the idea is that natural selection has not shaped our perceptions to be insights into the true structure and causal nature of objective reality, but has instead shaped our perceptions to be a species-specific user interface, fashioned to guide the behaviors that we need to survive and reproduce. Space and time are the desktop of our perceptual interface, and three-dimensional objects are icons on that desktop.

Our interface gives the impression that it reveals true cause and effect relations. When one billiard ball hits a second, it certainly looks as though the first causes the second to careen away. But this appearance of cause and effect is simply a useful fiction, just as it is for the icons on the computer desktop.

There is an obvious rejoinder: "If that cobra is just an icon of your interface with no causal powers, why don't you grab it by the tail?" The answer is straightforward: "I don't grab the cobra for the same reason I don't carelessly drag my file icon to the trash—I could lose a lot of work. I don't take my icons *literally*: The file, unlike its icon, is not literally blue or rectangular. But I do take my icons *seriously*."

Similarly, evolution has shaped us with a species-specific interface whose icons we must take seriously. If there is a cliff, don't step over. If there is a cobra, don't grab its tail. Natural selection has endowed us with perceptions that function to guide adaptive behaviors, and we ignore them at our own peril.

But, given that we must take our perceptions seriously, it does not follow that we must take them literally. Such an inference is natural, in the sense that most of us, even the brightest, make it automatically. When Samuel Johnson heard Berkeley's theory that "To be is to be perceived" he kicked a stone and said, "I refute it *thus*!" (Boswell, 1986) Johnson observed that one must take the stone seriously or risk injury. From this Johnson concluded that one must take the stone literally. But this inference is fallacious.

One might object that there still is an important sense in which our perceptual icon of, say, a cobra does resemble the true objective reality: The consequences for an observer of grabbing the tail of the cobra are precisely the consequences that would obtain if the objective reality were in fact a cobra. Perceptions and internal information-bearing structures are useful for fitness-preserving or enhancing behavior because there is some mutual information between the predicted utility of a behavior (like escaping) and its actual utility. If there's no mutual information and no mechanism for increasing mutual information, fitness is low and stays that way. Here we use mutual information in the sense of standard information theory (Cover and Thomas, 2006).

This point is well-taken. Our perceptual icons do give us genuine information about fitness, and fitness can be considered an aspect of objective reality. Indeed, in Gibson's ecological theory of perception, our perceptions primarily resonate to "affordances," those aspects of the objective world that have important consequences for fitness (Gibson, 1979). While we disagree with Gibon's direct realism and denial of information processing in perception, we agree with his emphasis on the tuning of perception to fitness.

So we must clarify the relationship between truth and fitness. In evolutionary theory it is as follows. If W denotes the objective world then, for a fixed organism, state, and action, we can think of a fitness function to be a function $f{:}W \rightarrow [0,1]$, which assigns to each state w of W a fitness value $f(w)$. If, for instance, the organism is a hungry cheetah and the action is eating, then f might assign a high fitness value to world state w in which fresh raw meat is available; but if the organism is a hungry cow then f might assign a low fitness value to the same state w.

If the true probabilities of states in the world are given by a probability measure m on W, then one can define a new probability measure mf on W, where for any event A of W, $mf(A)$ is simply the integral of f over A with respect to m; mf must of course be normalized so that $mf(W) = 1$.

And here is the key point. A perceptual system that is tuned to maximize the mutual information with m will not, in general, maximize mutual information with mf (Cover and Thomas, 2006). Being tuned to truth, i.e., maximizing mutual information with m, is not the same as being tuned to fitness, i.e., maximizing mutual information with mf. Indeed, depending on the fitness function f, a perceptual system tuned to truth might carry little or no information about fitness, and vice versa. It is in this sense that the interface theory of perception claims that our perceptions are tuned to fitness rather than truth.

There is another rejoinder: "The interface metaphor is nothing new. Physicists have told us for more than a century that solid objects are really mostly empty space. So an apparently solid stone isn't the true reality, but its atoms and subatomic particles are." Physicists have indeed said this since Rutherford published his theory of the atomic nucleus in 1911 (Rutherford, 1911). But the interface metaphor says something more radical. It says that space and time themselves are just a desktop, and that anything in space and time, including atoms and subatomic particles, are themselves simply icons. It's not just the moon that isn't there when one doesn't look, it's the atoms, leptons and quarks themselves that aren't there. Object permanence fails for microscopic objects just as it does for macroscopic.

This claim is, to contemporary sensibilities, radical. But there is a perspective on the intellectual evolution of humanity over the last few centuries for which the interface theory seems a natural next step. According to this perspective, humanity has gradually been letting go of the false belief that the way *H. sapiens* sees the world is an insight into objective reality.

Many ancient cultures, including the pre-Socratic Greeks, believed the world was flat, for the obvious reason that it looks that way. Aristotle became persuaded, on empirical grounds, that the earth is spherical, and this view gradually spread to other cultures. Reality, we learned, departed in important respects from some of our perceptions.

But then a geocentric model of the universe, in which the earth is at the center and everything revolves around it, still held sway.

Why? Because that's the way things look to our unaided perceptions. The earth looks like it's not moving, and the sun, moon, planets, and stars look like they circle a stationary earth. Not until the work of Copernicus and Kepler did we recognize that once again reality differs, in important respects, from our perceptions. This was difficult to swallow. Galileo was forced to recant in the Vatican basement, and Giordano Bruno was burned at the stake. But we finally, and painfully, accepted the mismatch between our perceptions and certain aspects of reality.

The interface theory entails that these first two steps were mere warm up. The next step in the intellectual history of *H. sapiens* is a big one. We must recognize that *all* of our perceptions of space, time and objects no more reflect reality than does our perception of a flat earth. It's not just this or that aspect of our perceptions that must be corrected, it is *the entire framework* of a space-time containing objects, the fundamental organization of our perceptual systems, that must be recognized as a mere species-specific mode of perception rather than an insight into objective reality.

By this time it should be clear that, if the arguments given here are sound, then the current Bayesian models of object perception need more than tinkering around the edges, they need fundamental transformation. And this transformation will necessarily have ramifications for scientific questions well-beyond the confines of computational models of object perception.

One example is the mind-body problem. A theory in which objects and space-time do not exist unperceived and do not have causal powers, cannot propose that neurons—which by hypothesis do not exist unperceived and do not have causal powers—cause any of our behaviors or conscious experiences. This is so contrary to contemporary thought in this field that it is likely to be taken as a *reductio* of the view rather than as an alternative direction of inquiry for a field that has yet to construct a plausible theory.

DEFINITION OF CONSCIOUS AGENTS

If our reasoning has been sound, then space-time and three-dimensional objects have no causal powers and do not exist unperceived. Therefore, we need a fundamentally new foundation from which to construct a theory of objects. Here we explore the possibility that consciousness is that new foundation, and seek a mathematically precise theory. The idea is that a theory of objects requires, first, a theory of subjects.

This is, of course, a non-trivial endeavor. Frank Wilczek, when discussing the interpretation of quantum theory, said, "The relevant literature is famously contentious and obscure. I believe it will remain so until someone constructs, within the formalism of quantum mechanics, an "observer," that is, a model entity whose states correspond to a recognizable caricature of conscious awareness . . . That is a formidable project, extending well-beyond what is conventionally considered physics" (Wilczek, 2006).

The approach we take toward constructing a theory of consciousness is similar to the approach Alan Turing took toward constructing a theory of computation. Turing proposed a simple but rigorous formalism, now called the *Turing machine* (Turing, 1937; Herken, 1988). It consists of six components: (1) a finite set of states, (2) a finite set of symbols, (3) a special blank symbol, (4) a finite set of input symbols, (5) a start state, (6) a set of halt states, and (7) a finite set of simple transition rules (Hopcroft et al., 2006).

Turing and others then conjectured that a function is algorithmically computable if and only if it is computable by a Turing machine. This "Church-Turing Thesis" can't be proven, but it could in principle be falsified by a counterexample, e.g., by some example of a procedure that everyone agreed was computable but for which no Turing machine existed. No counterexample has yet been found, and the Church-Turing thesis is considered secure, even definitional.

Similarly, to construct a theory of consciousness we propose a simple but rigorous formalism called a *conscious agent*, consisting of six components. We then state the *conscious agent thesis*, which claims that every property of consciousness can be represented by some property of a conscious agent or system of interacting conscious agents. The hope is to start with a small and simple set of definitions and assumptions, and then to have a complete theory of consciousness arise as a series of theorems and proofs (or simulations, when complexity precludes proof). We want a theory of consciousness *qua consciousness*, i.e., of consciousness on its own terms, not as something derivative or emergent from a prior physical world.

No doubt this approach will strike many as *prima facie* absurd. It is a commonplace in cognitive neuroscience, for instance, that most of our mental processes are *unconscious* processes (Bargh and Morsella, 2008). The standard account holds that well more than 90% of mental processes proceed without conscious awareness. Therefore, the proposal that consciousness is fundamental is, to contemporary thought, an amusing anachronism not worth serious consideration.

This critique is apt. It's clear from many experiments that each of us is indeed unaware of most of the mental processes underlying our actions and conscious perceptions. But this is no surprise, given the interface theory of perception. Our perceptual interfaces have been shaped by natural selection to guide, quickly and cheaply, behaviors that are adaptive in our niche. They have not been shaped to provide exhaustive insights into truth. In consequence, our perceptions have endogenous limits to the range and complexity of their representations. It was not adaptive to be aware of most of our mental processing, just as it was not adaptive to be aware of how our kidneys filter blood.

We must be careful not to assume that limitations of our species-specific perceptions are insights into the true nature of reality. My friend's mind is not directly conscious to me, but that does not entail that my friend is unconscious. Similarly, most of my mental processes are not directly conscious to me, but that does not entail that they are unconscious. Our perceptual systems have finite capacity, and will therefore inevitably simplify and omit. We are well-advised not to mistake our omissions and simplifications for insights into reality.

There are of course many other critiques of an approach that takes consciousness to be fundamental: How can such an approach explain matter, the fundamental forces, the Big Bang, the genesis and structure of space-time, the laws of physics, evolution by natural selection, and the many neural correlates of consciousness? These are non-trivial challenges that must be faced by the theory of conscious agents. But for the moment we

will postpone them and develop the theory of conscious agents itself.

Conscious agent is a technical term, with a precise mathematical definition that will be presented shortly. To understand the technical term, it can be helpful to have some intuitions that motivate the definition. The intuitions are just intuitions, and if they don't help they can be dropped. What does the heavy lifting is the definition itself.

A key intuition is that consciousness involves three processes: *perception*, *decision*, and *action*.

In the process of perception, a conscious agent interacts with the world and, in consequence, has conscious experiences.

In the process of decision, a conscious agent chooses what actions to take based on the conscious experiences it has.

In the process of action, the conscious agent interacts with the world in light of the decision it has taken, and affects the state of the world.

Another intuition is that we want to avoid unnecessarily restrictive assumptions in constructing a theory of consciousness. Our conscious visual experience of nearby space, for instance, is approximately Euclidean. But it would be an unnecessary restriction to require that *all* of our perceptual experiences be represented by Euclidean spaces.

However it does seem necessary to discuss the *probability* of having a conscious experience, of making a particular decision, and of making a particular change in the world through action. Thus, it seems necessary to assume that we can represent the world, our conscious experiences, and our possible actions with probability spaces.

We also want to avoid unnecessarily restrictive assumptions about the *processes* of perception, decision, and action. We might find, for instance, that a particular decision process maximizes expected utility, or minimizes expected risk, or builds an explicit model of the self. But it would be an unnecessary restriction to require this of all decisions.

However, when considering the processes of perception, decision and action, it does seem necessary to discuss *conditional probability*. It seems necessary, for instance, to discuss the conditional probability of deciding to take a specific action given a specific conscious experience, the conditional probability of a particular change in the world given that a specific action is taken, and the conditional probability of a specific conscious experience given a specific state of the world.

A general way to model such conditional probabilities is by the mathematical formalism of Markovian kernels (Revuz, 1984). One can think of a Markovian kernel as simply an indexed list of probability measures. In the case of perception, for instance, a Markovian kernel might specify that if the state of the world is w_1, then here is a list of the probabilities for the various conscious experiences that might result, but if the state of the world is w_2, then here is a different list of the probabilities for the various conscious experiences that might result, and so on for all the possible states of the world. A Markovian kernel on a finite set of states can be written as matrix in which the entries in each row sum to 1.

A Markovian kernel can also be thought of as an *information channel*. Cover and Thomas, for instance, define "a discrete channel to be a system consisting of an input alphabet X and output alphabet Y and a probability transition matrix $p(x|y)$ that expresses the probability of observing the output symbol y given that we send the symbol x" (Cover and Thomas, 2006). Thus, a discrete channel is simply a Markovian kernel.

So, each time a conscious agent interacts with the world and, in consequence, has a conscious experience, we can think of this interaction as a message being passed from the world to the conscious agent over a channel. Similarly, each time the conscious agent has a conscious experience and, in consequence, decides on an action to take, we can think of this decision as a message being passed over a channel within the conscious agent itself. And when the conscious agent then takes the action and, in consequence, alters the state of the world, we can think of this as a message being passed from the conscious agent to the world over a channel. In the discrete case, we can keep track of the number of times each channel is used. That is, we can count the number of messages that are passed over each channel. Assuming that all three channels (perception, decision, action) all work in lock step, we can use one counter, N, to keep track of the number of messages that are passed.

These are some of the intuitions that underlie the definition of conscious agent that we will present. These intuitions can be represented pictorially in a diagram, as shown in **Figure 1**. The channel P transmits messages from the world W, leading to conscious experiences X. The channel D transmits messages from X, leading to actions G. The channel A transmits messages from G that are received as new states of W. The counter N is an integer that keeps track of the number of messages that are passed on each channel.

In what follows we will be using the notion of a measurable space. Recall that a measurable space, $(X, \mathbf{X})$, is a set X together with a collection $\mathbf{X}$ of subsets of X, called *events*, that satisfies three properties: (1) X is in $\mathbf{X}$; (2) $\mathbf{X}$ is closed under complement (i.e., if a set A is in $\mathbf{X}$ then the complement of A is also in $\mathbf{X}$); and (3) $\mathbf{X}$ is closed under countable union. The collection of events $\mathbf{X}$ is a σ-algebra (Athreya and Lahiri, 2006). A probability measure assigns a probability to each event in $\mathbf{X}$.

With these intuitions, we now present the formal definition of a conscious agent where, for the moment, we simply assume that the world is a measurable space $(W, \mathbf{W})$.

Definition 1. A *conscious agent*, C, is a six-tuple

$$C = ((X, \mathbf{X}), (G, \mathbf{G}), P, D, A, N)), \tag{1}$$

where:

(1) $(X, \mathbf{X})$ and $(G, \mathbf{G})$ are measurable spaces;
(2) $P : W \times \mathbf{X} \to [0, 1]$, D: $X \times \mathbf{G} \to [0, 1]$, A: $G \times \mathbf{W} \to [0, 1]$ are Markovian kernels; and
(3) N is an integer.

For convenience we will often write a conscious agent C as

$$C = (X, G, P, D, A, N), \tag{2}$$

omitting the σ-algebras.

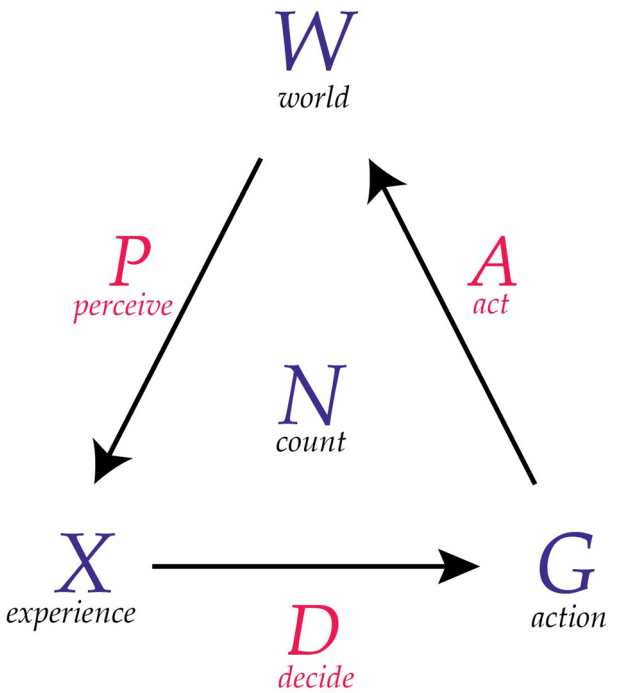


FIGURE 1 | A diagram of a conscious agent. A conscious agent has six components as illustrated here. The maps *P*, *D*, and *A* can be thought of as communication channels.

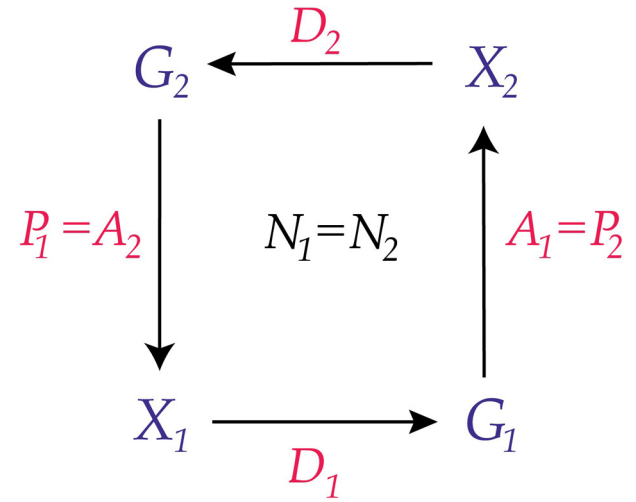


FIGURE 2 | Two conscious agents, C_1 and C_2. Each is part of the world *W* for the other conscious agent. The lower part of the diagram represents C_1 and the upper part represents C_2. This creates an *undirected combination* of C_1 and C_2, a concept we define in section The Combination Problem.

Given that *P*, *D*, and *A* are channels, each has a *channel capacity*, viz., a highest rate of bits per channel use, at which information can be sent across the channel with arbitrarily low chance of error (Cover and Thomas, 2006).

The formal structure of a conscious agent, like that of a Turing machine, is simple. Nevertheless, we will propose, in the next section, a "conscious-agent thesis" which, like the Church-Turing thesis, claims wide application for the formalism.

CONSCIOUS REALISM

One glaring feature of the definition of a conscious agent is that it involves the world, *W*. This is not an arbitrary choice; *W* is required to define the perceptual map *P* and action map *A* of the conscious agent.

This raises the question: What is the world? If we take it to be the space-time world of physics, then the formalism of conscious agents is dualistic, with some components (e.g., *X* and *G*) referring to consciousness and another, viz., *W*, referring to a physical world.

We want a non-dualistic theory. Indeed, the monism we want takes consciousness to be fundamental. The formalism of conscious agents provides a precise way to state this monism.

Hypothesis 1. *Conscious realism*: The world *W* consists entirely of conscious agents.

Conscious realism is a precise hypothesis that, of course, might be precisely wrong. We can explore its theoretical implications in the normal scientific manner to see if they comport well with existing data and theories, and make predictions that are novel, interesting and testable.

TWO CONSCIOUS AGENTS

Conscious realism can be expressed mathematically in a simple form. Consider the elementary case, in which the world *W* of one conscious agent,

$$C_1 = (X_1, G_1, P_1, D_1, A_1, N_1), \tag{3}$$

contains just C_1 and one other agent,

$$C_2 = (X_2, G_2, P_2, D_2, A_2, N_2), \tag{4}$$

and vice versa. This is illustrated in **Figure 2**.

Observe that although *W* is the world it cannot properly be called, in this example, the *external* world of C_1 or of C_2 because C_1 and C_2 are each part of *W*. This construction of *W* requires the compatibility conditions

$$P_1 = A_2, \tag{5}$$

$$P_2 = A_1, \tag{6}$$

$$N_1 = N_2. \tag{7}$$

These conditions mean that the perceptions of one conscious agent are identical to the actions of the other, and that their counters are synchronized. To understand this, recall that we can think of P_1, P_2, A_1, and A_2 as *information channels*. So interpreted, conditions (5) and (6) state that the action channel of one agent is the same information channel as the perception channel of the other agent. Condition (7) states that the channels of both agents operate in synchrony.

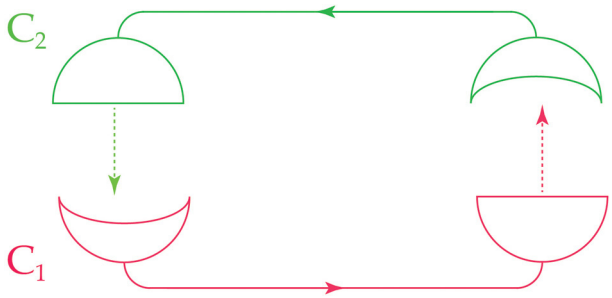


FIGURE 3 | Two adjacent conscious agents, C_1 and C_2. Each agent receives messages from the other (indicated by the concave receivers) and sends messages to the other (indicated by the semicircular transmitters). Arrows show the direction of information flow.

If two conscious agents C_1 and C_2 satisfy the commuting diagram of **Figure 2**, then we say that they are *joined* or *adjacent*: the experiences and actions of C_1 affect the probabilities of experiences and actions for C_2 and vice versa. **Figure 3** illustrates the ideas so far.

We can simplify the diagrams further and simply write C_1—C_2 to represent two adjacent conscious agents.

THREE CONSCIOUS AGENTS

Any number of conscious agents can be joined. Consider the case of three conscious agents,

$$C_i = (X_i, G_i, P_i, D_i, A_i, N_i),\ i = 1, 2, 3. \tag{8}$$

This is illustrated in **Figure 4**, and compactly in **Figure 5**.

Because C_1 interacts with C_2 and C_3, its perceptions are affected by both C_2 and C_3. Thus, its perception kernel, P_1, must reflect the inputs of C_2 and C_3. We write it as follows:

$$P_1 = P_{12} \otimes P_{13} : (G_2 \times G_3) \times \mathbf{X}_1 \to [0, 1], \tag{9}$$

where

$$\mathbf{X}_1 = \boldsymbol{\sigma}(\mathbf{X_{12}} \times \mathbf{X_{13}}), \tag{10}$$

$(X_{12}, \mathbf{X}_{12})$ is the measurable space of perceptions that C_1 can receive from C_2, and $(X_{13}, \mathbf{X}_{13})$ is the measurable space of perceptions that C_1 can receive from C_3, and $\boldsymbol{\sigma}(\mathbf{X}_{12} \times \mathbf{X}_{13})$ denotes the σ-algebra generated by the Cartesian product of $\mathbf{X}_{12}$ and $\mathbf{X}_{13}$. The tensor product P_1 of (9) is given by the formula

$$P_1\left((g_2, g_3), (x_{12}, x_{13})\right) = P_{12}(g_2, x_{12})P_{13}(g_3, x_{13}), \tag{11}$$

where $g_2 \in G_2$, $g_3 \in G_3$, $x_{12} \in \mathbf{X}_{12}$, and $x_{13} \in \mathbf{X}_{13}$. Note that (11) allows that the perceptions that C_1 gets from C_2 could be entirely different from those it gets from C_3, and expresses the probabilistic independence of these perceptual inputs. In general, $X_{\mathbf{12}}$ need not be identical to $X_{\mathbf{13}}$, since the kinds of perceptions that C_1 can receive from C_2 need not be the same as the kinds of perceptions that C_1 can receive from C_3.

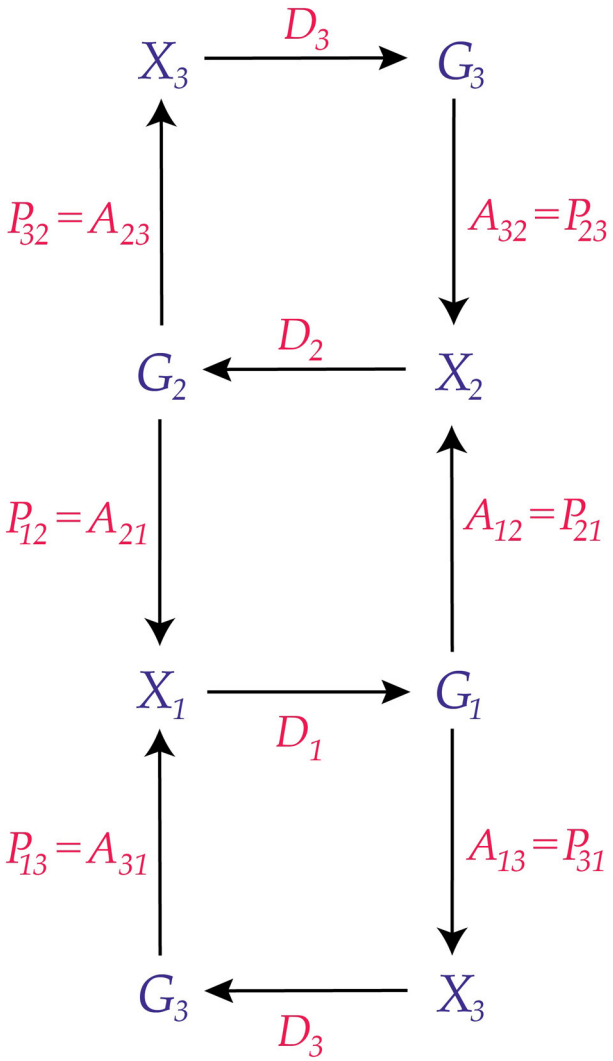


FIGURE 4 | Three adjacent conscious agents. The third agent is replicated at the **top** and **bottom** of the diagram for visual simplicity.

Because C_1 interacts with C_2 and C_3, its actions affect both. However, the way C_1 acts on C_2 might differ from how it acts on C_3, and the definition of its action kernel, A_1, must allow for this difference of action. Therefore, we define the action kernel, A_1, to be the tensor product

$$A_1 = A_{12} \otimes A_{13} : G_1 \times \boldsymbol{\sigma}(\mathbf{X}_2 \times \mathbf{X}_3) \to [0, 1], \tag{12}$$

where

$$G_1 = G_{12} \times G_{13}, \tag{13}$$

$(G_{12}, \mathbf{G}_{12})$ is the measurable space of actions that C_1 can take on C_2, and $(G_{13}, \mathbf{G}_{13})$ is the measurable space of actions that C_1 can take on C_3.

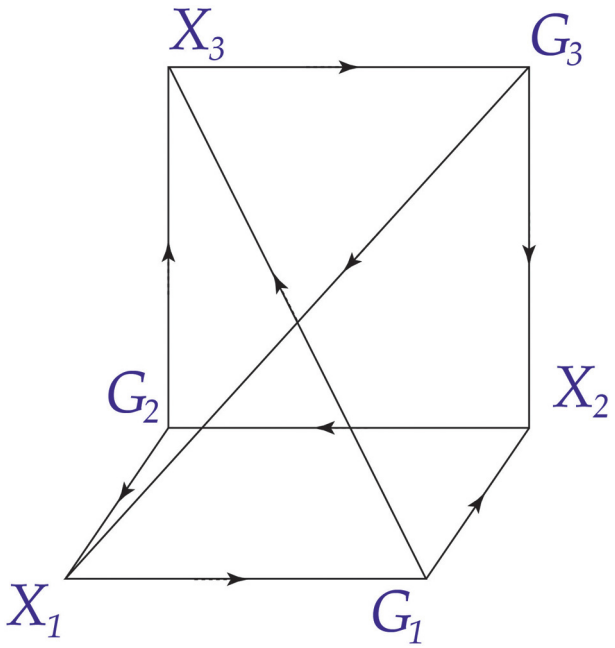


FIGURE 5 | Three adjacent conscious agents. This is a compact representation of the diagram in **Figure 4**.

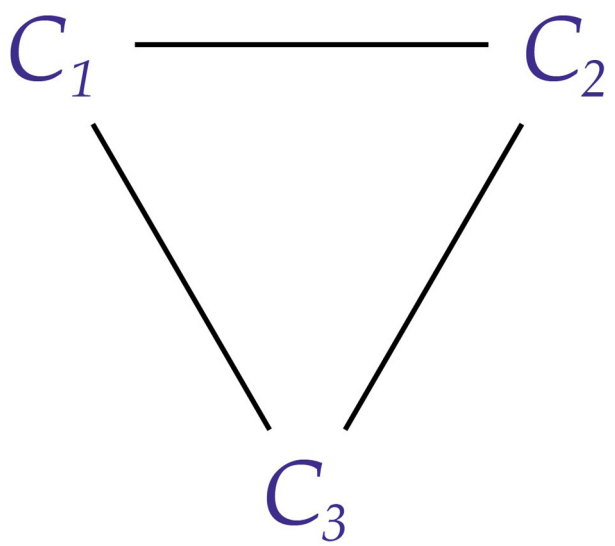


FIGURE 6 | Three conscious agents whose graph is complete.

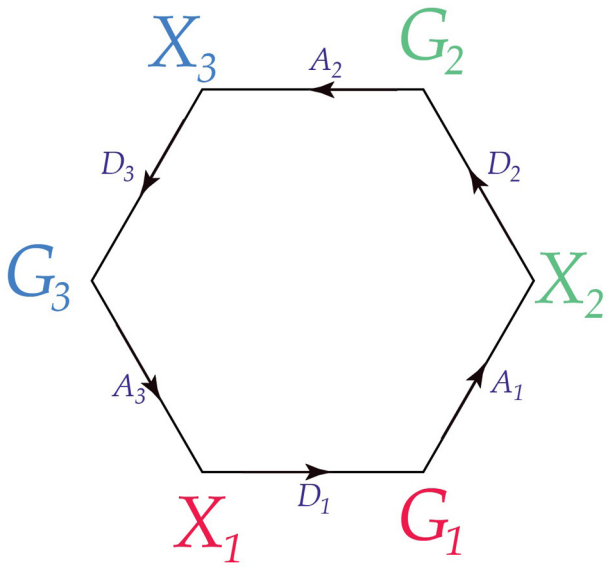


FIGURE 7 | Three conscious agents with directed joins. Here we assume $A_1 = P_2$, $A_2 = P_3$, and $A_3 = P_1$.

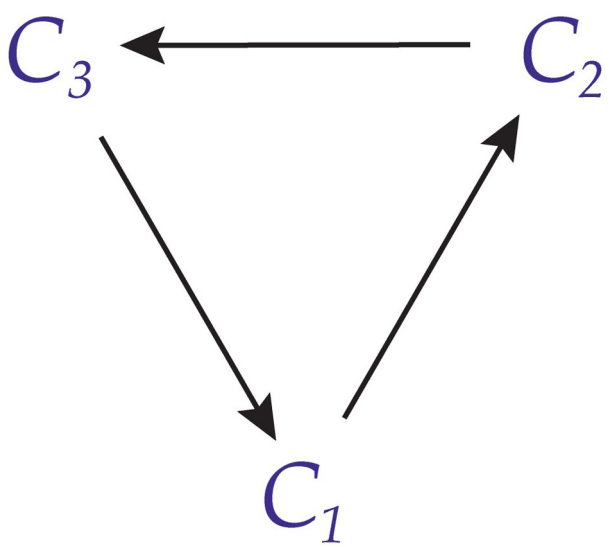


FIGURE 8 | Simplified graph of three conscious agents with directed joins.

In this situation, the three conscious agents have the property that every pair is adjacent; we say that the graph of the three agents is *complete*. This is illustrated in **Figure 6**.

So far we have considered joins that are undirected, in the sense that if C_1 sends a message to C_2 then C_2 sends a message to C_1. However, it is also possible for conscious agents to have *directed joins*. This is illustrated in **Figure 7**. In this case, C_1 sends a message to C_2 and receives a message from C_3, but receives no message from C_2 and sends no message to C_3. Similar remarks hold, *mutatis mutandis*, for C_2 and C_3.

Figure 7 can be simplified as shown in **Figure 8**.

Directed joins can model the standard situation in visual perception, in which there are multiple levels of visual representations, one level building on others below it. For instance, at one level there could be the construction of 2D motions based on a solution to the correspondence problem; at the next level there could be a computation of 3D structure from motion, based on the 2D motions computed at the earlier level (Marr, 1982). So

an agent C_1 might solve the correspondence problem and pass its solution to C_2, which solves the structure-from-motion problem, and then passes its solution to C_3, which does object recognition.

We can join any number of conscious agents into any multigraph, where nodes denote agents and edges denote directed or undirected joins between agents (Chartrand and Ping, 2012). The nodes can have any finite degree, i.e., any finite number of edges. As a special case, conscious agents can be joined to form deterministic or non-deterministic cellular automata (Ceccherini-Silberstein and Coornaert, 2010) and universal Turing machines (Cook, 2004).

DYNAMICS OF TWO CONSCIOUS AGENTS

Two conscious agents

$$C_1 = (X_1, G_1, P_1, D_1, A_1, N_1), \tag{14}$$

and

$$C_2 = (X_2, G_2, P_2, D_2, A_2, N_2), \tag{15}$$

can be joined, as illustrated in **Figure 2**, to form a dynamical system. Here we discuss basic properties of this dynamics.

The state space, E, of the dynamics is $E = X_1 \times G_1 \times X_2 \times G_2$, with product σ-algebra $\mathbf{E}$. The idea is that for the current step, $t \in N$, of the dynamics, the state can be described by the vector $(x_1(t), g_1(t), x_2(t), g_2(t))$, and based on this state four actions happen simultaneously: (1) agent C_1 experiences the perception $x_1(t) \in X_1$ and decides, according to D_1, on a specific action $g_1(t) \in G_1$ to take at step $t + 1$; (2) agent C_1, using A_1, takes the action $g_1(t) \in G_1$; (3) agent C_2 experiences the perception $x_2(t) \in X_2$ and decides, according to D_2, on a specific action $g_2(t) \in G_2$ to take at step $t + 1$; (4) agent C_2, using A_2, takes the action $g_2(t) \in G_2$.

Thus, the state evolves by a kernel

$$L : E \times \mathbf{E} \to [0, 1], \tag{16}$$

which is given, for state $e = (x_1(t), g_1(t), x_2(t), g_2(t)) \in E$ at time t and event $B \in \mathbf{E}$, comprised of a measurable set of states of the form $(x_1(t + 1), g_1(t + 1), x_2(t + 1), g_2(t + 1))$, by

$$L(e, B) = \int_B A_2(g_2(t), dx_1(t + 1))D_1(x_1(t), dg_1(t + 1))A_1(g_1(t), dx_2(t + 1))D_2(x_2(t), dg_2(t + 1)). \tag{17}$$

This is not kernel composition; it is simply multiplication of the four kernel values. The idea is that at each step of the dynamics each of the four kernels acts simultaneously and independently of the others to transition the state $(x_1(t), g_1(t), x_2(t), g_2(t))$ to the next state $(dx_1(t + 1), dg_1(t + 1), dx_2(t + 1), dg_2(t + 1))$.

FIRST EXAMPLE OF ASYMPTOTIC BEHAVIOR

For concreteness, consider the simplest possible case where (1) X_1, G_1, X_2, and G_2 each have only two states which, using Dirac notation, we denote $|0\rangle$ and $|1\rangle$, and (2) each of the kernels A_2, D_1, A_1, and D_2 is a 2×2 identity matrix.

There are total of $2^4 = 16$ possible states for the dynamics of the two agents, which we can write as $|0000\rangle$, $|0001\rangle$, $|0010\rangle$, ... $|1111\rangle$, where the leftmost digit is the state of X_1, the next digit the state of G_1, the next of X_2, and the rightmost of G_2.

The asymptotic (i.e., long-term) dynamics of these two conscious agents can be characterized by its absorbing sets and their periods. Recall that an absorbing set for such a dynamics is a smallest set of states that acts like a roach motel: once the dynamics enters the absorbing set it never leaves, and it forever cycles periodically through the states within that absorbing set. It is straightforward to verify that for the simple dynamics of conscious agents just described, the asymptotic behavior is as follows:

(1) $\{|0000\rangle\}$ is absorbing with period 1;
(2) $\{|1111\rangle\}$ is absorbing with period 1;
(3) $\{|0101\rangle, |1010\rangle\}$ is absorbing with period 2;
(4) $\{|0001\rangle, |1000\rangle, |0100\rangle, |0010\rangle\}$ is absorbing with period 4, and cycles in that order;
(5) $\{|0011\rangle, |1001\rangle, |1100\rangle, |0110\rangle\}$ is absorbing with period 4, and cycles in that order;
(6) $\{|0111\rangle, |1011\rangle, |1101\rangle, |1110\rangle\}$ is absorbing with period 4, and cycles in that order.

SECOND EXAMPLE OF ASYMPTOTIC BEHAVIOR

If we alter this dynamics by simply changing the kernel D_1 from an identity matrix to the matrix $D_1 = ((0, 1), (1, 0))$, then the asymptotic behavior changes to the following:

(1) $\{|0000\rangle, |0100\rangle, |0110\rangle, |0111\rangle, |1111\rangle, |1011\rangle, |1001\rangle, |1000\rangle\}$ is absorbing with period 8, and cycles in that order;
(2) $\{|0001\rangle, |1100\rangle, |0010\rangle, |0101\rangle, |1110\rangle, |0011\rangle, |1101\rangle, |1010\rangle\}$ is absorbing with period 8, and cycles in that order.

If instead of changing D_1 we changed D_2 (or A_1 or A_2) to ((0,1),(1,0)), we would get the same asymptotic behavior. Thus, in general, an asymptotic behavior corresponds to an equivalence class of interacting conscious agents.

The range of possible dynamics of pairs of conscious agents is huge, and grows as one increases the richness of the state space E and, therefore, the set of possible kernels. The possibilities increase as one considers dynamical systems of three or more conscious agents, with all the possible directed and undirected joins among them, forming countless connected multi-graphs or amenable groups.

With this brief introduction to the dynamics of conscious agents we are now in a position to state another key hypothesis.

Hypothesis 2. *Conscious-agent thesis*. Every property of consciousness can be represented by some property of a dynamical system of conscious agents.

THE COMBINATION PROBLEM

Conscious realism and the conscious-agent thesis are strong claims, and face a tough challenge: Any theory that claims consciousness is fundamental must solve the *combination problem* (Seager, 1995; Goff, 2009; Blamauer, 2011; Coleman, 2014). William Seager describes this as "the problem of explaining how

the myriad elements of 'atomic consciousness' can be combined into a new, complex and rich consciousness such as that we possess" (Seager, 1995).

William James saw the problem back in 1890: "Where the elemental units are supposed to be feelings, the case is in no wise altered. Take a hundred of them, shuffle them and pack them as close together as you can (whatever that may mean); still each remains the same feeling it always was, shut in its own skin, windowless, ignorant of what the other feelings are and mean. There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness belonging to the group as such should emerge. And this 101st feeling would be a totally new fact; the 100 original feelings might, by a curious physical law, be a signal for its creation, when they came together; but they would have no substantial identity with it, nor it with them, and one could never deduce the one from the others, or (in any intelligible sense) say that they evolved it. ... The private minds do not agglomerate into a higher compound mind" (James, 1890/2007).

There are really two combination problems. The first is the combination of phenomenal *experiences*, i.e., of qualia. For instance, one's taste experiences of salt, garlic, onion, basil and tomato are somehow combined into the novel taste experience of a delicious pasta sauce. What is the relationship between one's experiences of the ingredients and one's experience of the sauce?

The second problem is the combination of *subjects* of experiences. In the sauce example, a single subject experiences the ingredients and the sauce, so the problem is to combine experiences within a single subject. But how can we combine subjects themselves to create a new unified subject? Each subject has its point of view. How can different points of view be combined to give a new, single, point of view?

No rigorous theory has been given for combining phenomenal experiences, but there is hope. Sam Coleman, for instance, is optimistic but notes that "there will have to be some sort of qualitative blending or pooling among the qualities carried by each ultimate: if each ultimate's quality showed up as such in the macro-experience, it would lack the notable homogeneity of (e.g., color experience, and plausibly some mixing of basic qualities is required to obtain the qualities of macro-experience" (Coleman, 2014).

Likewise, no rigorous theory has been given for combining subjects. But here there is little hope. Thomas Nagel, for instance, says "Presumably the components out of which a point of view is constructed would not themselves have to have points of view" (Nagel, 1979). Coleman goes further, saying, "it is impossible to explain the generation of a macro-subject (like one of us) in terms of the assembly of micro-subjects, for, as I show, subjects cannot combine" (Coleman, 2014).

So at present there is the hopeful, but unsolved, problem of combining experiences and the hopeless problem of combining subjects.

The theory of conscious agents provides two ways to combine conscious agents: *undirected combinations* and *directed combinations*. We prove this, and then consider the implications for solving the problems of combining experiences and combining subjects.

Theorem 1. (*Undirected Join Theorem*.) An undirected join of two conscious agents creates a new conscious agent.

Proof. (*By construction*.) Let two conscious agents

$$C_1 = ((X_1, \mathbf{X}_1), (G_1, \mathbf{G}_1), P_1, D_1, A_1, N_1), \tag{18}$$

and

$$C_2 = ((X_2, \mathbf{X}_2), (G_2, \mathbf{G}_2), P_2, D_2, A_2, N_2), \tag{19}$$

have an undirected join. Let

$$C = ((X, \mathbf{X}), (G, \mathbf{G}), P, D, A, N)) \tag{20}$$

where

$$X = X_1 \times X_2, \tag{21}$$

$$G = G_1 \times G_2, \tag{22}$$

$$P = P_1 \otimes P_2 : G^T \times \mathbf{X} \to [0, 1], \tag{23}$$

$$D = D_1 \otimes D_2 : X \times \mathbf{G} \to [0, 1], \tag{24}$$

$$A = A_1 \otimes A_2 : G \times \mathbf{X}^T \to [0, 1], \tag{25}$$

$$N = N_1 = N_2, \tag{26}$$

where superscript T indicates transpose, e.g., $X^T = X_2 \times X_1$; where $\mathbf{X}$ is the σ-algebra generated by the Cartesian product of $\mathbf{X}_1$ and $\mathbf{X}_2$; where $\mathbf{G}$ is the σ-algebra generated by $\mathbf{G}_1$ and $\mathbf{G}_2$; and where the Markovian kernels P, D, and A are given explicitly, in the discrete case, by

$$\begin{aligned} P((g_2, g_1), (x_1, x_2)) &= P_1 \otimes P_2((g_2, g_1), (x_1, x_2)) \\ &= P_1(g_2, x_1)P_2(g_1, x_2), \end{aligned} \tag{27}$$

$$\begin{aligned} D((x_1, x_2), (g_1, g_2)) &= D_1 \otimes D_2((x_1, x_2), (g_1, g_2)) \\ &= D_1(x_1, g_1)\, D_2(x_2, g_2), \end{aligned} \tag{28}$$

$$\begin{aligned} A((g_1, g_2), (x_2, x_1)) &= A_1 \otimes A_2((g_1, g_2), (x_2, x_1)) \\ &= A_1(g_1, x_2)\, A_2(g_2, x_1), \end{aligned} \tag{29}$$

where $g_1 \in G_1$, $g_2 \in G_2$, $x_1 \in X_1$, and $x_2 \in X_2$. Then C satisfies the definition of a conscious agent. □

Thus, the undirected join of two conscious agents (illustrated in **Figure 2**) creates a single new conscious agent that we call their *undirected combination*. It is straightforward to extend the construction in Theorem 1 to the case in which more than two conscious agents have an undirected join. In this case the joined agents create a single new agent that is their undirected combination.

Theorem 2. (*Directed Join Theorem*.) A directed join of two conscious agents creates a new conscious agent.

Proof. (*By construction*.) Let two conscious agents

$$C_1 = ((X_1, \mathbf{X}_1), (G_1, \mathbf{G}_1), P_1, D_1, A_1, N_1), \tag{30}$$

and

$$C_2 = ((X_2, \mathbf{X}_2), (G_2, \mathbf{G}_2), P_2, D_2, A_2, N_2), \quad (31)$$

have the directed join $C_1 \rightarrow C_2$. Let

$$C = ((X, \mathbf{X}), (G, \mathbf{G}), P, D, A, N)) \quad (32)$$

where

$$X = X_1, \quad (33)$$
$$G = G_2, \quad (34)$$
$$P = P_1, \quad (35)$$
$$D = D_1A_1D_2 : X_1 \times \mathbf{G}_2 \rightarrow [0, 1], \quad (36)$$
$$A = A_2, \quad (37)$$
$$N = N_1 = N_2, \quad (38)$$

where $D_1A_1D_2$ denotes kernel composition. Then C satisfies the definition of a conscious agent. □

Thus, the directed join of two conscious agents creates a single new conscious agent that we call their *directed combination*. It is straightforward to extend the construction in Theorem 2 to the case in which more than one conscious agent has a directed join to C_2. In this case, all such agents, together with C_2, create a new agent that is their directed combination.

Given Theorems 1 and 2, we make the following

Conjecture 3: (*Combination Conjecture.*) Given any pseudograph of conscious agents, with any mix of directed and undirected edges, then any subset of conscious agents from the pseudograph, adjacent to each other or not, can be combined to create a new conscious agent.

How do these theorems address the problems of combining experiences and subjects? We consider first the combination of experiences.

Suppose C_1 has a space of possible perceptual experiences X_1, and C_2 has a space of possible perceptual experiences X_2. Then their undirected join creates a new conscious agent C that has a space of possible perceptual experiences $X = X_1 \times X_2$. In this case, C has possible experiences that are not possible for C_1 or C_2. If, for instance, C_1 can see only achromatic brightness, and C_2 can see only variations in hue, then C can see hues of varying brightness. Although C's possible experiences X are the Cartesian product of X_1 and X_2, nevertheless C might exhibit perceptual dependence between X_1 and X_2, due to feedback inherent in an undirected join (Maddox and Ashby, 1996; Ashby, 2000).

For a directed join $C_1 \rightarrow C_2$, the directed-combination agent C has a space of possible perceptual experiences $X = X_1$. This might suggest that no combination of experiences takes place. However, C has a decision kernel D that is given by the kernel product $D_1A_1D_2$. This product integrates (in the literal sense of integral calculus) over the entire space of perceptual experiences X_2, making these perceptual experiences an integral part of the decision process. This comports well with evidence that there is something it is like to make a decision (Nahmias et al., 2004; Bayne and Levy, 2006), and suggests the intriguing possibility that the phenomenology of decision making is intimately connected with the spaces of perceptual experiences that are integrated in the decision process. This is an interesting prediction of the formalism of conscious agents, and suggests that solution of the combination problem for experience will necessarily involve the integration of experience with decision-making.

We turn now to the combination of subjects. Coleman describes subjects as follows: "The idea of being a subject goes with being an experiential entity, something conscious of phenomenal qualities. That a given subject has a particular phenomenological point of view can be taken as saying that there exists a discrete 'sphere' of conscious-experiential goings-on corresponding to this subject, with regard to which other subjects are distinct in respect of the phenomenal qualities they experience, and they have no direct (i.e., experiential) access to the qualitative field enjoyed by the first subject. A subject, then, can be thought of as a point of view annexed to a private qualitative field" (Coleman, 2014).

A conscious agent C_i is a subject in the sense described by Coleman. It has a distinct sphere, X_i, of "conscious-experiential goings-on" and has no direct experiential access to the sphere, X_j, of experiences of any other conscious agent C_j. Moreover, a conscious agent is a subject in the further sense of being an *agent*, i.e., making decisions and taking actions on its own. Thus, according to the theory being explored here a subject, a point of view, is a six-tuple that satisfies the definition of a conscious agent.

The problem with combining subjects is, according to Goff, that "It is never the case that the existence of a number (one or more) of subjects of experience with certain phenomenal characters a priori entails the existence of some other subject of experience" (Goff, 2009).

Coleman goes further, saying that "The combination of subjects is a demonstrably incoherent notion, not just one lacking in a priori intelligibility . . . " (Coleman, 2014). He explains why: ". . . a set of points of view have nothing to contribute as such to a single, unified successor point of view. Their essential property defines them against it: in so far as they are points of view they are experientially distinct and isolated—they have different streams of consciousness. The diversity of the subject-set, of course, derives from the essential oneness of any given member: since each subject is essentially a oneness, a set of subjects are essentially diverse, for they must be a set of onenesses. Essential unity from essential diversity . . . is thus a case of emergence . . . "

The theory of conscious agents proposes that a subject, a point of view, is a six-tuple that satisfies the definition of conscious agent. The directed and undirected join theorems give constructive proofs of how conscious agents and, therefore, points of view, can be combined to create a new conscious agent, and thus a new point of view. The original agents, the original subjects, are not destroyed in the creation of the new agent, the new subject. Instead the original subjects structurally contribute in an understandable, indeed mathematically definable, fashion to the structure and properties of the new agent. The original agents are, indeed, influenced in the process, because they interact with each other. But they retain their identities. And the new agent has new properties not enjoyed by the constituent agents, but which are intelligible from the structure and interactions of the constituent

agents. In the case of undirected combination, for instance, we have seen that the new agent can have periodic asymptotic properties that are not possessed by the constituent agents but that are intelligible—and thus not emergent in a brute sense—from the structures and interactions of the constituent agents.

Thus, in short, the theory of conscious agents provides the first rigorous theoretical account of the combination of subjects. The formalism is rich with deductive implications to be explored. The discussion here is just a start. But one hint is the following. The undirected combination of two conscious agents is a single conscious agent whose world, W, is itself. This appears to be a model of *introspection*, in which introspection emerges, in an intelligible fashion, from the combination of conscious agents.

MICROPHYSICAL OBJECTS

We have sketched a theory of subjects. Now we use it to sketch a theory of objects, beginning with the microscopic and proceeding to the macroscopic.

The idea is that space-time and objects are among the symbols that conscious agents employ to represent the properties and interactions of conscious agents. Because each agent is finite, but the realm of interacting agents is infinite, the representations of each agent, in terms of space-time and objects, must omit and simplify. Hence the perceptions of each agent must serve as an interface to that infinite realm, not as an isomorphic map.

Interacting conscious agents form dynamical systems, with asymptotic (i.e., long-term) behaviors. We propose that microphysical objects represent asymptotic properties of the dynamics of conscious agents, and that space-time is simply a convenient framework for this representation. Specifically, we observe that the harmonic functions of the space-time chain that is associated with the dynamics of a system of conscious agents are identical to the wave function of a free particle; particles are vibrations not of strings but of interacting conscious agents.

Consider, for concreteness, the system of two conscious agents of section Dynamics of Two Conscious Agents, whose dynamics is governed by the kernel L of (17). This dynamics is clearly Markovian, because the change in state depends only on the current state. The *space-time chain* associated to L has, by definition, the kernel

$$Q : (E \times \mathbb{N}) \times (\mathbf{E} \otimes 2^{\mathbb{N}}) \to [0, 1], \tag{39}$$

given by

$$Q((e,\ n),\ A \times \{m\}) = \begin{cases} L(e,\ A) & \text{if } m = n + 1, \\ 0, & \text{otherwise,} \end{cases} \tag{40}$$

where e ∈ E, n,m ∈ $\mathbb{N}$, and A ∈ **E** (Revuz, 1984).

Then it is a theorem (Revuz, 1984) that, if Q is quasi-compact (this is true when the state space is finite, as here), the asymptotic dynamics of the Markov chain takes on a cyclical character:

- There are a finite number of invariant events or absorbing sets: once the chain lands in any of these, it stays there forever. And the union of these events exhausts the state space E. We will index these events with the letter ρ.
- Each invariant event ρ is partitioned into a finite number d_ρ of "asymptotic" events, indexed by ρ and by $\delta = 1, \ldots, d_\rho$, so that once the chain enters the asymptotic event δ, it will then proceed, with certainty, to $\delta + 1$, $\delta + 2$, and so on, cyclically around the set of asymptotic events for the invariant event ρ.

Then there is a correspondence between eigenfunctions of L and harmonic functions of Q (Revuz, 1984, p. 210) We let

$$\lambda_{\rho,k} = \exp(2i\pi k/d_\rho), \tag{41}$$

and

$$f_{\rho,k} = \sum_{\delta=1}^{d_\rho} (\lambda_{\rho,k})^\delta U_{\rho,\delta} \tag{42}$$

where ρ is the index over the invariant events (i.e., absorbing sets), the variable k is an integer modulo d_ρ, and $U_{\rho,\delta}$ is the indicator function of the asymptotic event with index ρ, δ. For instance, in the example of section First Example of Asymptotic Behavior, there are 6 absorbing sets, so $\rho = 1, 2, \ldots, 6$. The first absorbing set has only one state, so $d_1 = 1$. Similarly, $d_2 = 1$, $d_3 = 2$, $d_4 = d_5 = d_6 = 4$. The function $U_{1,1}$ has the value 1 on the state $|0000\rangle$ and 0 for all other states; $U_{5,3}$ has the value 1 on the state $|1100\rangle$ and 0 for all other states.

Then it is a theorem that

$$Lf_{\rho,k} = \lambda_{\rho,k} f_{\rho,k}, \tag{43}$$

i.e., that $f_{\rho,k}$ is an eigenfunction of L with eigenvalue $\lambda_{\rho,k}$, and that

$$g_{\rho,k}(\cdot\,, n) = (\lambda_{\rho,k})^{-n} f_{\rho,k}, \tag{44}$$

is Q-harmonic (Revuz, 1984). Then, using (41–42), we have

$$\begin{aligned} g_{\rho,k}(\cdot, n) &= \exp(2i\pi k/d_\rho)^{-n} \sum_{\delta=1}^{d_\rho} \exp(2i\pi k/d_\rho)^\delta U_{\rho,\delta} \\ &= \sum_{\delta=1}^{d_\rho} \exp(2i\pi k\frac{\delta}{d_\rho} - 2i\pi k\frac{n}{d_\rho}) U_{\rho,\delta} \\ &= \sum_{\delta=1}^{d_\rho} \mathrm{cis}(2\pi\frac{k\delta}{d_\rho} - 2\pi\frac{kn}{d_\rho}) U_{\rho,\delta} \\ &= \sum_{\delta=1}^{d_\rho} \mathrm{cis}(2\pi\frac{\delta}{d_{\rho,k}} - 2\pi\frac{n}{d_{\rho,k}}) U_{\rho,\delta} \end{aligned} \tag{45}$$

where $d_{\rho,k} = d_\rho/k$. This is identical in form to the wavefunction of the free particle (Allday, 2009, §7.2.3):

$$\psi(x, t) = A \sum_x \mathrm{cis}(2\pi\frac{x}{\lambda} - 2\pi\frac{t}{T})\,|x\rangle \tag{46}$$

This leads us to identify $A \leftrightarrow 1$, $U_{\rho,\delta} \leftrightarrow |x\rangle$, $\delta \leftrightarrow x$, $n \leftrightarrow t$, and $d_{\rho,k} \leftrightarrow \lambda = T$. Then the momentum of the particle is $p = h/d_{\rho,k}$ and its energy is $E = hc/d_{\rho,k}$, where h is Planck's constant and c is the speed of light.

Thus, we are identifying (1) a wavefunction ψ of the free particle with a harmonic function g of a space-time Markov chain of interacting conscious agents, (2) the position basis $|x\rangle$ of the particle with indicator functions $U_{\rho,\delta}$ of asymptotic events of the agent dynamics, (3) the position index x with the asymptotic state index δ, (4) the time parameter t with the step parameter n, (5) the wavelength λ and period T with the number of asymptotic events $d_{\rho,k}$ in the asymptotic behavior of the agents, and (6) the momentum p and energy E as functions inversely proportional to $d_{\rho,k}$.

Note that wavelength and period are identical here: in these units, the speed of the wave is 1.

This identification is for non-relativistic particles. For the relativistic case we sketch a promising direction to explore, starting with the dynamics of two conscious agents in an undirected join. In this case, the state of the dynamics has six components: N_1, N_2, X_1, X_2, G_1, G_2. We identify these with the generating vectors of a geometric algebra $\mathbb{G}(2, 4)$ (Doran and Lasenby, 2003). The components N_1 and N_2 have positive signature, and the remaining have negative signature. $\mathbb{G}(2, 4)$ is the conformal geometric algebra for a space-time with signature (1, 3), i.e., the Minkowski space of special relativity. The conformal group includes as a subgroup the Poincare group of space-time translations and rotations; but the full conformal group is needed for most massless relativistic theories, and appears in theories of supersymmetry and supergravity. The Lie group SU(2, 2) is isomorphic to the rotor group of $\mathbb{G}(2, 4)$, which provides a connection to the twistor program of Roger Penrose for quantum gravity (Penrose, 2004).

Thus, the idea is to construct a geometric algebra $\mathbb{G}(2, 4)$ from the dynamics of two conscious agents, and from this to construct space-time and massless particles. Each time we take an undirected join of two conscious agents, we get a new geometric algebra $\mathbb{G}(2, 4)$ with new basis vectors as described above. Thus, we get a nested hierarchy of such geometric algebras from which we can build space-time from the Planck scale up to macroscopic scales. The metric would arise from the channel capacity of the joined agents.

The massive case involves symmetry breaking, and a promising direction to explore here involves hierarchies of stopping times in the Markovian dynamics of conscious agents. The idea is that one system of conscious agents might infrequently interact with another system, an interaction that can be modeled using stopping times. Such interactions can create new conscious agents, using the combination theorems presented earlier, whose "time" is moving more slowly than that of the original systems of agents involved in the combination. This hierarchy of stopping times proceeds all the way up to the slow times of our own conscious experiences as human observers (roughly 10^{40} times slower than the Planck time). The hierarchy of stopping times is linked to a hierarchy of combinations of conscious agents, leading up to the highest level of conscious agents that constitute us, and beyond.

OBJECTIONS AND REPLIES

Here we summarize helpful feedback from readers of earlier drafts, in the form of objections and replies.

(1) Your definition of conscious agents could equally well-apply to unconscious agents. Thus, your theory says nothing about consciousness.

Even if the definition could apply to unconscious agents, that would not preclude it from applying to consciousness, any more than using the integers to count apples would preclude using them to count oranges.

(2) How can consciousness be cast in a mathematical formalism without losing something essential?

The mathematics does lose something essential, viz., consciousness itself. Similarly, mathematical models of weather also lose something essential, viz., weather itself. A mathematical model of hurricanes won't create rain, and a mathematical model of consciousness won't create consciousness. The math is not the territory. But, properly constructed, mathematics reveals the structure of the territory.

(3) Why do you represent qualia by a probability space X?

Probability spaces can be used, of course, to represent a diverse range of content domains, from the outcomes of coin-flips to the long-term behavior of equity markets. But this does not preclude using probability spaces to represent qualia. A probability space is not itself identical to qualia (or to coin flips or equity markets). To propose that we represent the possible qualia of a conscious agent by a probability space is to propose that qualia convey *information*, since probability and information are (as Shannon showed) transforms of each other. It is also to propose that qualia need not, in general, exhibit other structures, such as metrics or dimensions. Now certain qualia spaces, such as the space of phenomenal colors, do exhibit metrical and dimensional properties. These properties are not precluded. They are allowed but not required. All that is required is that we can meaningfully talk about the information content of qualia.

The qualia X of a conscious agent C are private, in the sense that no other conscious agent C_i can directly experience X. Instead each C_i experiences its own qualia X_i. Thus, the qualia X are "inside" the conscious agent C. The "outside" for C is W, or more precisely, W-C.

(4) A conscious agent should have free will. Where is this modeled in your definition?

The kernel D represents the free will choices of the conscious agent C. For any particular quale x in X, the kernel D gives a probability measure on possible actions in the set G that the conscious agent might choose to perform. We take this probability measure to represent the free will choice of the conscious agent. Thus, we

interpret the probabilities as objective probabilities, i.e., as representing a true nondeterminism in nature. We are inclined to interpret all the other probabilities as subjective, i.e., as reflections of ignorance and degrees of belief.

(5) A conscious agent should have goals and goal-directed behaviors. Where are these modeled in your definition?

Goals and goal-directed behaviors are not in the definition of conscious agent. This allows the possibility of goal-free conscious agents, and reflects the view that goals are not a definitional property of consciousness. However, since one can construct universal Turing machines from dynamical systems of conscious agents, it follows that one can create systems of conscious agents that exhibit goal-directed behaviors. Goals experienced as conscious desires can be represented as elements of a qualia space X.

(6) Your theory doesn't reject object permanence, because conscious agents are the "objects" that give rise to our perceptions of size and shape, and those agents are permanent even when we're not looking.

Conscious realism proposes that conscious agents are there even when one is not looking, and thus rejects solipsism. But it also rejects object permanence, viz., the doctrine that 3D space and physical objects exist when they are not perceived. To claim that conscious agents exist unperceived differs from the claim that unconscious objects and space-time exist unperceived.

(7) If our perceptions of space-time and objects don't resemble objective reality, if they're just a species-specific interface, then science is not possible.

The interface theory of perception poses no special problems for science. The normal process of creating theories and testing predictions continues as always. A particularly simple theory, viz., that our perceptions resemble reality, happens to be false. Fine. We can develop other theories of perception and reality, and test them. Science always faces the problem, well-known to philosophers of science, that no collection of data uniquely determines the correct theory. But that makes science a creative and engaging process.

(8) Your proposal that consciousness, rather than physics, is fundamental places consciousness outside of science.

Absolutely not. The onus is on us to provide a mathematically rigorous theory of consciousness, to show how current physics falls out as a special case, and to make new testable predictions beyond those of current physics. To dismiss the physicalist theory that space-time and objects are fundamental is not to reject the methodology of science. It is just to dismiss a specific theory that is false.

(9) You argue that natural selection does not favor true perceptions. But this entails that the reliability of our cognitive faculties is low or inscrutable, and therefore constitutes a defeater for belief in natural selection. See Alvin Plantinga's argument on this (Plantinga, 2002).

Evolutionary games and genetic algorithms demonstrate that natural selection does not, in general, favor true perceptions. But this entails nothing about the reliability of our cognitive faculties more generally. Indeed, selection pressures might favor more accurate logic and mathematics, since these are critical for the proper estimation of the fitness consequences of actions. The selection pressures on each cognitive faculty must be studied individually before conclusions about reliability are drawn.

(10) The undirected join of conscious agents doesn't really solve the problem of combining subjects, because the decision kernel of the combination is just the product of the decision kernels of the two conscious agents that are combined. This product only models two separate agents making separate decisions, not two subjects combined into a single decision-making subject.

It's true that the decision kernel, D, of the combination starts out as a product, indicating independent decisions. But as the conscious agents in the combination continue to interact, the decisions become less and less independent. In the asymptotic limit, the decision kernel D^n as $n \to \infty$ of the combination cannot, in general, be written as a product. In this limit, the combination now has a single unified decision kernel, not decomposable as a product of the original decision kernels. And yet the two conscious agents in the combination still retain their identities. Thus, the undirected join models a combination process which starts off as little more than the product of the constituent agents but ends up with those agents fully entangled to form a new conscious agent with a genuinely new and integrated decision kernel.

(11) If I have an objection it is that the authors' proposal is maybe not crazy enough. I am with them 100% when they compare neurons to icons on a computer screen. But (if I have understood them correctly) they then go on to attribute absolute existence to consciousness. My own inclination is to propose that consciousness is also just an icon on a computer screen.

Conscious realism is the hypothesis that the objective world W consists of conscious agents. The theory of conscious agents is a mathematical theory of consciousness that quantifies over qualia that it assumes really exist. So this theory does assume the existence of consciousness.

However, it does not assume incorrigibility of qualia (to believe one has a quale is to have one) or infallibility about the contents of one's consciousness. Psychophysical studies provide clear evidence against incorrigibility and infallibility [see, e.g., the literature on change blindness (Simons and Rensink, 2005)]. Nor does it assume that the mathematics of conscious agents is itself identical to consciousness; a theory is just a theory.

One might try to interpret the theory of conscious agents as describing a psychophysical monism, in which matter and consciousness are two aspects of a more abstract reality. Such an

interpretation, if possible, might still be unpalatable to most physicalists since it entails that dynamical physical properties, such as position, momentum and spin, have definite values only when they are observed.

(12) One problem with section Evolution and Perception is that the authors never define either their notion of Truth, or their notion of Perception. They seem to believe that if you startle at any sound of rustling leaves (as a sort of sensitive predator avoidance system), then when you run from a real predator, you are not in any way in touch with the truth. But this is incorrect.

For sake of brevity, we omitted our definitions of truth and perception from this paper. But they are defined precisely in papers that study the evolution of perception in Monte Carlo simulations of evolutionary games and genetic algorithms (Mark et al., 2010; Hoffman et al., 2013; Marion, 2013; Mark, 2013).

Briefly, we define a *perceptual strategy* as a measurable function (or, more generally, a Markovian kernel) $p : W \to X$, where W is a measurable space denoting the objective world and X is a measurable space denoting an organism's possible perceptions. If $X = W$ and p is an isomorphism that preserves all structures on W, then p is a *naïve realist* perceptual strategy. If $X \subset W$ and p is structure preserving on this subset, then p is a *strong critical realist* strategy. If X need not be a subset of W and p is structure preserving, then p is a *weak critical realist* strategy. If X need not be a subset of W and p need not be structure preserving, then p is an *interface* strategy. These strategies form a nested hierarchy: naïve realist strategies are a subset of strong critical realist, which are a subset of weak critical realist, which are a subset of interface.

Naïve realist strategies see all and only the truth. Strong critical realist strategies see some, but in general not all, of the truth. Weak critical realist strategies in general see none of the truth, but the relationships among their perceptions genuinely reflect true relationships in the structure of the objective world W. Interface strategies in general see none of the truth, and none of the true relationships in the structure of W. Thus, our mathematical formulation of perceptual strategies allows a nuanced exploration of the role of truth in perception.

We let these perceptual strategies compete in hundreds of thousands of evolutionary games in hundreds of thousands of randomly chosen worlds, and find that strategies which see some or all of the truth have a pathetic tendency to go extinct when competing against interface strategies that are tuned to fitness rather than truth. The various truth strategies don't even get a chance to compete in the genetic algorithms, because they are not fit enough even to get on the playing field.

Thus, natural selection favors interface strategies that are tuned to fitness, rather than truth. If an organism with an interface perceptual strategy perceives, say, a predatory lion, then it really does perceive a lion in the same sense that someone having a headache really does have a headache. However, this does not entail that the objective world, W, contains an observer-independent lion, any more than a blue rectangular icon on a computer desktop entails that there is a blue rectangular file in the computer. There is something in the objective world W that triggers the organism to perceive a lion, but whatever that something is, it almost surely doesn't resemble a lion. A lion is simply a species-specific adaptive symbol, not an insight into objective reality.

(13) In section Evolution and Perception, the authors' argument seems to be: **Argument 1:** (1) Natural selection favors fitness in perceptual systems. (2) Fitness is incompatible with truth. (3) Therefore, natural selection favors perceptions that do not see truth in whole or in part.

With some minor tweaking, Argument 1 can be made valid. But premise 2 is completely implausible. If a tiger is charging you with lunch on his mind, truth works in the service of fitness. (The authors' treatment here raises the question of why we have perceptual systems at all and not just kaleidoscope eyes. They never address this.)

The authors would object that premise 2 is too strong. They don't subscribe to premise 2, they would say. They would perhaps hold out for Argument 2:

Argument 2: (1) Natural selection favors fitness in perceptual systems. (2) Fitness need not always coincide with truth. (3) Therefore, natural selection favors perceptions that do not see truth in whole or in part.

But Argument 2 is not valid and not tweakable into a valid argument. The conclusion is a lot stronger than the premises.

Worse, any weaker premise doesn't give the authors their needed/wanted radical thesis: Perception is not about truth, it is about having kids. Which they insist must be interpreted as Perception is never about truth, but about having kids. But this interpretation is obviously false. For one thing, if an ancient ancestor of ours (call her, Ug) is successful in having kids, she needs to know the truth: that she has kids! Why? Because Ug needs to take care of them!

We do not use either argument. We simply use Monte Carlo simulations of evolutionary games and genetic algorithms to study the evolution of perceptual strategies (as discussed in Objection 12). We find, empirically, that strategies tuned to truth almost always go extinct, or never even arise, in hundreds of thousands of randomly chosen worlds.

The key to understanding this finding is the distinction between fitness and truth. If W denotes the objective world (i.e., the truth), O denotes an organism, S the state of that organism, and A an action of that organism, then one can describe fitness as a function $f : W \times O \times S \times A \to \Re$. In other words, fitness depends not only on the objective truth W, but also on the organism, its state and the action. Thus, fitness and truth are quite distinct. Only if the fitness function happens to be a monotonic function of some structure in W, i.e., so that truth and fitness happen to coincide, will natural selection allow a truth strategy to survive. In the generic case, where truth and fitness diverge, natural selection sends truth strategies to extinction.

To phrase this as an argument of the kind given in the objection we would have **Argument 3**: (1) Natural selection favors fitness in perceptual systems. (2) Truth *generically* diverges from fitness.

(3) Therefore, natural selection *generically* favors perceptions that diverge from the truth.

The word *generically* here is a technical term. Some property holds generically if it holds everywhere except on a set of measure zero. So, for instance, the cartesian coordinates (x, y) of a point in the plane *generically* have a non-zero y coordinate. Here we are assuming an unbiased (i.e., uniform) measure on the plane, in which the measure of a set is proportional to its area. Since the set of points with a zero y coordinate is the x-axis line, and since lines have no area, it follows that generically a point in the plane has a non-zero y coordinate. Note, however, that there are *infinitely* many points with a zero y coordinate, even though this property is non-generic.

So our argument is that, for an appropriate unbiased measure, fitness functions *generically* diverge from truth, and thus natural selection generically favors perceptions that diverge from truth. This does not entail the stronger conclusion that natural selection *never* favors truth. That conclusion is indeed stronger than our premises and stronger than required for the interface theory of perception. Perhaps *H. sapiens* is lucky and certain aspects of our perceptual evolution has been shaped by a non-generic fitness function that does not diverge from truth. In this case some aspects of our perceptions *might* be shaped to accurately report the truth, in the same sense that your lottery ticket *might* be the winner. But the smart money would bet long odds against it. That's what non-generic means.

The account of the interface theory about Ug's perception of her kids is the same as the account in Objection 12 for the perception of lions. There are no public physical objects. Lions and kids are no more public and observer independent than are headaches. Lions and kids (and space-time itself) are useful species-specific perceptions that have been shaped by natural selection not to report the truth but simply to guide adaptive behavior. We must take them seriously, but it is a logical error to conclude that we must take them literally.

Although our eyes do not report the truth, they are not kaleidoscope eyes because they do report what matters: fitness.

(14) We see then that the authors are caught in version of the Liar: Science shows that perception never cares about truth. Let this statement be *L*. *L* is derived via perception. So is *L* (together with its perceptual base) true or false? If it is one, then it is the other. Contradiction.

This is not our argument. We claim that perception evolved by natural selection. Call this statement *E*. Now *E* is indeed informed by the results of experiments, and thus by our perceptions. We observe, from evolutionary game theory, that one mathematical prediction of *E* is that natural selection generically drives true perceptions to extinction when they compete with perceptions tuned to fitness.

Suppose *E* is true. Then our perceptions evolved by natural selection. This logically entails that our perceptions are generically about fitness rather than truth. Is this a contradiction? Not at all. It is a scientific hypothesis that makes testable predictions. For instance, it predicts that (1) physical objects have no causal powers and (2) physical objects have no dynamical physical properties when they are not observed. These predictions are in fact compatible with quantum theory, and are part of the standard interpretation of quantum theory.

Suppose *E* is false. Then our perceptions did not evolve by natural selection. At present, science has no other theory on offer for the development of our perceptual systems. So, in this case, science cannot at present make an informed prediction about whether our perceptions are true or not. But this is not a logical contradiction.

So there is no liar paradox. And there'd better not be. Science cannot be precluded *a priori* from questioning the veridicality of the perceptions of *H. sapiens*, any more than it can be precluded from questioning the veridicality of the perceptions of other species. David Marr, for instance, argues that "... it is extremely unlikely that the fly has any explicit representation of the visual world around him—no true conception of a surface, for example, but just a few triggers and some specifically fly-centered parameters ... " and that the fly's perceptual information "... is all very subjective" (Marr, 1982, p. 34). Science has no trouble investigating the veridicality of the perceptions of other species and concluding, e.g., in the case of the fly, that they fail to be veridical. Its methods apply equally well to evaluating the veridicality of the perceptions of *H. sapiens* (Koenderink et al., 2010; Koenderink, 2011b, 2013).

(15) Section The Interface Theory of Perception fares no better. Here they say Reality, we learned, departed in important respects from some of our perceptions. This is true. But it is true because other perceptions of ours won out because they were true. E.g., the Earth is not a flat disk or plane.

Other perceptions indeed won out—not because they are true but because they are adaptive in a wider range of contexts. Flat earth is adequate for many everyday activities, but if one wants to circumnavigate the earth by boat then a spherical earth is more adaptive. If one wants to control satellites in orbit or navigate strategic submarines then a spherical earth is inadequate and a more complex model is required.

Perceived 3D space is simply a species-specific perceptual interface, not an insight into objective reality; we have argued for this on evolutionary grounds, and researchers in embodied cognition have arrived at a similar conclusion (Laflaquiere et al., 2013; Terekhov and O'Regan, 2013). Space as modeled in physics extends perceived space via the action of groups, e.g., the Euclidean group, Poincare group, or arbitrary differentiable coordinate transformations (Singh and Hoffman, 2013). Any objects embedded in space, including earth and its 3D shape, are thus descriptions in a species-specific vocabulary, not insights into objective reality.

(16) Also, I don't understand their interface theory of perception. I not only take my icons seriously, but literally: they are icons. I'm prepared to wager the farm on this: they are indeed icons.

We would agree that icons are indeed icons. When I open my eyes and see a red apple, that red apple is indeed an icon of my perceptual interface. When I close my eyes that icon disappears; I

see just a mottled gray field. Now some physicalists would like to claim that even when my eyes are closed, an objective red apple still exists, indeed the very red apple that triggered my perceptual interface to have a red apple icon. It is this claim that is generically incorrect, if our perceptual systems evolved by natural selection.

(17) The authors make too much of the Humean idea that the appearance of cause and effect is simply a useful fiction (section The Interface Theory of Perception). They like all mammals and perhaps most animals cannot fail to see causation in the deepest aspects of their lives. The authors believe in causation as deeply as anyone in the world. Why? Because we are all hardwired to see causation. And while it is true that causation goes away at the quantum level, we have no reason to believe that it doesn't really exist at the macro level. These two levels don't live well together, but pretending that there's no such thing as causation is silly, at least it is silly without a lot of argument. Even Hume admitted that causation was perfectly real when he had left his study and went to play backgammon with his friends.

There is indeed good evidence that belief in causation is either innate or learned early in life (Carey, 2009; Keil, 2011). And of course we, the authors, are no exception; we, no less than others, have a psychological penchant toward causal reasoning about the physical world. But, equally, we no less than others have a psychological penchant toward assuming that space, time and physical objects are not merely icons of a species-specific perceptual interface, but are instead real insights into the true nature of objective reality. Science has a habit of correcting our penchants, even those deeply held. Evolutionary games and genetic algorithms convinced us, against our deeply held convictions to the contrary, that perceptions are, almost surely, interfaces not insights; they also convinced us that the appearance of causality among physical objects is a useful fiction.

Perceptual icons do, we propose, *inform* the behavior of the perceiver, and in this sense might be claimed to have causal powers. This sense of causality, however, differs from that typically attributed to physical objects.

Hume's ideas on causation had little influence on us, in part because exegesis of his ideas is controversial, including projectivist, reductionist and realist interpretations (Garrett, 2009).

Our views on causality are consistent with interpretations of quantum theory that abandon microphysical causality, such as the Copenhagen, quantum Bayesian and (arguably) many-worlds interpretations, (Allday, 2009; Fuchs, 2010; Tegmark, 2014). The burden of proof is surely on one who would abandon microphysical causation but still cling to macrophysical causation.

(18) Their treatment of the combination problem is worth reading. There is however a very large problem with their model: It relies on the Cartesian product of X_1 and X_2 (this is right after Conjecture 3). The Cartesian product is not conducive to real combination (this problem is all over mathematics, by the way—mathematicians don't care about it because they only care about high level abstractions). In section Objections and Replies, where they discuss objections to their model, they discuss this very objection (objection 10). Unfortunately, their resolution to this objection is mere handwaving: But as the conscious agents in the combination continue to interact, the decisions become less and less independent. This is mere wishful thinking. The authors have no reason to believe this less and less business and they've given the reader no reason to think this either. In fact, if this less and less business were true, their model wouldn't require the Cartesian product in the first place. Frankly, this objection and their failure to handle it guts their model. In this same paragraph, in the next couple of sentences, the authors just assert (using proof by blatant assertion) that in some undefined limit, a true new conscious entity emerges. This makes the complex presentation of their model otiose. Why not just write a haiku asserting that the combination problem is not a problem?

The limit we speak of (for the emergence of a new combined conscious agent) is the *asymptotic limit*. Asymptotic behavior is a precise technical concept in the theory of Markov chains (see, e.g., Revuz, 1984, chapter 6). We have given, in sections First Example of Asymptotic Behavior and Second Example of Asymptotic Behavior, concrete examples of undirected joins for which, asymptotically, a new combined conscious agent is created that is not just a Cartesian product of the original agents.

Intuitively, the reason that the undirected combination of two agents creates a new agent that is not just a product is that there is feedback between the two agents (this is illustrated in **Figure 2**). Thus, the decisions and actions of one agent influence those of the other. This influence is not fully felt in the first step of the dynamics, but in the asymptotic limit of the dynamics it completely dominates, carving the state space of the dynamics into various absorbing sets with their own periodic behaviors, in a fashion that is not reducible to a simple product of the original two agents.

The degree to which the new conscious agent is not reducible to a simple product of the original agents can be precisely quantified using, for instance, the measure of *integrated information* developed by Tononi and others (Tononi and Edelman, 1998; Tononi and Spoorns, 2003; Tononi, 2008; Tononi and Koch, 2008; Barrett and Seth, 2011). It is straightforward to compute, for instance, that the new agent in Second Example of Asymptotic Behavior has 2 bits of integrated information, i.e., of new information that is not reducible to that of the two original agents. Thus, there is a precise and quantifiable sense in which the undirected combination of conscious agents creates a new conscious agent with its own new information.

We should note, however, that our use here of Tononi's measure of integrated information does not imply that we endorse his theory of consciousness. Tononi is a reductive functionalist, proposing that consciousness is *identical* to integrated information and that qualia are *identical* to specific informational relationships (Tononi, 2008). Consistent with this view he asserts, for instance, that spectrum inversion is impossible (Tononi, 2008, footnote 8). However, a recent theorem proves that *all* reductive functionalist theories of consciousness are false (Hoffman, 2006). A fortiori, Tononi's theory is false. His measure of integrated

information and his analyses of informational relationships are valuable. But his next move, of *identifying* consciousness with integrated information, is provably false. He could fix this by making the weaker claim that consciousness is *caused by* or *results from* integrated information. His theory would no longer be necessarily false. But then he would need to offer a scientific theory about how integrated information causes or gives rise to consciousness. No such theory is currently on offer and, we suspect, no such theory is possible.

(19) The paper explicitly commits a fallacy: it privileges the authors' take on reality while denying that there is any such thing as reality. For example: The authors say "There are no public physical objects. Lions and kids are no more public and observer independent than are headaches. Lions and kids (and space-time itself) are useful species-specific perceptions that have been shaped by natural selection not to report the truth but simply to guide adaptive behavior. We must take them seriously, but it is a logical error to conclude that we must take them literally."

Natural selection, which the authors clearly think is the truth, is just as susceptible to their arguments as headaches or truth itself. So by their own reasoning, natural selection is not true; neither are their computer programs/models. So the reader doesn't have to take natural selection or their models either seriously or literally. So their paper is now exposed as self-refuting.

If we indeed proposed a "take on reality while denying that there is any such thing as reality," we would of course be self-refuting. However, we do not deny that there is any such thing as reality. We cheerfully admit that there is a reality. We simply inquire into the relationship between reality and the perceptions of a particular species, *H. sapiens*. Such inquiry is surely within the purview of science. Moreover all currently accepted theories in science, including evolutionary theory, are appropriate tools for such inquiry.

We find that evolutionary theory entails a low probability that our perceptions are veridical, and thus a high probability that reality is not isomorphic to our perceptions, e.g., of spacetime and objects. This prompts us to propose a new theory of reality, which we have done by defining conscious agents and proposing conscious realism, viz., that reality consists of interacting conscious agents.

This proposal invites us to revisit evolutionary theory itself. The standard formulation of evolutionary theory, i.e., the neo-Darwinian synthesis, is couched in terms of spacetime and objects (such as organisms and genes), which we now take to be a species-specific perceptual representation, not an insight into reality. But we are not forced into self-refutation at this point. It is open to us to formulate a new *generalized* theory of evolution that operates on what we now take to be reality, viz., interacting systems of conscious agents.

A key constraint on our new evolutionary theory is this: When the new evolutionary theory is projected onto the spacetime perceptual interface of *H. sapiens* we must get back the standard evolutionary theory. Thus, we do not take the standard evolutionary theory to be true, but instead to be a "boundary condition" on the new evolutionary theory. Standard evolutionary theory is simply how the new evolutionary theory appears when it is shoehorned into the perceptual framework that *H. sapiens* happens to have.

The process we are describing here is standard procedure in science. We always use our current best theory as a ladder to a better theory, whereupon we can, if necessary, kick away the ladder. However, we needn't take our best theory to be true. It's simply the best ladder we have to our next theory. We are here adopting a philosophy of instrumentalism in regards to scientific theories.

The development of a new generalized theory of evolution is not just an abstract possibility, but is in fact one of our current projects. We are investigating the possibility of keeping the core ideas of standard evolutionary theory that are sometimes referred to as "Universal Darwinism," ideas that include abstract notions of variation, selection and retention. We plan to apply Universal Darwinism to interacting systems of conscious agents to model their evolution.

The new limited resource that is the source of competition would be *information*, which is the measure we use to quantify the channel capacity of conscious agents. This is a promising direction, since information is equivalent to energy, and information can be converted into energy (Toyabe et al., 2010). Limited energy resources, e.g., in the form of food, are a clear source of competition in standard evolutionary theory.

The new evolutionary theory that we construct should explain why the standard evolutionary theory was a good ladder to the new theory, and why we are justified in kicking away that ladder.

(20) The authors say, "In short, natural selection does not favor perceptual systems that see the truth in whole or in part. Instead, it favors perceptions that are fast, cheap, and tailored to guide behaviors needed to survive and reproduce. Perception is not about truth, it's about having kids." This is a false dichotomy.

The distinction between truth and fitness, between truth and having more kids, is not a false dichotomy to evolutionary biologists. It is a distinction that is central to their theory. The same objectively true world can have an infinite variety of different fitness functions, corresponding to the variety of organisms, states and actions. A steak that conveys substantial fitness benefits to a hungry lion conveys no benefits to a cow. Each distinct fitness function drives natural selection in a different direction.

(21) In response to the claim that "Your definition of conscious agents could equally well-apply to unconscious agents; thus, your theory says nothing about consciousness." the authors reply that "Even if the definition could apply to unconscious agents, that would not preclude it from applying to consciousness, any more than using the integers to count apples would preclude using them to count oranges."

However, the very fact that the integers can be used to count apples and oranges and peace treaties, etc., is precisely WHY the integers are not a theory of either apples or oranges or peace treaties, etc. The same is true of definitions.

> If my definition of integer applies equally well to the complex numbers as well as to the integers, then I do not have a definition of integers. Instead I have a definition of complex numbers. So their definition is useless; all they've done is define an agent. Consciousness is not present, except accidentally.

The integers are not considered a theory of peace treaties because they don't have the appropriate mathematical structure to model peace treaties—not because they can be used to count apples and peace treaties.

If one has a mathematical structure that is rich enough to provide a useful theory of some subject, this does not entail that the same structure *cannot* be a useful theory of a different subject. The group SU(3), for instance, models an exact symmetry of quark colors and an approximate symmetry of flavors. No physicist would insist that because SU(3) is a useful theory of quark color it cannot also be a useful theory of flavor. A given Markovian kernel P can model a stochastic dynamics, but also a communication channel. The fact that P applies to both does not entail that it's a theory of neither.

Similarly, a measurable space X might properly represent the conscious color experiences of a human observer, and also the unconscious color judgments of a robotic vision system designed to mimic that observer. No vision scientist would insist that because X properly represents the unconscious color judgments of the robotic vision system that therefore X cannot model the conscious color experiences of the human observer.

Scientists do not reject a model because it has multiple domains of useful application. They do reject a model if its structure is inappropriate to the domain, or if it makes predictions that are empirically false. These are the appropriate grounds to judge whether the formalism of conscious agents provides an adequate model for consciousness. The possibility that this formalism applies well to other domains does not entail that it cannot apply to consciousness.

CONCLUSION

Belief in object permanence commences at 3 months of age and continues for a lifetime. It inclines us to assume that objects exist without subjects to perceive them, and therefore that an account of objects can be given without a prior account of subjects.

However, studies with evolutionary games and genetic algorithms indicate that selection does not favor veridical perceptions, and that therefore the objects of our perceptual experiences are better understood as icons of a species-specific interface rather than as an insight into the objective structure of reality. This requires a fundamental reformulation of the theoretical framework for understanding objects.

This reformulation cannot assume that physical objects have genuine causal powers, nor that space-time is fundamental, since objects and space-time are simply species-specific perceptual adaptions.

If we assume that conscious subjects, rather than unconscious objects, are fundamental, then we must give a mathematically precise theory of such subjects, and show how objects, and indeed all physics, emerges from the theory of conscious subjects. This is, of course, a tall order. We have taken some first steps by (1) proposing the formalism of conscious agents, (2) using that formalism to find solutions to the combination problem of consciousness, and (3) sketching how the asymptotic dynamics of conscious agents might lead to particles and space-time itself. Much work remains to flesh out this account. But if it succeeds, *H. sapiens* might just replace object permanence with objects of consciousness.

ACKNOWLEDGMENTS

For helpful discussions and comments on previous drafts we thank Marcus Appleby, Wolfgang Baer, Deepak Chopra, Federico Faggin, Pete Foley, Stuart Hameroff, David Hoffman, Menas Kafatos, Joachim Keppler, Brian Marion, Justin Mark, Jeanric Meller, Julia Mossbridge, Darren Peshek, Manish Singh, Kyle Stephens, and an anonymous reviewer.

REFERENCES

Allday, J. (2009). *Quantum Reality: Theory and Philosophy.* Boca Raton, FL: CRC Press.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *J. Math. Psychol.* 44, 310–329. doi: 10.1006/jmps.1998.1249

Athreya, K. B., and Lahiri, S. N. (2006). *Measure Theory and Probability Theory.* New York, NY: Springer.

Baillargeon, R., and DeVos, J. (1991). Object permanence in young infants: further evidence. *Child Dev.* 62, 1227–1246. doi: 10.2307/1130803

Bargh, J. A., and Morsella, E. (2008). The unconscious mind. *Perspect. Psychol. Sci.* 3, 73–79. doi: 10.1111/j.1745-6916.2008.00064.x

Barrett, A. B., and Seth, A. K. (2011). Practical measures of integrated information for time-series data. *PLoS Comput. Biol.* 7:1. doi: 10.1371/journal.pcbi.1001052

Bayne, T., and Levy, N. (2006). "The feeling of doing: deconstructing the phenomenology of agency," in *Disorders of Volition,* eds N. Sebanz and W. Prinz (Cambridge, MA: MIT Press), 49–68.

Blackmore, S. (1999). *The Meme Machine.* Oxford: Oxford University Press.

Blamauer, M. (2011). Is the panpsychist better off as an idealist? Some Leibnizian remarks on consciousness and composition. *Eidos* 15, 48–75.

Boswell, J., (1986). *The Life of Samuel Johnson,* ed H. Christopher (New York, NY: Penguin Classics), 122.

Bower, T. G. R. (1974). *Development in Infancy.* San Francisco, CA: Freeman.

Byrne, A., and Hilbert, D. R. (2003). Color realism and color science. *Behav. Brain Sci.* 26, 3–26. doi: 10.1017/S0140525X03000013

Carey, S. (2009). *The Origin of Concepts.* New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195367638.001.0001

Ceccherini-Silberstein, T., and Coornaert, M. (2010). *Cellular Automata and Groups.* Berlin: Springer. doi: 10.1007/978-3-642-14034-1

Chartrand, G., and Ping, Z. (2012). *A First Course in Graph Theory.* Mineola, NY: Dover.

Coleman, S. (2014). The real combination problem: panpsychism, micro-subjects, and emergence. *Erkenntnis* 79, 19–44. doi: 10.1007/s10670-013-9431-x

Cook, M. (2004). Universality in elementary cellular automata. *Complex Syst.* 15, 1–40.

Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory.* New York, NY: Wiley.

Coyne, J. A. (2009). *Why Evolution is True.* New York, NY: Viking.

Crick, F. (1994). *The Astonishing Hypothesis: the Scientific Search for the Soul.* New York, NY: Scribners.

Dawkins, R. (2009). *The Greatest Show on Earth: the Evidence for Evolution.* New York, NY: Free Press.

Dennett, D. (1995). *Darwin's Dangerous Idea.* New York, NY: Touchstone.

Doran, C., and Lasenby, A. (2003). *Geometric Algebra for Physicists.* New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511807497

Edelman, G. M. (2004). *Wider than the Sky: the Phenomenal Gift of Consciousness.* New Haven, CT: Yale University Press.

Feldman, J. (2013). Tuning your priors to the world. *Top. Cogn. Sci.* 5, 13–34. doi: 10.1111/tops.12003

Fodor, J., and Piattelli-Palmarini, M. (2010). *What Darwin got Wrong, Farrar, Straus and Giroux.* New York, NY: Farrar, Straus and Giroux.

Fuchs, C. (2010). *QBism, the Perimeter of Quantum Bayesianism*. Available online at: arXiv:1003.5209v51.

Futuyma, D. J. (2010). Two critics without a clue. *Science* 328, 692–693. doi: 10.1126/science.1189090

Garrett, D. (2009). "Hume," in *The Oxford Handbook of Causation*, eds H. Beebee, C. Hitchcock, and P. Menzies (New York, NY: Oxford University Press), 53–71.

Geisler, W. S., and Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cogn. Sci.* 27, 379–402. doi: 10.1207/s15516709cog2703_3

Geisler, W. W., and Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philos. Trans. R. Soc. B* 357, 419–448. doi: 10.1098/rstb.2001.1055

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin.

Goff, P. (2009). Why panpsychism doesn't help us explain consciousness. *Dialectica* 63, 289–311. doi: 10.1111/j.1746-8361.2009.01196.x

Hameroff, S. R. (1998). Quantum computation in brain microtubules? The Penrose-Hameroff "Orch OR" model of consciousness". Philos. Trans. R. Soc. A 356, 1869–1896. doi: 10.1098/rsta.1998.0254

Herken, R. (ed.). (1988). *The Universal Turing Machine: a Half-century Survey.* New York, NY: Springer.

Hoffman, D. D. (1998). *Visual Intelligence: How We Create What We See*. New York, NY: W. W. Norton.

Hoffman, D. D. (2006). The scrambling theorem: a simple proof of the logical possibility of spectrum inversion. *Conscious. Cogn.* 15, 31–45. doi: 10.1016/j.concog.2005.06.002

Hoffman, D. D. (2009). "The interface theory of perception," in *Object Categorization: Computer and Human Vision Perspectives*, eds S. Dickinson, M. Tarr, A. Leonardis, and B. Schiele (New York, NY: Cambridge University Press), 148–165. doi: 10.1017/CBO9780511635465.009

Hoffman, D. D. (2011). "The construction of visual reality," in *Hallucinations: Theory and Practice*, eds J. Blom and I. Sommer (New York, NY: Springer), 7–15.

Hoffman, D. D. (2012). "The sensory desktop," in *This Will Make You Smarter: New Scientific Concepts to Improve Your Thinking*, ed J. Brockman (New York, NY: Harper Perennial), 135–138.

Hoffman, D. D. (2013). "Public objects and private qualia: the scope and limits of psychophysics," in *The Wiley-Blackwell Handbook of Experimental Phenomenology*, ed L. Albertazzi (New York, NY: Wiley-Blackwell), 71–89.

Hoffman, D. D., and Singh, M. (2012). Computational evolutionary perception. *Perception* 41, 1073–1091. doi: 10.1068/p7275

Hoffman, D. D., Singh, M., and Mark, J. T. (2013). "Does natural selection favor true perceptions," in *Proceedings of the SPIE* (Bellingham, WA).

Hopcroft, J. E., Motwani, R., and Ullman, J. D. (2006). *Introduction to Automata Theory, Languages, and Computation, 3rd Edn.* Upper Saddle River, NJ: Prentice Hall.

James, W. (1890). *The Principles of Psychology,* Vols 1, 2. London: MacMillan. doi: 10.1037/10538-000

James, W., (1890/2007). *The Principles of Psychology,* Vol. 1, New York, NY: Cosimo.

Keil, F.C. (2011). Science starts early. *Science* 33, 1022–1023. doi: 10.1126/science.1195221

Knill, D., and Richards, W. A. (eds.). (1996). *Perception as Bayesian Inference*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511984037

Koenderink, J. J. (2011a). "Vision and information," in *Perception Beyond Inference: The Information Content of Visual Processes*, eds L. Albertazzi, G. Tonder, and D. Vishnawath (Cambridge, MA: MIT Press), 27–58.

Koenderink, J. J. (2011b). "Vision as a user interface," in *Proceedings of SPIE, Human Vision and Electronic Imaging XVI,* Vol. 7865 (Bellingham, WA). doi: 10.1117/12.881671

Koenderink, J. J. (2013). "World, environment, umwelt, and inner-world: a biological perspective on visual awareness," in *Proceedings of SPIE, Human Vision and Electronic Imaging XVIII*, Vol. 8651 (Bellingham, WA). doi: 10.1117/12.2011874

Koenderink, J. J., van Doorn, A., de Ridder, H., and Oomes, S. (2010). Visual rays are parallel. *Perception* 39, 1163–1171. doi: 10.1068/p6530

Laflaquiere, A., Terekhov, A. V., Gas, B., and O'Regan, K. O. (2013). "Learning an internal representation of the end-effector configuration space," in *IEEE Conference on Intelligent Robots and Systems* (Tokyo), 1230–1235.

Lieberman, E., Hauert, C., and Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature* 433, 312–316. doi: 10.1038/nature03204

Locke, J. (1690). *An Essay Concerning Human Understanding*. London: Tho Basset.

Maddox, W. T. S., and Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *J. Exp. Psychol.* 22, 795–817.

Marion, B. B. (2013). *The Impact of Utility on the Evolution of Perceptions*. Ph.D. Dissertation, University of California, Irvine, CA.

Mark, J. T. (2013). *Evolutionary Pressures on Perception: When does Natural Selection Favor Truth?* Ph.D. Dissertation, University of California, Irvine, CA.

Mark, J. T., Marion, B. B., and Hoffman, D. D. (2010). Natural selection and veridical perceptions. *J. Theor. Biol.* 266, 504–515. doi: 10.1016/j.jtbi.2010.07.020

Marr, D. (1982). *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco, CA: Freeman.

Mausfeld, R. (2002). "The physicalist trap in perception theory," in *Perception and the Physical World: Psychological and Philosophical Issues in Perception*, eds D. Heyer and R. Mausfeld (New York, NY: John Wiley & Sons), 75–112.

McGinn, C. (1989). Can we solve the mind-body problem? *Mind* 98, 349–366. doi: 10.1093/mind/XCVIII.391.349

Mitchell, M. (1998). *An Introduction to Genetic Algorithms.* Cambridge, MA: Bradford Books MIT Press.

Nagel, T. (ed.). (1979). "Panpsychism," in *Mortal Questions* (New York, NY: Cambridge University Press), 181–195.

Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2004). The phenomenology of free will. *J. Conscious. Stud.* 11, 162–179.

Nowak, M. (2006). *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, MA: Belknap Press of Harvard University Press.

Pais, A. (1979). Einstein and the quantum theory. *Rev. Mod. Phys.* 51, 863–914. (Quote on p. 907). doi: 10.1103/RevModPhys.51.863

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Penrose, R. (2004). *The Road to Reality*, New York, NY: Knopf.

Piaget, J. (1954). *The Construction of Reality in the Child.* New York, NY: Basic. doi: 10.1037/11168-000

Plantinga, A. (2002). *Naturalism Defeated?* Ithaca, NY: Cornell University Press.

Revuz, D. (1984). *Markov Chains*, Amsterdam: North-Holland.

Rutherford, E. (1911). The scattering of α and β particle by matter and the structure of the atom. *Philos. Mag.* 6, 21.

Seager, W. (1995). Consciousness, information, and panpsychism. *J. Conscious. Stud.* 2, 272–288.

Simons, D. J., and Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends Cogn. Sci.* 9, 16–23. doi: 10.1016/j.tics.2004.11.006

Singh, M., and Hoffman, D. D. (2013). "Natural selection and shape perception: shape as an effective code for fitness," in *Shape Perception in Human and Computer Vision: an Interdisciplinary Perspective*, eds S. Dickinson and Z. Pizlo (New York, NY: Springer), 171–185.

Taylor, C. C. W. (1999). "The atomists," in *The Cambridge Companion to Early Greek Philosophy,* ed A. A. Long (New York, NY: Cambridge University Press), 181–204. doi: 10.1017/CCOL0521441226.009

Tegmark, M. (2014). *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality.* New York, NY: Knopf.

Terekhov, A., and O'Regan, K.O. (2013). *Space as an Invention of Biological Systems*. Available online at: arxiv.org/abs/1308.2124.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. doi: 10.2307/25470707

Tononi, G., and Edelman, G. (1998). Consciousness and complexity. *Science* 282, 1846–1851. doi: 10.1126/science.282.5395.1846

Tononi, G., and Koch, C. (2008). The neural correlates of consciousness: an update. *Ann. N.Y. Acad. Sci.* 1124, 239–261. doi: 10.1196/annals.1440.004

Tononi, G., and Spoorns, O. (2003). Measuring information integration. *BMC Neurosci.* 4:31. doi: 10.1186/1471-2202-4-31

Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E., and Sano, M. (2010). Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* 6, 988–992. doi: 10.1038/nphys1821

Turing, A.M. (1937). On computable numbers, with an application to the Entscheidungs problem. *Proc. Lond. Math. Soc.* 2, 42, 230–265. doi: 10.1112/plms/s2-42.1.230

Wilczek, F. (2006). *Fantastic Realities: 49 Mind Journeys and a Trip to Stockholm*. Singapore: World Scientific.

Yuille, A., and Bülthoff, H. (1996). "Bayesian decision theory and psychophysics," in *Perception as Bayesian inference*, eds D. Knill and W. Richards (New York, NY: Cambridge University Press), 123–162.

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 January 2014; accepted: 23 May 2014; published online: 17 June 2014.
Citation: Hoffman DD and Prakash C (2014) Objects of consciousness. Front. Psychol. ***5****:577. doi: 10.3389/fpsyg.2014.00577*
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.
Copyright © 2014 Hoffman and Prakash. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read, for greatest visibility

COLLABORATIVE PEER-REVIEW
Designed to be rigorous – yet also collaborative, fair and constructive

FAST PUBLICATION
Average 85 days from submission to publication (across all journals)

COPYRIGHT TO AUTHORS
No limit to article distribution and re-use

TRANSPARENT
Editors and reviewers acknowledged by name on published articles

SUPPORT
By our Swiss-based editorial team

IMPACT METRICS
Advanced metrics track your article's impact

GLOBAL SPREAD
5'100'000+ monthly article views and downloads

LOOP RESEARCH NETWORK
Our network increases readership for your article

Frontiers
EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on