

NEW TRENDS IN FALSE MEMORY RESEARCH

EDITED BY: Maria Soledad Beato, Sara Cadavid, Beatriz Martín-Luengo
and Rui Paulo

PUBLISHED IN: Frontiers in Psychology





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88976-314-6

DOI 10.3389/978-2-88976-314-6

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: frontiersin.org/about/contact

NEW TRENDS IN FALSE MEMORY RESEARCH

Topic Editors:

Maria Soledad Beato, University of Salamanca, Spain

Sara Cadavid, Rosario University, Colombia

Beatriz Martín-Luengo, National Research University Higher School of
Economics, Russia

Rui Paulo, Bath Spa University, United Kingdom

Citation: Beato, M. S., Cadavid, S., Martín-Luengo, B., Paulo, R., eds. (2022). New Trends in False Memory Research. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-314-6

Table of Contents

- 05 *Reducing False Recognition in the Deese-Roediger/McDermott Paradigm: Related Lures Reveal How Distinctive Encoding Improves Encoding and Monitoring Processes***
Mark J. Huff, Glen E. Bodner and Matthew R. Gretz
- 15 *The Role of Language Proficiency in False Memory: A Mini Review***
Mar Suarez and Maria Soledad Beato
- 24 *The Dynamic and Fragile Nature of Eyewitness Memory Formation: Considering Stress and Attention***
Alia N. Wulff and Ayanna K. Thomas
- 36 *Manipulations of List Type in the DRM Paradigm: A Review of How Structural and Conceptual Similarity Affect False Memory***
Jennifer H. Coane, Dawn M. McBride, Mark J. Huff, Kai Chang, Elizabeth M. Marsh and Kendal A. Smith
- 51 *Does Alexithymia Affect Memory for a Crime? The Relationship Between Alexithymia, Executive Functions, and Memories***
Fabiana Battista, Tiziana Lanciano and Antonietta Curci
- 63 *Prospective Memory and Positivity Bias in the COVID-19 Health Crisis: The Effects of Aging***
Alaitz Aizpurua, Malen Migueles and Ainara Aranberri
- 74 *Aging Does Not Enhance Social Contagion Effect***
Susana Carnero-Sierra and Julio Menor
- 80 *Working Memory Beats Age: Evidence of the Influence of Working Memory on the Production of Children's Emotional False Memories***
Chiara Mirandola and Francesca Pazzaglia
- 87 *Imagination Reduces False Memories for Everyday Action Sentences: Evidence From Pragmatic Inferences***
María J. Maraver, Ana Lapa, Leonel Garcia-Marques, Paula Carneiro and Ana Raposo
- 97 *Generalization and False Memory in an Acquired Equivalence Paradigm: The Influence of Physical Resemblance Across Related Episodes***
Caitlin R. Bowman, Maria-Alejandra de Araujo Sanchez, William Hou, Sarina Rubin and Dagmar Zeithamova
- 116 *Feelings of Contrast at Test Reduce False Memory in the Deese/Roediger-McDermott Paradigm***
Sara Cadavid, Maria Soledad Beato, Mar Suarez and Pedro B. Albuquerque
- 126 *Questioning the Role of Forward Associative Strength in False Memories: Evidence From Deese-Roediger-McDermott Lists With Three Critical Lures***
Maria Soledad Beato and Jason Arndt
- 134 *Transcranial Direct Current Stimulation Over the Right Anterior Temporal Lobe Does Not Modulate False Recognition***
María Angeles Alonso, Antonio M. Díez-Álamo, Carlos J. Gómez-Ariza, Emiliano Díez and Angel Fernandez

- 146** *Lie for Me: Developmental Trends in Acquiescing to a Blatantly False Statement*
Amelia Courtney Hritz and Stephen J. Ceci
- 156** *False Memories in Native and Foreign Languages*
Aleksandra Dolgoarshinnaia and Beatriz Martin-Luengo
- 168** *Identification Performance During Quarantine by COVID-19 Pandemic: Influence of Emotional Variables and Sleep Quality*
Facundo A. Urreta Benitez, Candela S. Leon, Matías Bonilla, Pablo Ezequiel Flores-Kanter and Cecilia Forcato
- 182** *Mediators and Moderators of Reinforced Self-Affirmation as a Method for Reducing the Memory Misinformation Effect*
Malwina Szpitalak and Romuald Polczyk
- 196** *The Trajectory of Targets and Critical Lures in the Deese/Roediger–McDermott Paradigm: A Systematic Review*
Patricia I. Coburn, Kirandeep K. Dogra, Iarenjit K. Rai and Daniel M. Bernstein
- 211** *Novel Approaches and Cognitive Neuroscience Perspectives on False Memory and Deception*
Michael P. Toglia, Joseph Schmuller, Britni G. Surprenant, Katherine C. Hooper, Natasha N. DeMeo and Brett L. Wallace



Reducing False Recognition in the Deese-Roediger/McDermott Paradigm: Related Lures Reveal How Distinctive Encoding Improves Encoding and Monitoring Processes

Mark J. Huff^{1*}, Glen E. Bodner² and Matthew R. Gretz¹

¹School of Psychology, The University of Southern Mississippi, Hattiesburg, MS, United States, ²College of Education, Psychology and Social Work, Flinders University, Adelaide, SA, Australia

OPEN ACCESS

Edited by:

Maria Soledad Beato,
University of Salamanca, Spain

Reviewed by:

Emiliano Díez,
University of Salamanca, Spain
Chunyan Guo,
Capital Normal University, China

*Correspondence:

Mark J. Huff
mark.huff@usm.edu

Specialty section:

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

Received: 03 September 2020

Accepted: 03 November 2020

Published: 20 November 2020

Citation:

Huff MJ, Bodner GE and
Gretz MR (2020) Reducing False
Recognition in the Deese-Roediger/
McDermott Paradigm: Related Lures
Reveal How Distinctive Encoding
Improves Encoding and
Monitoring Processes.
Front. Psychol. 11:602347.
doi: 10.3389/fpsyg.2020.602347

In the Deese-Roediger/McDermott (DRM) paradigm, distinctive encoding of list items typically reduces false recognition of critical lures relative to a read-only control. This reduction can be due to enhanced item-specific processing, reduced relational processing, and/or increased test-based monitoring. However, it is unclear whether distinctive encoding reduces false recognition in a selective or global manner. To examine this question, participants studied DRM lists using a distinctive item-specific anagram generation task and then completed a recognition test which included both DRM critical lures and either strongly related lures (Experiment 1) or weakly related lures (Experiment 2). Compared to a read-control group, the generate groups showed increased correct recognition and decreased false recognition of all lure types. We then estimated the separate contributions of encoding and retrieval processes using signal-detection indices. Generation improved correct recognition by both increasing encoding of memory information for list words and by increasing memory monitoring at test. Generation reduced false recognition by reducing the encoding of memory information and by increasing memory monitoring at test. The reduction in false recognition was equivalent for critical lures and related lures, indicating that generation globally reduces the encoding of related non-presented items at study (not just critical lures), while globally increasing list-theme-based monitoring at test.

Keywords: DRM illusion, distinctive encoding, false recognition, generation, distinctiveness

INTRODUCTION

Researchers have long been interested in techniques that can improve memory accuracy. Many of these techniques involve encoding tasks that induce a “deeper” level of processing of study materials (Craik and Lockhart, 1972; Craik, 2002). Examples include pleasantness ratings (Hunt and Einstein, 1981), generation (Slamecka and Graf, 1978; Bertsch et al., 2007), production (Conway and Gathercole, 1987; MacLeod and Bodner, 2017), and survival processing (Nairne et al., 2007; Nairne, 2015). Other techniques focus on enhancing retrieval-based processes such as ensuring a match between cues at study and test (Morris et al., 1977; Blaxton, 1989)

and instructing participants to stringently monitor their retrievals (Brainerd et al., 2001; Huff et al., 2011). Although these techniques can improve correct memory, their effects on memory errors and, in turn, on overall memory accuracy are as important. Here, we explore how one technique improves overall memory accuracy by shaping encoding and monitoring processes.

Memory errors are generally grouped into omission errors, which include forgetting and encoding failures, and commission errors, which refer to the remembering of events differently than their original presentation. One of the most robust and researched commission errors arises in the Deese/Roediger-McDermott (DRM; Deese, 1959; Roediger and McDermott, 1995) false memory paradigm. In the DRM paradigm, participants study lists of associates (e.g., *sour, candy, sugar*, etc.) that converge upon a single non-studied critical lure (e.g., *sweet*) that is later falsely reported or endorsed. The DRM false memory illusion is robust. False recall can exceed 50% (Roediger et al., 2001b), and false recognition can approximate hit rates for correctly studied list items (e.g., Lampinen et al., 1999; Dodson and Schacter, 2001). Additionally, participants often report conscious recollection of critical lures as appearing on studied lists (Payne et al., 1996), a pattern termed *phantom recollection* given that the critical lures were internally generated (Brainerd et al., 2003).

Several theories have been proposed to account for the DRM illusion (see Gallo, 2006 for review), most prominently the fuzzy-trace theory (Brainerd and Reyna, 2002; Reyna et al., 2016) and the activation-monitoring theory (Roediger et al., 2001a). Fuzzy-trace theory posits that two memory representations of study lists – verbatim and gist – are encoded. The verbatim representation contains memory for the specific items and any accompanying contextual details, whereas the gist representation contains the general meaning of the item or a group of related items in DRM lists. The DRM illusion must occur through a persistent gist representation because the critical lures do not have a verbatim representation. Activation-monitoring theory posits that the DRM illusion is the result of a two-stage process. First, the critical lure is implicitly activated during encoding through automatic spreading activation of associated study items (Collins and Loftus, 1975). Second, a source-monitoring failure occurs at test such that activation of the lure is misattributed to the studied list (Johnson et al., 1993). It is often difficult to disentangle these accounts because DRM list items both have (1) strong thematic coherence leading to extraction of a strong gist representation and (2) strong associations with the critical lure based on associative strength norms (Roediger et al., 2001a; Nelson et al., 2004). To circumvent this confound, researchers have had to employ different list types (e.g., homograph or mediated false memory lists) to reduce thematic coherence while maintaining associative strength (Hutchison and Balota, 2005; Huff et al., 2012). Studies taking these approaches suggest that both mechanisms can play a role (see Huff et al., 2015b).

With the goal of improving overall memory accuracy, researchers have identified several methods for reducing the DRM illusion, including study list repetitions (Benjamin, 2001), warnings (Gallo et al., 2001; McCabe and Smith, 2002), and

requiring participants to specify the source of their retrievals at test (Multhaup and Conner, 2002). Relevant to the present study, study tasks that encourage distinctive processing have been very fruitful, including perceptual manipulations, such as presenting study list words in unique fonts (Arndt and Reder, 2003) or paired with pictures (Israel and Schacter, 1997; Schacter et al., 1999; but see Smith and Hunt, 2020), and distinctive encoding tasks, such as mental images (Foley et al., 2006; Gunter et al., 2007; Robin, 2010; Oliver et al., 2016; Bodner et al., 2017), pleasantness ratings (Gunter et al., 2007; Huff and Bodner, 2013), and generation from anagram cues (McCabe and Smith, 2006; Huff et al., in press). Anagram generation, explored in our study, often yields an increase in correct recognition and a decrease in false recognition relative to a non-distinctive control task, a pattern termed a *mirror effect* (Glanzer and Adams, 1990; see Huff et al., 2015b for a review).

The benefits of distinctive processing induced by encoding tasks such as generation have generally been ascribed to two processes – one that occurs at encoding and the other at retrieval. The *impoverished relational encoding account* (Hockley and Cristi, 1996; Hege and Dodson, 2004) posits that distinctive processing disrupts encoding of the thematic meaning of the list or the implicit activation of the critical lure. The *distinctiveness heuristic*, on the other hand, posits that participants employ a test-based decision strategy in which recollection of distinctive details can be diagnostic that a study item was originally studied. Here, the absence of distinctive details can disqualify a test item from being reported as studied through a recall-to-reject process (Schacter et al., 1999; Gallo, 2004, 2010).

Several methods have been used to separate encoding and retrieval processes (see Huff et al., 2015b for a review and discussion). We have advocated for using a signal-detection approach when memory is tested *via* recognition (Gunter et al., 2007; Huff and Bodner, 2013; Bodner et al., 2017; Huff et al., in press). The primary advantage of the signal-detection approach is that it yields separate indices of the effects of manipulations on encoding (i.e., the amount of memory information encoded for a given type of test item) and retrieval (i.e., the extensiveness of participants' memory monitoring at test).

Using the signal-detection approach, Huff and Bodner (2013) compared the effects of different types of encoding manipulations on encoding and monitoring indices. In each experiment, the distinctive groups received item-specific processing instructions, a pleasantness-rating task, or an anagram-generation task and their memory was compared to a control (read-only) group. Each distinctive task group showed a mirror effect pattern in correct and false recognition relative to its control group. For correct recognition, the signal-detection indices of encoded memory information and monitoring were both greater following the distinctive tasks. For false recognition, monitoring for critical lures was greater in the distinctive task groups, consistent with use of a distinctiveness heuristic (Schacter et al., 1999). Encoded memory information was also lower in the distinctive tasks, consistent with impoverished relational encoding (Hege and Dodson, 2004). In addition, a meta-analysis confirmed that distinctive

tasks reduce the DRM illusion due to enhancement of both encoding and monitoring processes (Huff et al., 2015a).

Although we have learned much about how distinctive tasks operate to reduce false recognition, it is unclear whether their effects on encoding and retrieval processes operate globally (i.e., reducing false recognition of all lures that are related to a studied list) or are effective only on reducing false recognition of critical lures. This issue warrants attention given that the critical lures are qualitatively different than the other DRM list items. Critical lures have a high number of semantic associates (hence, their use as DRM critical lures), and they also tend to be higher in word frequency and concreteness – characteristics that can affect recognition accuracy (Balota and Neely, 1980; Roediger et al., 2001b). Indeed, false alarms to critical lures from non-studied lists (i.e., critical lure controls) are typically 5–7% greater than false alarms to list words from non-studied lists (Huff et al., 2015b). The reduction in false recognition enjoyed following distinctive encoding may, therefore, be restricted to critical lures due to their unique characteristics, rather than occurring globally to different types of recognition lures.

To determine whether reductions in false recognition are specific to critical lures or operate globally, the recognition tests in our experiments included a set of related lures from the DRM lists, in addition to the standard DRM critical lures. According to the impoverished relational encoding account, distinctive processing should reduce associative/thematic processing at study, and this reduction should affect any lure that shares a semantic association with the study list. Similarly, the distinctiveness heuristic is a global monitoring strategy and should similarly affect all test items, given that there is little evidence of within-test criterion shifts in recognition (Wixted and Stretch, 2000). Thus, although critical lures possess lexical and semantic characteristics that make them unique relative to other related lures, distinctive tasks should reduce false recognition globally for all lures that are related to the study list.

A few studies have tested recognition of related lures, separate from critical lures (e.g., Roediger and McDermott, 1995; Miller and Wolford, 1999; Miller et al., 2011; Smith and Hunt, 2020). Smith and Hunt (2020; Experiment 1) compared participants who viewed list items that were auditorily presented alongside a related picture to produce distinctive encoding (cf. Israel and Schacter, 1997) or who read/heard list items in isolation. After each list, participants completed a free recall test followed by a final recognition test that included both critical lures and weakly related lures (i.e., low associate DRM list items not presented in the study lists). False recognition of critical lures was lower in the distinctive picture group than the control; however, there was no difference in false recognition of weakly related lures. This pattern contrasts the notion that impoverished relational encoding and the distinctiveness operate globally, given that distinctive tasks had no effect on false recognition of weakly related lures. However, Smith and Hunt's participants completed a recall test prior to the final recognition test, which may have contaminated recognition (see Huff et al., 2018, for review). Moreover, Smith and Hunt did not find that picture

encoding improved correct recognition, unlike for other distinctive tasks, suggesting that picture encoding may not be as effective as other distinctive tasks. In short, the lack of reduction in false recognition for weakly related lures may be due to the initial recall test and/or use of an ineffective distinctive task.

Huff and Aschenbrenner (2018) studied how distinctive item-specific encoding instructions influenced correct and false recognition for categorized word lists rather than DRM lists. Their recognition task included categorically related critical lures. Distinctive instructions produced a mirror effect pattern. The signal-detection approach revealed that distinctive instructions increased memory monitoring for related lures relative to the read group, but encoded memory information was equivalent to the read group. Item-specific processing reduced false recognition of categorized lures, akin to the reduction found in studies using DRM lists. However, categorized lures differ from critical lures in that they overlap in semantic features rather than being associatively related to their study list. Thus, it remains possible that a reduction in false recognition may extend to other lure types in the DRM paradigm.

In summary, to date, there has not been a definitive answer as to whether distinctive tasks produce a global reduction in false recognition or a reduction that is specific to critical lures. Therefore, our primary goal was to examine the effects of distinctive encoding (*via* generation from anagram cues) on false recognition of both critical lures and related lures relative to a read-only control task. Previous work (Huff and Bodner, 2013, 2019) has indicated that the generation of individual anagrams (e.g., *terhad* → *thread*) induces distinctive item-specific processing. We, therefore, expected that generation would produce a mirror effect by improving correct recognition of studied list items (i.e., a generation effect; Slamecka and Graf, 1978; Bertsch et al., 2007) and by reducing false recognition of critical lures (Huff et al., 2015a). The key question was whether distinctive encoding also reduces false alarms for related lures. To examine this issue, across experiments, we varied the strength of the related lures we tested. In Experiment 1, we tested one strongly related lure from each studied DRM list. In Experiment 2, we tested one weakly related lure from each studied DRM list.

The signal-detection approach was then used to determine whether the anticipated reductions in false recognition for both lure types were due to encoding and/or monitoring-processes. If distinctive encoding reduced false recognition by leading to impoverished relational encoding, our estimate of the amount of memory information encoded should be lower for both critical lures and related lures in the generation group relative to the read group. Similarly, if the distinctiveness heuristic operates globally, our estimate of memory monitoring at test should be greater for both critical lures and related lures in the generation group relative to the read group. Indeed, the latter comparisons will indicate whether monitoring focuses on critical lures or is applied similarly to all related items. The distinctiveness heuristic assumes a global monitoring process, yet to our knowledge, this assumption has not been tested by including related lures at test.

EXPERIMENT 1: STRONGLY RELATED LURES

Experiment 1 examined the effects of a distinctive anagram-solution task on correct and false recognition relative to a read-only control group. Critically, the recognition test included both DRM critical lures and strongly related lures. Based on prior findings (e.g., Huff and Bodner, 2013), generation was expected to increase correct recognition and to reduce false recognition of critical lures. Our novel questions were (1) does generation also reduce false recognition of other theme-related lures? and (2) if so, does generation do so by decreasing global memory information for related lures and/or by increasing global monitoring at test? If distinctive generation operates globally, reduced encoding of memory information and increased monitoring at test should occur for both lure types.

Materials and Methods

Participants

Native English-speaking undergraduates from The University of Southern Mississippi participated for course credit. They were randomly assigned to the read or generate group. Five participants were excluded due to an unusual predominance of “old” responses across item types, leaving 64 participants (32 per group) for analysis. A sensitivity analysis using GPower 3 (Faul et al., 2007) indicated that this sample size had sufficient power (0.80) to detect medium-to-large sized effects and greater (Cohen’s $d \geq 0.70$).

Materials

The 20 DRM lists with the highest backward associative strength (BAS) from Roediger et al. (2001b) were used. Lists were divided into two counterbalanced sets of 10 lists in which one set was studied and the other was new. The top 12 associates from each list were used. The second highest associate in each list was designated a strongly related lure and was only included in the recognition test, leaving 11 words per DRM list. Lists were organized in descending BAS (Table 1; materials for our experiments are provided in our OSF project: www.osf.io/k73r4). For the generate group, anagrams were created by swapping either the first and third or second and fourth letters (cf. Gunter et al., 2007; Huff and Bodner, 2013). The eighty-item recognition test included 20 studied list items (from positions 1 and 8 in each list), 10 DRM critical lures from studied lists, 10 strongly related lures from study lists, 20 list item controls (from positions 1 and 8 in the non-studied set), 10 DRM critical lure controls, and 10 strongly related lure controls (from the non-studied set). Test items were newly randomized for each participant.

Procedure

Participants were tested individually with an experimenter present using a computer running SuperLab software (Cedrus Corporation). They were instructed that they would study lists of items for an upcoming memory test. During the study

TABLE 1 | Example study list items and backward associative strength (BAS) values for the critical lure “Shirt” with strongly and weakly related lures in Experiments 1 and 2.

List item	BAS
Blouse	0.647
Sleeves*	0.347
Collar	0.342
Shorts	0.252
Button	0.240
Pants	0.185
Polo	0.177
Jersey	0.174
Vest	0.143
Cuffs	0.143
Tie*	0.074
Pocket	0.058

*Strongly related lure used in Experiment 1. *Weakly related lure used in Experiment 2.

phase, read group participants read each word aloud and the experimenter advanced to the next word using a keyboard. Generate group participants were presented anagrams and were instructed to swap letters to generate a solution which they then read aloud (after Huff and Bodner, 2013, Experiment 3). If participants were unable to solve the anagram after a few seconds, the experimenter provided a hint (the first letter of the solution). If participants remained unable to solve the anagram after a few more seconds, the experimenter provided the solution and asked the participant to repeat it aloud. Thus, all participants read all list words aloud. The experimenter coded each trial as “correct,” “hint,” or “pass.”

The study phase began with an 8-item practice list; the experimenter provided feedback when necessary and answered questions about the tasks. Participants then studied the 10 DRM lists. Each list was separated by the words “next list.” The self-paced recognition test followed. Participants were told that words would be presented one at a time, and for each word, they should press the “old” or “new” labeled keys to indicate that the word was studied or not studied, respectively.

Results

Table 2 presents the mean proportion of “old” responses and mean signal-detection indices on the recognition test as a function of item type for the read and generate groups. The correct anagram completion rate (“correct” or “hint”) typically exceeded 95%, so analyses were not conditionalized on correct solution at study. The mean response time for correct anagram solutions (including hints) was 7.65 s ($SD = 3.10$). All comparisons were $p < 0.05$ unless noted otherwise. Estimates of effect size are provided for all significant comparisons using partial-eta squared (η_p^2) for analyses of variance (ANOVAs) or Cohen’s d for t -tests. Confidence intervals for effect size estimates (lower limit, upper limit), based on Smithson (2003), were computed using the MBESS package in R. For signal-detection analyses, false alarm rates of 0 and hit rates of 1 were adjusted using Macmillan and Creelman’s (1991) $1/2n$ correction. The reliability of non-significant comparisons was further tested using a

TABLE 2 | Mean (95% CI) proportion of “Old” responses and signal-detection indices as a function of item type/index and group/list type for test lists with strongly related lures (Experiment 1), weakly related lures (Experiment 2), and means pooled across experiments.

Encoding group/ item type/index	Experiment 1: strongly related lure		Experiment 2: weakly related lure		Pooled experiments	
	Read	Gen	Read	Gen	Read	Gen
<i>N</i>	32	32	34	34	66	66
List items	0.76 (0.05)	0.85 (0.05)	0.83 (0.03)	0.85 (0.03)	0.80 (0.03)	0.85 (0.03)
List item controls	0.15 (0.04)	0.10 (0.03)	0.10 (0.04)	0.07 (0.03)	0.12 (0.03)	0.08 (0.02)
List items <i>d'</i>	1.96 (0.19)	2.53 (0.22)	2.41 (0.20)	2.68 (0.19)	2.19 (0.15)	2.60 (0.15)
List items λ	1.17 (0.19)	1.36 (0.15)	1.39 (0.17)	1.57 (0.14)	1.28 (0.13)	1.47 (0.11)
Critical lures	0.58 (0.05)	0.46 (0.09)	0.66 (0.07)	0.51 (0.08)	0.62 (0.04)	0.48 (0.06)
Critical lure controls	0.18 (0.04)	0.13 (0.04)	0.18 (0.06)	0.13 (0.05)	0.18 (0.04)	0.13 (0.03)
Critical lures <i>d'</i>	1.20 (0.19)	1.00 (0.27)	1.45 (0.21)	1.19 (0.26)	1.33 (0.14)	1.10 (0.19)
Critical lures λ	0.97 (0.16)	1.15 (0.16)	0.99 (0.20)	1.17 (0.16)	0.98 (0.13)	1.17 (0.11)
Related lures	0.27 (0.05)	0.16 (0.05)	0.20 (0.05)	0.12 (0.04)	0.23 (0.04)	0.14 (0.03)
Related lure controls	0.13 (0.05)	0.11 (0.05)	0.06 (0.03)	0.02 (0.02)	0.09 (0.03)	0.06 (0.03)
Related lures <i>d'</i>	0.49 (0.17)	0.19 (0.23)	0.48 (0.19)	0.38 (0.14)	0.49 (0.13)	0.29 (0.13)
Related lures λ	1.17 (0.18)	1.25 (0.17)	1.42 (0.13)	1.57 (0.06)	1.30 (0.11)	1.41 (0.10)

Bayesian estimate of the strength of evidence supporting the null hypothesis (Wagenmakers, 2007; Masson, 2011). This analysis compares a model that assumes a significant effect to a model assuming a null effect. This Bayesian analysis yields a probability estimate termed p_{BIC} (Bayesian information criterion), which indicates the likelihood that the null hypothesis is supported. The p_{BIC} analysis is highly sensitive to sample size and thus provides a way of gauging the strength of evidence for reported null effects.

Correct Recognition

A comparison of the hit rate for studied list items across the read and generate groups showed a reliable generation effect (0.85 vs. 0.76), $t(62) = 2.67$, $SEM = 0.03$, $d = 0.68$ (0.16, 1.17). The same analysis was performed for list item *d'*, our estimate of encoded memory information (Huff and Bodner, 2013). Here, *d'* values were computed by taking the difference between the *z*-score for the hit rate for list items minus the *z*-score for the false alarm rate to list item controls. This analysis indicated that the generate group had encoded more memory information about the list items than the read group (2.53 vs. 1.96), $t(62) = 3.75$, $SEM = 0.15$, $d = 0.95$ (0.42, 1.45). A final comparison examined lambda, an index of test-based monitoring. Lambda was computed by taking the *z*-score of 1 minus the false alarm rate for list item controls. Memory monitoring for studied words was similar across the generate and read groups (1.36 vs. 1.17), $t(62) = 1.49$, $SEM = 0.12$, $p = 0.14$, $p_{\text{BIC}} = 0.72$.

False Recognition

A mixed 2×2 ANOVA compared false recognition as a function of lure type (critical vs. strongly related) and group (generate vs. read). As expected, false recognition was greater for critical lures than for strongly related lures (0.52 vs. 0.21), $F(1, 62) = 191.09$, $MSE = 0.02$, $\eta_p^2 = 0.76$ (0.65, 0.81). The main effect of group indicated that our distinctive generation task reduced false recognition of related lures overall relative

to reading (0.31 vs. 0.43), $F(1, 62) = 8.20$, $MSE = 0.05$, $\eta_p^2 = 0.12$ (0.02, 0.24), consistent with Huff and Bodner (2013). Most importantly, the reduction in false recognition was similar for both lure types, $F < 1$, $p_{\text{BIC}} = 0.87$.

Next, we examined the effect of generation on our signal-detection estimates of encoded memory information and memory monitoring for lures. For each type of lure, the encoded memory information *d'* index was computed as the difference in *z*-score for lures from the studied lists (treated as hits) vs. the corresponding lures from the non-studied lists (treated as false alarms). The 2×2 ANOVA indicated that more memory information had been encoded for critical lures than for strongly related lures (1.10 vs. 0.34), $F(1, 62) = 94.57$, $MSE = 0.07$, $\eta_p^2 = 0.60$ (0.47, 0.69). There was a general trend for generation to reduce the amount of memory information encoded for lures relative to reading (0.60 vs. 0.84), $F(1, 62) = 3.17$, $MSE = 0.62$, $p = 0.08$, $\eta_p^2 = 0.05$ (0.00, 0.16), $p_{\text{BIC}} = 0.62$, but it was not significant. The interaction was non-significant, $F < 1$, $p_{\text{BIC}} = 87$. Finally, estimates of memory monitoring were also compared using the same ANOVA. Interestingly, monitoring at test was greater for strongly related than for critical lures (1.21 vs. 1.06), $F(1, 62) = 5.15$, $MSE = 0.13$, $\eta_p^2 = 0.08$ (0.01, 0.20); we return to this result in our General Discussion section. Monitoring for lures was not significantly greater in the generate group than the read group (1.20 vs. 1.07), $F(1, 62) = 1.60$, $MSE = 0.33$, $p = 0.21$, $p_{\text{BIC}} = 0.78$. The interaction was non-significant, $F < 1$, $p_{\text{BIC}} = 0.84$.

Discussion

Our distinctive encoding task – anagram generation – increased correct recognition and reduced false recognition, replicating previous research (e.g., Huff and Bodner, 2013). Our novel finding was that generation reduced false recognition similarly for both critical lures and strongly related lures. Turning to our signal-detection analyses, for correct recognition, generation improved encoded memory information for list items (as in Huff and Bodner, 2013) but did not significantly increase

memory monitoring (unlike in Huff and Bodner, 2013). For false recognition, generation did not significantly reduce encoded memory information about lures, nor did it significantly increase memory monitoring at test (again, unlike in Huff and Bodner, 2013). In sum, although generation reduced false recognition, contrary to our expectations, it did not significantly reduce the encoding of lures at study or increase the monitoring for lures at test.

EXPERIMENT 2: WEAKLY RELATED LURES

Experiment 2 revisited the influences of distinctive processing on false recognition, this time using weakly related lures – the type used in studies that have assessed false recognition for related lures (Roediger and McDermott, 1995; Miller and Wolford, 1999; Smith and Hunt, 2020). The reduced association between weakly related lures and studied lists provides a more stringent test of the generality of the global reduction in false recognition following generation and thus should help us pinpoint its locus. In particular, if the generate group engages in stricter monitoring at test, they might be able to weed out critical lures more effectively than weakly related lures. Experiment 2 also sought to clarify whether false recognition reductions due to generation are attributable to increased memory information at encoding and/or increased monitoring at test for both lure types – given that the results of Experiment 1 did not clearly adjudicate among these two loci.

Materials and Methods

Participants

Additional participants from the Experiment 1 pool were randomly assigned to the read or generate groups. As per Experiment 1, three participants were excluded due to unusually high rates of “old” responses, leaving 68 participants (34 per group).

Materials and Procedure

The only changes in Experiment 2 were that (1) the strongly related lures from Experiment 1 were reinserted in their corresponding DRM list (position 2) and (2) the eleventh associate from each DRM study list was removed and this set served as the weakly related lures on the recognition test (Table 1). The procedure was identical to that of Experiment 1. Correct anagram solution rates were again quite high (95% or greater), and the mean response time for correct anagram solutions (including hints) was 6.13 s ($SD = 1.54$).

Results

Correct Recognition

The effects of generation on correct recognition were in the expected direction for each measure but did not reach significance (cf. Experiment 1; see also Huff et al., 2015b). This was true for hits (0.85 vs. 0.83), $t < 1$, $p_{BIC} = 0.84$, encoded memory information (d' ; 2.68 vs. 2.41), $t(66) = 1.89$, $SEM = 0.14$,

$p = 0.06$, $d = 0.47$ (−0.03, 0.94), $p_{BIC} = 0.58$, and memory monitoring (λ ; 1.57 vs. 1.39), $t(66) = 1.65$, $SEM = 0.11$, $p = 0.10$, $d = 0.41$ (−0.08, 0.88), $p_{BIC} = 0.68$.

False Recognition

False recognition was greater for critical lures than weakly related lures (0.58 vs. 0.16), $F(1, 66) = 220.99$, $MSE = 0.03$, $\eta_p^2 = 0.77$ (0.68, 0.82). More importantly, false recognition was lower in the generate group than in the read group (0.32 vs. 0.43), $F(1, 66) = 10.39$, $MSE = 0.04$, $\eta_p^2 = 0.14$ (0.03, 0.26). But most importantly, as in Experiment 1, the generation effect on false recognition was consistent across lure types, as indicated by a non-significant interaction, $F(1, 66) = 1.80$, $MSE = 0.03$, $p = 0.18$, $p_{BIC} = 0.77$.

Turning to our signal-detection measures, more memory information was encoded for critical lures than weakly related lures (1.32 vs. 0.43), $F(1, 66) = 78.59$, $MSE = 0.34$, $\eta_p^2 = 0.54$ (0.40, 0.64), as expected. As in Experiment 1, there was a non-significant trend for generation to reduce the amount of memory information encoded for lures relative to reading (0.79 vs. 0.97), $F(1, 66) = 2.74$, $MSE = 0.40$, $p = 0.10$, $\eta_p^2 = 0.04$ (0.00, 0.14), $p_{BIC} = 0.67$. The interaction with lure type was again non-significant, $F < 1$, $p_{BIC} = 0.86$. Memory monitoring at test was higher for weakly related lures than critical lures (1.49 vs. 1.08), $F(1, 66) = 45.22$, $MSE = 0.13$, $\eta_p^2 = 0.41$ (0.25, 0.52), as was true for strongly related lures in Experiment 1. Overall monitoring was only marginally greater in the generate than read group (1.37 vs. 1.21), $F(1, 66) = 3.58$, $MSE = 0.26$, $p = 0.06$, $\eta_p^2 = 0.05$ (0.00, 0.16), $p_{BIC} = 0.60$. The interaction was again non-significant, $F < 1$, $p_{BIC} = 0.89$.

Pooled Analysis of Experiments 1 and 2

In general, the patterns in Experiments 1 and 2 were highly similar, but several of the effects of generation were marginal or non-significant (and were also associated with lower p_{BIC} values). Therefore, we pooled our experiments to enable more powerful tests of the effects of generation, particularly on encoded memory information and memory monitoring at test. This pooling provided sufficient power to detect medium-sized effects and larger (Cohen's $d \geq 0.45$; Faul et al., 2007).¹

Correct Recognition

The pooled analysis aligned with the significant generation effects in Experiment 1. Generation increased hits relative to reading (0.85 vs. 0.80), $t(130) = 2.63$, $SEM = 0.02$, $d = 0.46$ (0.11, 0.80), and this generation effect was due to both increased encoding of memory information for list items at study

¹Experiment, when included as a factor, interacted with related lure type on false recognition, $F(1, 128) = 8.74$, $MSE = 0.13$, $\eta_p^2 = 0.06$ (0.01, 0.14), due to greater false recognition for strongly than weakly related lures (0.21 vs. 0.16), $t(130) = 2.05$, $SEM = 0.03$, $d = 0.36$ (0.01, 0.70), and also interacted in memory monitoring, $F(1, 128) = 10.38$, $MSE = 0.02$, $\eta_p^2 = 0.06$ (0.02, 0.16), due to greater monitoring for list items in weakly than strongly related lure lists (1.48 vs. 1.26), $t(130) = 3.98$, $SEM = 0.07$, $d = 0.70$ (0.09, 0.79). The other main effects and interactions were not reliable, $F_s < 2.80$, $p_s > 0.10$, $p_{BIC} > 0.73$. These expected interactions validate our lure strength manipulation. For brevity, pooled analyses excluded the Experiment factor.

(2.60 vs. 2.19), $t(130) = 3.86$, $SEM = 0.11$, $d = 0.68$ (0.32, 1.02), and increased memory monitoring for list items at test (1.47 vs. 1.28), $t(130) = 2.18$, $SEM = 0.08$, $d = 0.38$ (0.03, 0.72).

False Recognition

False recognition (averaged across critical lures and related lures) was lower in the generate group than in the read group (0.48 vs. 0.62), $F(1, 130) = 18.67$, $MSE = 0.05$, $\eta_p^2 = 0.13$ (0.13, 0.22). This reduction was equivalent for the two lure types, $F(1, 130) = 1.88$, $MSE = 0.02$, $p = 0.17$, $p_{BIC} = 0.81$ for the interaction. These patterns replicated the individual experiments but are reported here for completeness.

The pooled analysis yielded much clearer results regarding the effects of generation on the signal-detection measures of false recognition. Across lure types, generation significantly reduced encoded memory information relative to reading (0.69 vs. 0.91), $F(1, 130) = 5.83$, $MSE = 0.51$, $\eta_p^2 = 0.04$ (0.00, 0.11), and this reduction was similar for critical lures and related lures, $F < 1$, $p_{BIC} = 0.92$ for the interaction. Memory monitoring was also significantly greater in the generate group than in the read group (1.29 vs. 1.14), $F(1, 130) = 4.76$, $MSE = 0.30$, $\eta_p^2 = 0.04$ (0.00, 0.10), and this increase in monitoring was again similar for critical and related lures, $F < 1$, $p_{BIC} = 0.89$ for the interaction.

Discussion

In Experiment 2, generation did not significantly improve correct recognition over reading, unlike Experiment 1 (and unlike in Huff and Bodner, 2013). This is not unprecedented: The generation effect is typically small in between-group designs (Bertsch et al., 2007), and we recently reported a null effect of the same generation task in free recall (Huff and Bodner, 2019). However, generation successfully reduced false recognition of both critical lures and weakly related lures. Here, our signal-detection indices of memory information and memory monitoring showed only marginal effects of generation. Given the similarities in design and logic of Experiments 1 and 2, we, therefore, conducted a pooled analysis. The basic recognition analyses showed that distinctive processing in the generate group led to increased correct recognition and reduced false recognition, and critically, the latter reduction was similar for critical and related lure types. Our signal-detection analyses further clarified that for correct recognition, generation increased memory information encoded for list items and increased test-based memory monitoring. For false recognition, generation decreased encoded memory information for lures and increased memory monitoring. Most importantly, all of these effects were invariant across lure types. Collectively, these patterns are consistent with Huff and Bodner (2013) and reveal that distinctive encoding reduces false recognition by (1) globally reducing encoding of related lures at study and (2) globally increasing monitoring for related lures at test.

GENERAL DISCUSSION

The goal of this research was to help pinpoint how distinctive encoding tasks influence encoding and monitoring processes

in the DRM false memory paradigm. Overall, relative to a read-only control, an item-specific anagram generation task improved correct recognition and reduced false recognition. Critically, the reduction in false recognition for critical lures extended to both strongly related (Experiment 1) and weakly related (Experiment 2) lures. Our signal-detection analyses evaluated the effect of generation on separate estimates of encoding- and test-based processes. Across experiments, generation increased the amount of encoded memory information for studied list items and decreased the amount of associative/relational memory information encoded for lures relative to the read group, a pattern consistent with an impoverished relational encoding account (Hege and Dodson, 2004). Generation also increased the amount of memory monitoring at test for all test items including related lures, suggesting that participants are monitoring test items more stringently, a pattern consistent with a distinctiveness heuristic account (Schacter et al., 1999). Thus, impoverished relational encoding and use of a distinctiveness heuristic contribute to the reduction of false recognition collectively, and furthermore, we have learned that both processes operate globally rather than targeting encoding or monitoring specifically for critical lures – items that differ qualitatively from other related lures.

The effects of distinctive tasks on encoding and monitoring patterns reported in these previous studies (Huff and Bodner, 2013; Huff et al., 2015b) were based solely on false recognition of critical lures, leaving it unclear whether these processes operate globally. The lack of lure-type interactions in the present study indicate that distinctive processing operates broadly and have similar effects on strongly and weakly related lures. Indeed, this global pattern on recognition is consistent with other evidence indicating that participants adopt a consistent response criterion on a recognition test (Wickens and Hirshman, 2000; Wixted and Stretch, 2000; Gallo et al., 2001).

Although generation generally produced similar effects on false recognition of both lure types, we obtained an interesting difference between lure types in our monitoring estimate. Specifically, monitoring was lower for critical lures than for either strongly or weakly related lures. These monitoring differences could reflect inherent differences between critical lures and other list items (and thus than our related lures) in terms of their frequency or concreteness. Indeed, critical lures from non-studied lists yield a higher false alarm rate than list words from non-studied lists (Roediger and McDermott, 1995; Fenn et al., 2009). Given that the baseline false alarm rate to controls is used to compute monitoring estimates, monitoring estimates would, therefore, be lower for critical lures than related lures.

Alternatively, test-based semantic priming might contribute to the greater false alarm rate to critical lure controls than to related lure controls. On the recognition test, participants received three types of control items, the critical lure control, the related lure control, and list item controls from non-studied lists. Because the order of test items was random, list item controls preceded the critical lure controls for some lists and participants; this may have increased the familiarity of the critical lure controls and thus may have contributed to false alarms.

Indeed, this *test-induced priming* has been reported on recognition tests when related test items precede lures (Marsh et al., 2004; Coane and McBride, 2006). False alarms would likely be greater for critical than related controls due the stronger associative strength between list items and critical lures, yielding a reduced monitoring estimate for critical lures. Consistent with both possibilities, false alarms were higher for critical lure controls than for related lure controls across experiments, 0.16 vs. 0.08, $t(130) = 6.06$, $SEM = 0.01$, $d = 0.57$ (0.32, 0.82), resulting in lower monitoring estimates for critical lures. Importantly, however, lexical/semantic item differences and test-induced priming likely would be similar for generate and read groups. Thus, it is unlikely that these item differences contributed to the monitoring differences between our generate and read groups.

Our study also provides clarity regarding the effects of distinctive processing on related lures. As reviewed above, Smith and Hunt (2020) included related lures in a recognition test following either a distinctive picture encoding task or an auditory control task. Their study did not find an effect of distinctive study on recognition of related lures. However, a free-recall test was completed prior to the recognition test. Initial recall testing has been found to encourage organizational/relational processing that mitigates the effects of distinctive item-specific processing on a subsequent recognition test (Burns, 1993; Zaromb and Roediger, 2010). Our findings are more consistent with those of Huff and Aschenbrenner (2018), who found a false recognition reduction for categorically related lures, indicating that distinctive encoding tasks can be effective with other types of related lures.

One limitation of our design warrants mention. Across experiments, we swapped out whether a strong or weak list word was present in the study list or served as the related lure. As a result, the study lists in Experiment 2 might have been more potent for producing false recognition than those in Experiment 1, due to greater backward associative strength (BAS; e.g., Roediger et al., 2001b). Despite the slight difference in study list composition, across Experiments 1 and 2, neither the mean BAS of the study lists (0.19 vs. 0.23) nor false recognition of critical lures (0.52 vs. 0.58) differed significantly, $t(38) = 1.40$, $SEM = 0.02$, $p = 0.17$, $p_{BIC} = 0.70$, and $t(130) = 1.61$, $SEM = 0.04$, $p = 0.11$, $p_{BIC} = 0.76$, respectively. Thus, differences in list composition did not reliably affect BAS or subsequent false recognition.

Although signal-detection measures can provide insightful estimates regarding encoding and monitoring, they are not without shortcomings. For one, the measures are only quantitative in nature and can only indicate the relative increase or decrease in encoding and monitoring relative to a read-only control. Discriminability is taken as a metric of the amount of encoded memory information and lambda is as a metric of monitoring, but these indices do not specify how participants implement these processes. For instance, encoded memory information could reflect the amount of gist-based information extracted from the study list (Brainerd and Reyna, 2002) or the strength of the associative network created at study (Roediger et al., 2001a). Likewise, increased monitoring could reflect enhanced

monitoring for the distinctive features presented at study, consistent with diagnostic monitoring (Gallo, 2004) and recollection-rejection processes (Brainerd et al., 2001). Accumulating evidence indicates that participants are able to attribute critical lures to particular tasks (e.g., Hicks and Hancock, 2002; Bodner et al., 2017), indicating that they are monitoring for distinctive details at test, however, additional research is needed to explore how qualitative memory processes map onto these signal-detection indices. Second, both encoding and monitoring are offline estimates computed from hit and false alarm rates. Huff and Aschenbrenner (2018) addressed this limitation by fitting the drift diffusion accumulation model (Ratcliff, 1978) which uses both recognition test responses and response latencies to estimate two latent parameters: drift rate (the rate with which evidence accumulates to make a recognition decision) and boundary separation (the amount of memory evidence needed to make a response). These parameters were used to estimate encoded memory information and monitoring, respectively. When compared to signal-detection indices, the effects of distinctive encoding on drift rate and boundary separation were found to parallel the effects on discriminability and lambda, providing convergent validity that signal-detection indices, at least, partially capture online memory processes.

Finally, distinctive encoding tasks are not likely to be pure with respect to their allowance for item-specific vs. relational processing (Jacoby, 1991). Even though our generation task focused participants on individual anagrams, false recognition of DRM critical lures in the generation groups remained robust in both experiments, indicating that some associative or relational processing of study items persists (Huff et al., 2015b). Although false recognition was lower for related lures than for critical lures, our generation task was unable to eliminate false recognition even for weaker related lures. This observation affirms the dogged nature of associative false recognition: It can be reduced, but it cannot readily be eliminated (Schacter et al., 1999; Benjamin, 2001; McCabe and Smith, 2002).

Conclusion

Given the interest in techniques for reducing false memory in both basic and applied areas, it is important to assess the collective contributions of encoding and retrieval processes to these reductions as well as to potential increases in correct memory. Using the DRM paradigm, our research establishes that distinctive encoding using a generation task can increase correct recognition while simultaneously reducing false recognition of critical lures and other related lures. We found that encoding and monitoring processes appear to operate similarly on both lure types, suggesting that distinctive tasks work to globally disrupt relational encoding while also globally increasing test-based monitoring.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University of Southern Mississippi Institutional Review Board (Protocol #16091503). The patients/participants provided their written informed consent to participate in this study.

REFERENCES

- Arndt, J., and Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *J. Mem. Lang.* 48, 1–15. doi: 10.1016/S0749-596X(02)00518-1
- Balota, D. A., and Neely, J. H. (1980). Test-expectancy and word frequency effects in recall and recognition. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 576–587. doi: 10.1037/0278-7393.6.5.576
- Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 941–947. doi: 10.1037/0278-7393.27.4.941
- Bertsch, S., Pesta, B. J., Wiscott, R., and McDaniel, M. A. (2007). The generation effect: a meta-analytic review. *Mem. Cogn.* 35, 201–210. doi: 10.3758/BF03193441
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: support for a transfer-appropriate processing framework. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 657–668. doi: 10.1037/0278-7393.15.4.657
- Bodner, G. E., Huff, M. J., Lamontagne, R. W., and Azad, T. (2017). Getting at the source of distinctive encoding effects in the DRM paradigm: evidence from signal-detection measures and source judgments. *Memory* 25, 647–655. doi: 10.1080/09658211.2016.1205094
- Brainerd, C. J., Payne, D. G., Wright, R., and Reyna, V. F. (2003). Phantom recall. *J. Mem. Lang.* 48, 445–467. doi: 10.1016/S0749-596X(02)00501-6
- Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Curr. Dir. Psychol. Sci.* 11, 164–169. doi: 10.1111/1467-8721.00192
- Brainerd, C. J., Wright, R., Reyna, V. F., and Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 307–327. doi: 10.1037/0278-7393.27.2.307
- Burns, D. J. (1993). Item gains and losses during hypermnesic recall: implications for item-specific-relational information distinction. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 163–173. doi: 10.1037/0278-7393.19.1.163
- Coane, J. H., and McBride, D. M. (2006). The role of test structure in creating false memories. *Mem. Cogn.* 34, 1026–1036. doi: 10.3758/BF03193249
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Conway, M. A., and Gathercole, S. E. (1987). Modality and long-term memory. *J. Mem. Lang.* 26, 341–361. doi: 10.1016/0749-596X(87)90118-5
- Craik, F. I. M. (2002). Levels of processing: past, present... and future? *Memory* 10, 305–318. doi: 10.1080/09658210244000135
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Dodson, C. S., and Schacter, D. L. (2001). “If I had said it I would have remembered it”: reducing false memories with a distinctiveness heuristic. *Psychon. Bull. Rev.* 8, 155–161. doi: 10.3758/BF03196152
- Faul, F., Erdfelder, E., Lang, A. -G., and Buchner, A. (2007). GPower 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L. 3rd, and Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learn. Mem.* 16, 509–513. doi: 10.1101/lm.1500808
- Foley, M. A., Wozniak, K. H., and Gillum, A. (2006). Imagination and false memory inductions: investigating the role of process, content and source of imaginations. *Appl. Cogn. Psychol.* 20, 1119–1141. doi: 10.1002/acp.1265
- Gallo, D. A. (2004). Using recall to reduce false recognition: diagnostic and disqualifying monitoring. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 120–128. doi: 10.1037/0278-7393.30.1.120

AUTHOR CONTRIBUTIONS

Conceptualization of the research study was conducted by MH and GB. Data collection and analyses were conducted by MH and MG. Initial writing was completed by MH. All authors contributed to the article and approved the submitted version.

- Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York, NY: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Mem. Cogn.* 38, 833–848. doi: 10.3758/MC.38.7.833
- Gallo, D. A., Roediger, H. L. 3rd, and McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychon. Bull. Rev.* 8, 579–589. doi: 10.3758/BF03196194
- Glanzer, M., and Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 5–16. doi: 10.1037/0278-7393.16.1.5
- Gunter, R. W., Bodner, G. E., and Azad, T. (2007). Generation and mnemonic encoding induce a mirror effect in the DRM paradigm. *Mem. Cogn.* 35, 1083–1092. doi: 10.3758/BF03193480
- Hege, A. C. G., and Dodson, C. S. (2004). Why distinctive information reduces false memories: evidence for both impoverished relational encoding and distinctiveness heuristic accounts. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 787–795. doi: 10.1037/0278-7393.30.4.787
- Hicks, J. L., and Hancock, T. W. (2002). Backward associative strength determines source attributions given to false memories. *Psychon. Bull. Rev.* 9, 801–815. doi: 10.3758/BF03196339
- Hockley, W. E., and Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Mem. Cogn.* 24, 202–216. doi: 10.3758/BF03200881
- Huff, M. J., and Aschenbrenner, A. J. (2018). Item-specific processing reduces false recognition in older and younger adults: separating encoding and retrieval using signal-detection and the diffusion model. *Mem. Cogn.* 46, 1287–1301. doi: 10.3758/s13421-018-0837-1
- Huff, M. J., and Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1246–1256. doi: 10.1037/a0031338
- Huff, M. J., and Bodner, G. E. (2019). Item-specific and relational processing both improve recall accuracy in the DRM paradigm. *Q. J. Exp. Psychol.* 72, 1493–1506. doi: 10.1177/1747021818801427
- Huff, M. J., Bodner, G. E., and Fawcett, J. M. (2015a). Effects of distinctive encoding on correct and false memory: a meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychon. Bull. Rev.* 22, 349–365. doi: 10.3758/s13423-014-0648-8
- Huff, M. J., Bodner, G. E., and Gretz, M. R. (in press). Distinctive encoding of a subset of DRM lists yields benefits, but also costs and spillovers. *Psychol. Res.* doi: 10.1007/s00426-019-01241-y
- Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., and Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: effects of initial recall, recognition, and the ironic effect of guessing. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1720–1730. doi: 10.1037/a0028476
- Huff, M. J., McNabb, J., and Hutchison, K. A. (2015b). List blocking and longer retention intervals reveal an influence of gist processing for lexically ambiguous critical lures. *Mem. Cogn.* 43, 1193–1207. doi: 10.3758/s13421-015-0533
- Huff, M. J., Meade, M. L., and Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory* 19, 317–330. doi: 10.1080/09658211.2011.568494
- Huff, M. J., Yates, T. J., and Balota, D. A. (2018). Evaluating the contributions of task expectancy in the testing and guessing benefits on recognition memory. *Memory* 26, 1065–1083. doi: 10.1080/09658211.2018.1467929
- Hunt, R. R., and Einstein, G. O. (1981). Relational and item-specific information in memory. *J. Verbal Learn. Verbal Behav.* 20, 497–514. doi: 10.1016/S0022-5371(81)90138-9

- Hutchison, K. A., and Balota, D. A. (2005). Decoupling semantic and associative information in false memories: explorations with semantically ambiguous and unambiguous critical lures. *J. Mem. Lang.* 52, 1–28. doi: 10.1016/j.jml.2004.08.003
- Israel, L., and Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychon. Bull. Rev.* 4, 577–581. doi: 10.3758/BF03214352
- Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *J. Mem. Lang.* 30, 513–541. doi: 10.1016/0749-596X(91)90025-F
- Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3
- Lampinen, J. M., Neuschatz, J. S., and Payne, D. G. (1999). Source attributions and false memories: a test of the demand characteristics account. *Psychon. Bull. Rev.* 6, 130–135. doi: 10.3758/BF03210820
- MacLeod, C. M., and Bodner, G. E. (2017). The production effect in memory. *Curr. Dir. Psychol. Sci.* 26, 390–395. doi: 10.1177/0963721417691356
- Macmillan, N. A., and Creelman, C. D. (1991). *Detection theory: A user's guide*. New York, NY: Cambridge University Press.
- Marsh, E. J., McDermott, K. B., and Roediger, H. L. 3rd. (2004). Does test-induced priming play a role in the creation of false memories? *Memory* 12, 44–55. doi: 10.1080/09658210244000405
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav. Res. Methods* 43, 679–690. doi: 10.3758/s13428-010-0049-5
- McCabe, D. P., and Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Mem. Cogn.* 30, 1065–1077. doi: 10.3758/BF03194324
- McCabe, D. P., and Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory* 14, 570–583. doi: 10.1080/09658210600624564
- Miller, M. B., Guerin, S. A., and Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1228–1235. doi: 10.1037/a0024539
- Miller, M. B., and Wolford, G. L. (1999). Theoretical commentary: the role of criterion shift in false memory. *Psychol. Rev.* 106, 398–405. doi: 10.1037/0033-295X.106.2.398
- Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *J. Verbal Learn. Verbal Behav.* 16, 519–533. doi: 10.1016/S0022-5371(77)80016-9
- Multhaup, K. S., and Conner, L. A. (2002). The effects of considering nonlist sources on the Deese-Roediger-McDermott memory illusion. *J. Mem. Lang.* 47, 214–228. doi: 10.1016/S0749-596X(02)00007-4
- Nairne, J. S. (2015). “Adaptive memory: novel findings acquired through forward engineering” in *Remembering: Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby*. eds. D. S. Lindsay, C. M. Kelley, A. P. Yonelinas and H. L. Roediger (New York, NY: Psychology Press).
- Nairne, J. S., Thompson, S. R., and Pandeirada, J. N. S. (2007). Adaptive memory: survival processing enhances retention. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 263–273. doi: 10.1037/0278-7393.33.2.263
- Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behav. Res. Methods Instrum. Comput.* 36, 402–407. doi: 10.3758/BF03195588
- Oliver, M. C., Bayes, R. B., and Zabrocky, K. M. (2016). False memories and the DRM paradigm: effects of imagery, list, and test type. *J. Gen. Psychol.* 143, 33–48. doi: 10.1080/00221309.2015.1110558
- Payne, D. G., Elie, C. J., Blackwell, J. M., and Neuschatz, J. S. (1996). Memory illusions: recalling, recognizing, and recollecting events that never occurred. *J. Mem. Lang.* 35, 261–285. doi: 10.1006/JMLA.1996.0015
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59
- Reyna, V. F., Corbin, J. C., Weldon, R. B., and Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *J. Appl. Res. Mem. Cogn.* 5, 1–9. doi: 10.1016/j.jarmac.2015.12.003
- Robin, F. (2010). Imagination and false memories. *Imagin. Cogn. Pers.* 30, 407–424. doi: 10.2190/IC.30.4.e
- Roediger, H. L. 3rd., Balota, D. A., and Watson, J. M. (2001a). “Spreading activation and arousal of false memories” in *The nature of remembering: Essays in honor of Robert G. Crowder*. eds. H. L. Roediger, J. S. Nairne, I. Neath and A. M. Surprenant (Washington, DC: American Psychological Association), 95–115.
- Roediger, H. L. 3rd., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278/7393.21.4.803
- Roediger, H. L. 3rd., Watson, J. M., McDermott, K. B., and Gallo, D. A. (2001b). Factors that determine false recall: a multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407. doi: 10.3758/BF03196177
- Schacter, D. L., Israel, L., and Racine, C. (1999). Suppressing false recognition in younger and older adults: the distinctiveness heuristic. *J. Mem. Lang.* 40, 1–24. doi: 10.1006/jmla.1998.2611
- Slamecka, N. J., and Graf, P. (1978). The generation effect: delineation of a phenomenon. *J. Exp. Psychol. Hum. Learn. Mem.* 4, 592–604. doi: 10.1037/0278-7393.4.6.592
- Smith, R. E., and Hunt, R. R. (2020). When do pictures reduce false memory? *Mem. Cogn.* 48, 623–644. doi: 10.3758/s13421-019-00995-5
- Smithson, M. (2003). *Confidence intervals*. Thousand Oaks, CA: SAGE Publications.
- Wagenmakers, E. -J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3458/BF03194105
- Wickens, T. D., and Hirshman, E. (2000). False memories and statistical decision theory: comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychol. Rev.* 107, 377–383. doi: 10.1037/0033-295X.107.2.377
- Wixted, J. T., and Stretch, V. (2000). The case against a criterion shift account of false memory. *Psychol. Rev.* 107, 368–376. doi: 10.1037/0033-295X.107.2.368
- Zaromb, F. M., and Roediger, H. L. 3rd. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Mem. Cogn.* 38, 995–1008. doi: 10.3758/MC.38.8.995

Conflict of Interest: 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 © 2020 Huff, Bodner and Gretz. 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) and the copyright owner(s) 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.



The Role of Language Proficiency in False Memory: A Mini Review

Mar Suarez[†] and Maria Soledad Beato^{*†}

Faculty of Psychology, University of Salamanca, Salamanca, Spain

OPEN ACCESS

Edited by:

Hui Chen,
Zhejiang University, China

Reviewed by:

Patricia E. Román,
Loyola University Andalusia, Spain
Naveen Kashyap,
Indian Institute of Technology
Guwahati, India

*Correspondence:

Maria Soledad Beato
msol@usal.es

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 27 January 2021

Accepted: 18 March 2021

Published: 09 April 2021

Citation:

Suarez M and Beato MS (2021)
The Role of Language Proficiency
in False Memory: A Mini Review.
Front. Psychol. 12:659434.
doi: 10.3389/fpsyg.2021.659434

Memory errors and, specifically, false memories in the Deese/Roediger–McDermott paradigm have been extensively studied in the past decades. Most studies have investigated false memory in monolinguals' native or first language (L1), but interest has also grown in examining false memories in participants' second language (L2) with different proficiency levels. The main purpose of this manuscript is to review the current state of knowledge on the role of language proficiency on false memories when participants encode and retrieve information in the same language. To do so, a systematic literature search was conducted, and the available studies were reviewed. These studies differed in, for example, age, language proficiency, or material characteristics, including both high and low associative strength lists, and they reported different results. In this review, we attempted to make sense of the apparently contradictory results by carefully identifying participants' language dominance and L2 proficiency. Specifically, the results indicated that, first, people are more prone to produce false memories in their dominant than in their non-dominant language. This result generalizes to lists with high and low associative strength, as well as to participants of different ages. Second, false memories do not differ between two languages when speakers are equally proficient in both languages. Finally, highly proficient L2 speakers produce more false memories in their L2 than speakers with lower L2 proficiency. The results of this review will be considered in the light of the theoretical frameworks of false memories and bilingual language processing.

Keywords: false memories, false recognition, DRM paradigm, bilingualism, language proficiency

INTRODUCTION

In recent years, there has been considerable interest in investigating the malleable nature of human memory. Memory is a reconstructive process that is prone to errors (Kolodner, 1983) and this fact has implications in real-world settings such as eyewitness testimonies (Aizpurua et al., 2009; Havard and Memon, 2013; Loftus, 2018) or the clinical practice (Otgaar et al., 2017; Turk et al., 2020). Among all the potential memory errors, a substantial body of research has focused on understanding false memories, that is, memories for events that did not occur. Specifically, in this review we will focus on a particular type of false memory: the associative memory illusion (see Gallo, 2006, 2010 for review).

False memories have been mostly investigated in monolinguals of very diverse languages, such as Germanic languages (e.g., English, Stadler et al., 1999; Dutch, Van Damme and d'Ydewalle, 2009), Romance languages (e.g., Portuguese, Albuquerque, 2005; Spanish, Beato and Díez, 2011; French, Dubuisson et al., 2012) and Slavic languages (e.g., Polish, Ulatowska and Olszewska, 2013).

Even non-Indo-European languages have been employed to study the associative memory illusion (e.g., Chinese, Chen et al., 2008; Japanese, Kawasaki and Yama, 2006), with a robust false memory effect in all of them. However, false memories have not only been studied in first languages (L1), but also in second languages (L2) with different proficiency levels (e.g., Anastasi et al., 2005; Arndt and Beato, 2017; Beato and Arndt, 2021).

It is important to note that language proficiency may vary with language usage and experience, and, therefore, it is not a constant feature. Bilingualism and monolingualism would be the two ends of a continuum with no clear division between them due to a lack of consensus on the definition of bilingualism (Edwards, 2004) and a high variability in its measurement (Surrain and Luk, 2019). In this review, our purpose is not to define bilingualism, but rather to study false memories across the proficiency continuum on participants with some knowledge of a second language.

Different questions have been investigated in the literature regarding false memories in various languages. First, some research has focused on false memories when languages are switched between encoding and retrieval (i.e., between-language false memory) (see Graves and Altarriba, 2014 for review). Second, another line of research has been interested in whether language and memory processes differ between bilinguals and monolinguals (e.g., Bialystok et al., 2020). Third, and the aim of this review, increasing interest has been centered on whether language proficiency influenced false memories when participants encode and retrieve information in the same language (i.e., within-language false memory). To this aim, we reviewed all the available articles investigating this topic that emerged from a systematic literature search. In particular, our goals were to (1) examine false recognition in the L1 versus L2, centering our attention on language dominance and L2 proficiency, and (2) discuss the findings in terms of the theoretical frameworks.

FALSE MEMORIES: THE DRM PARADIGM

One of the most widely used paradigms to study false memories is the Deese/Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger and McDermott, 1995). In this paradigm participants study lists of words associated to a non-studied word (i.e., critical lure). For example, participants study the words *hot*, *ice*, *snow*, *warm*, *winter*, and *weather*, all of them associated to the critical lure *cold*, based on free association norms (e.g., Nelson et al., 1998). At the test, participants often falsely recall and/or recognize the critical lures as studied items.

The DRM paradigm has been extensively used to study the mechanisms underlying false memories by manipulating variables such as backward and forward associative strength (e.g., Brainerd and Wright, 2005; Arndt, 2012, 2015; Beato and Arndt, 2014, 2017), presentation rate (e.g., Seamon et al., 1998; Smith and Kimball, 2012; Sadler et al., 2018), number of words associated to the critical lure (e.g., Arndt, 2010; Flegal and Reuter-Lorenz, 2014), presentation modality (e.g., Mao et al., 2010; Boldini et al., 2013), retrieval time (e.g., Giammattei

and Arndt, 2012; Carneiro et al., 2014), attentional demands (e.g., Pérez-Mata et al., 2002; Otgaar et al., 2012), distinctive encoding (e.g., Huff et al., 2015, 2020), warning instructions (e.g., Watson et al., 2004; Carneiro and Fernandez, 2010; Coane et al., 2016), identifiability of the critical lure (e.g., Neuschatz et al., 2003; Carneiro et al., 2009; Beato and Cadavid, 2016), or emotional valence (e.g., Bookbinder and Brainerd, 2016; Hellenthal et al., 2019; Chang et al., 2020), among many others. All these experimental manipulations confirmed the robustness of this paradigm to produce false memories.

Furthermore, the DRM paradigm has also been employed to study false memories in different clinical populations such as patients with schizophrenia (e.g., Bhatt et al., 2010; Favre et al., 2020), Alzheimer's disease (e.g., Malone et al., 2019; Howe and Akhtar, 2020), or autism spectrum disorder (e.g., Wojcik et al., 2018). Additionally, it has been used throughout development in children (e.g., Carneiro et al., 2007; Brainerd et al., 2008; Knott et al., 2011), and older adults (e.g., McCabe et al., 2009; Devitt and Schacter, 2016).

Not only behavioral research has been conducted on false memories. Some efforts have also been made to identify the neural correlates of false memories (see Schacter and Slotnick, 2004 for review) using techniques such as functional magnetic resonance imaging (e.g., Abe et al., 2013), event-related potentials (e.g., Curran et al., 2001; Beato et al., 2012; Cadavid and Beato, 2016), near infrared spectroscopy (e.g., Kubota et al., 2006), positron emission tomography (Schacter et al., 1996), or transcranial direct current stimulation (e.g., Díez et al., 2017).

The two main theoretical explanations of the false memory effect in the DRM paradigm are the fuzzy-trace theory (FFT; Reyna and Brainerd, 1995; Brainerd and Reyna, 2002) and the activation-monitoring framework (AMF; Roediger et al., 2001) (see also global matching models, Arndt and Hirshman, 1998). According to FFT, two types of information are encoded during the study of the DRM lists: verbatim traces (i.e., perceptual features of the event) and gist traces (i.e., meaning-based information of the event). If gist memory traces are retrieved, false memories are more likely to occur because critical lures tend to match the meaning information extracted from the list. Therefore, the critical lure's meaning would be familiar. That familiarity triggered by the critical lure might be countervailed by retrieving verbatim traces of the studied items, a process referred to as recollection rejection (Brainerd et al., 2003), that would reduce false memories. For its part, AMF suggests that false memories are produced by the combination of activation and monitoring processes. When a DRM list is studied these words are activated and the activation is spread throughout the semantic network to associatively related words, namely, the critical lure, increasing the likelihood to produce false memories. In order to counteract that activation, monitoring processes might be engaged. Monitoring processes are defined as decision processes that use different types of information to determine the source of the activation and so false memories can be reduced (Gallo, 2006). As will be discussed later, both theories (i.e., FFT and AMF) could potentially explain the results of the present review, but in slightly different ways.

TWO LANGUAGES IN ONE BRAIN

A central question in bilingual research is how two languages are represented in one brain (Heredia and Brown, 2006). Various models have been developed to seek an answer (e.g., bilingual interactive activation model, Dijkstra and Van Heuven, 2002; inhibitory control model, Green, 1998; distributed feature model, Van Hell and De Groot, 1998; revised hierarchical model, Kroll and Stewart, 1994) and, despite differing in the exact nature of L1 and L2 representations, these models share a consensual view about two assumptions relevant for the present review. First, both languages access a shared conceptual system (Francis, 1999, 2020; Francis et al., 2019) and, second, associations between word forms and their concepts are stronger in L1 than in L2 (e.g., Gollan et al., 2008). To further elaborate on these ideas, the revised hierarchical model (RHM)¹, referred above, will be considered.

The RHM assumes two different levels of representation, the lexical and conceptual level, with independent lexical representations for each language and a shared conceptual store. First, at the lexical level, although both languages are stored independently, they are interconnected with stronger connections from L2 to L1 than from L1 to L2 (Kroll et al., 2002). The explanation is found in the fact that the L2 is acquired by creating links between L2 words and the correspondent L1 translation at the lexical level, leading to stronger connection from L2 to L1. By contrast, links from L1 to L2 are weaker due to a lack of translation practice in that direction (Kroll and Stewart, 1994). Second, at the conceptual level, the links between words and concepts (i.e., conceptual links) are assumed to be stronger in L1 than L2. This means that the concept store is fully activated quicker from L1 than from L2 lexical representations. Nonetheless, once L2 learners become more proficient, the conceptual links from L2 words to the concepts become stronger (Perea et al., 2008).

Although the RHM interprets that the first language acquired (L1) is the dominant language, it is noteworthy that bilingual memory is a dynamic system influenced by language usage (Heredia and Altarriba, 2001). Thus, as Heredia (1997) suggested, the L1 might lose strength while the L2 might become the dominant language as a function of exposure, hence the L1 and L2 should be interpreted as the dominant and non-dominant language, respectively, disregarding which language was learned first.

FALSE MEMORIES AND LANGUAGE PROFICIENCY

In this review, we investigated the role of language proficiency in false memory when information was encoded and retrieved in L1 versus L2 (i.e., within-language conditions)². To this end, we identified participants' dominant and non-dominant language and their L2-proficiency level. This was crucial to

understand and discuss the different results, but it was not an easy task. Specifically, regarding language dominance, sometimes it was difficult to identify which language was dominant (not necessarily the L1) based on the available information. Furthermore, regarding language proficiency, it was difficult to compare this variable across studies due to, first, different facets of bilingual experience being reported (e.g., usage or years of academic training; see Surrain and Luk, 2019). Second, there were differences in language proficiency (see **Table 1**), with some studies employing highly proficient bilinguals that used both languages in everyday life (e.g., Cabeza and Lennartson, 2005), while others included participants whose only L2 experience was in a classroom setting (e.g., Arndt and Beato, 2017). Third, although most of the studies included young adults ($M = 24.04$ years across experiments), children were also tested (Howe et al., 2008), possibly leading to age related differences in L1 and L2 proficiency.

Focusing now on the results, when comparing false memory in the L1 and L2, some studies found $L1 > L2$, others $L1 < L2$, or even, $L1 = L2$ (see **Table 1**). That is, although, to our knowledge, only eight works have investigated this topic, all possible results have been reported. However, as referred above, language dominance can make sense of the apparently contradictory results. In other words, if we compare false memory in the dominant and non-dominant languages, instead of considering the order of language acquisition (i.e., L1 versus L2), consistent conclusions can be drawn. To further elaborate on this idea, the three patterns of results observed in the reviewed studies will be explained below.

First, beginning with the most common result, the studies that found significantly higher false recognition in L1 than L2 (Anastasi et al., 2005, Experiments 3 and 4; Sahlin et al., 2005; Howe et al., 2008; Arndt and Beato, 2017; Beato and Arndt, 2021) reported that the dominant and non-dominant language were the L1 and L2, respectively. That is, false recognition was higher in the dominant than in the non-dominant language in these studies. Kawasaki-Miyaji et al. (2003) and Marmolejo et al. (2009) seemingly point in the same direction, with higher false recognition in L1 (dominant) than in L2 (non-dominant) (0.71 versus 0.62, and 0.80 versus 0.73, respectively), although they did not directly test this comparison statistically.

This pattern of results has also been found in 6-, 8-, and 12-year-old children (Howe et al., 2008). This study not only showed that false recognition increased with age in both languages, but also that all age groups were more likely to produce false recognition in L1 (dominant) than in L2 (non-dominant). Furthermore, the effect of language dominance on false memory was obtained in most of the studies using DRM lists strongly related to the critical lure, but Beato and Arndt (2021) found this effect with lists weakly related to the critical lure. Namely, higher false memories were reported in the dominant than in the non-dominant language in both adults and children, and with DRM lists that had high and low associative strength between the studied words and the critical lure.

Second, a study showed higher false recognition in L2 than L1 (Anastasi et al., 2005, Experiment 2). In this case, we can consider the L2 as the dominant language, since most of the participants frequently used this language at work (75%) and with friends

¹As some concerns have been raised regarding the RHM (e.g., Brysbaert and Duyck, 2010), see Kroll et al. (2010) for a critical review and assessment of this model.

²Note that some reviewed studies also included other experimental conditions.

TABLE 1 | Summary of the reviewed studies on the role of language proficiency in false recognition.

Authors, year		Languages		Participants		Language proficiency and background	Results: false recognition		
		L1	L2	No.	Age (M)		L1	L2	Conclusion
Kawasaki-Miyaji et al., 2003		Japanese	English	74	University students (N/A)	L1: dominant language L2: 7 years of academic training Participants lived in Japan	0.71 ¹	0.62 ¹	L1 > L2 ²
Anastasi et al., 2005	Exp 1	Japanese, Spanish, German	English	12	University students (N/A)	L1: dominant language L2: 8.51 years of experience/academic training Participants were exchange students in the United States	0.55	0.49	L1 = L2 ³
	Exp 2	Spanish	English	22	Young adults (30.70)	L1: 75% at home, 45% at work, and 60% with friends L2 (dominant language): 50% at home, 75% at work, and 80% with friends ⁴ Participants lived in the United States	0.52	0.70	L1 < L2 ³
	Exp 3	Spanish	English	20	Young adults (29.70)	L1 (dominant language): 100% at home, work, and with friends L2: living in an L2 speaking country Participants lived in the United States	0.59	0.44	L1 > L2 ³
	Exp 4	English	Spanish	24	University students (N/A)	L1 (dominant language): 100% at home, work, and with friends L2: no formal instruction Participants lived in the United States	0.68	0.16	L1 > L2 ³
Cabeza and Lennartson, 2005		English	French	30	University students (N/A)	L1 and L2: high proficiency and used in everyday life Participants lived in Canada	0.41	0.40	L1 = L2
Sahlin et al., 2005		English	Spanish	20	University students (20.00)	L1 (dominant language): proficiency self-report = 5/5 L2: proficiency self-report = 4.55/5 Participants lived in the United States	0.75 ⁵	0.62 ⁵	L1 > L2 ³
Howe et al., 2008		English	French	40 32 30 20	6 years old 8 years old 12 years old 20 years old	L1 (children and adults): dominant language L2 (children): L2-immersion school. This was the only L2-speaking environment L2 (adults): participants were formally studying French All lived in an L1 community in Canada	0.74 ¹	0.61 ¹	L1 > L2

(Continued)

TABLE 1 | Continued

Authors, year		Languages		Participants		Language proficiency and background	Results: false recognition		
		L1	L2	No.	Age (M)		L1	L2	Conclusion
Marmolejo et al., 2009		English	Spanish	60	University students (20.63)	L1 (dominant language): proficiency self-report = 9.35/10 L2: proficiency self-report = 8.40/10 Participants lived in the United States	0.80	0.73	L1 > L2 ²
Arndt and Beato, 2017	Exp 1	English	Spanish	28	University students (19.75)	L1: dominant language L2: participants were formally studying a third-term Spanish course. Proficiency self-report = 6.32/10 Participants lived in the United States	0.29	0.15	L1 > L2
	Exp 2	Spanish	English	156	University students (22.48)	L1: dominant language L2: studied on primary and secondary school. Proficiency self-report = 5.25/10 Participants lived in Spain	0.30	0.14	L1 > L2
	Exp 3	Spanish	English	52	University students (25.69)	L1: dominant language L2: participants were formally studying English, 26 at elementary (low) level and 26 at advanced (high) level. Proficiency self-report: low = 4.31/10, high = 7/10 Participants lived in Spain	0.35	0.18 High: 0.22 Low: 0.13	L1 > L2 High > Low
Beato and Arndt, 2021	Exp 1	Spanish	English	90	University students (21.76)	L1: dominant language L2: studied on primary and secondary school. Proficiency self-report = 6.02/10 Participants lived in Spain	0.33	0.21	L1 > L2
	Exp 2	Spanish	English	164	Young adults (29.69)	L1: dominant language L2: participants were formally studying English, 58 at elementary (low) level, 59 intermediate (mid) level, and 47 advanced (high) level. Proficiency self-report: low = 5.36/10, mid = 6.54/10, high = 6.89/10 Participants lived in Spain	0.28	0.16 High: 0.19 Mid: 0.12 Low: 0.12	L1 > L2 High > Mid High > Low Mid = Low

False recognition rates are reported as mean proportions. L1, first language; L2, second language; N/A, not available.

¹Means were provided by the first author in Kawasaki-Miyaji et al. (2003) and estimated from Figure 4 in Howe et al. (2008).

²The comparison was not tested statistically.

³Analyses were conducted on corrected scores in Anastasi et al. (2005) and sensitivity scores in Sahlin et al. (2005).

⁴Some participants reported using both languages in various environments.

⁵Only means for the first study-test trial are reported to be comparable to the other studies.

(80%), and even half of participants used it at home. Therefore, here we can also conclude that false memories were higher in the dominant than in the non-dominant language.

Third, we identified two studies where false recognition was similar in the L1 and L2 (Anastasi et al., 2005, Experiment 1; Cabeza and Lennartson, 2005). In these cases, L1 and L2 proficiency seem to be similar. Specifically, Cabeza and Lennartson (2005) reported highly proficient speakers that used both languages in everyday life with no dominance difference specified. For their part, Anastasi et al. (2005, Experiment 1) included participants whose dominant language seems to be the L1, but it is also important to consider that participants had several years of L2 academic training, besides living in the L2-speaking country. As L2 acquisition and its associated brain changes are highly related to the amount of L2 immersion (Pliatsikas et al., 2017), it is reasonable to think that, in this case, participants could reach high levels of L1 and L2 proficiency at the moment of testing. Thus, in these two studies, we would expect false recognition not to differ significantly between L1 and L2 (since L1 and L2 proficiency would be similar), and this was exactly the result found in both experiments.

Lastly, two studies investigated false memories in participants that differed in L2 proficiency (Arndt and Beato, 2017; Beato and Arndt, 2021). Specifically, the authors found that greater language proficiency in the non-dominant language increased false memories. This result is in line with the above reported greater false recognition in the dominant (high proficiency, in these studies) than in the non-dominant (low proficiency, in these studies) language.

DISCUSSION

The reviewed studies suggest that, regardless of age and the associative strength of the lists, false memories are higher in participants' dominant language than in their non-dominant one, just as false memories are greater in high than low L2 proficiency participants. Only when proficiency in the L1 and L2 is similar, false memories do not differ. These results could be accommodated by theoretical accounts from very different research areas, false memory and bilingual language processing.

Regarding the theoretical framework of the false memory effect, both the AMF and the FFT, mentioned above, could explain the current data despite claiming different mechanisms underlying the effect. On the one hand, according to FFT, the extraction of gist representations improves throughout development because participants become better in processing word meaning and connecting meaning across different words (Brainerd and Reyna, 2002). Given the parallel between how false memory differ across the proficiency continuum and the developmental trajectory of false memory (e.g., Carneiro and Fernandez, 2010; Arndt and Beato, 2017), this prediction can be used to explain the present results. That is, gist memory would be hindered when processing L2-words as participants have less experience, leading to a decrease in false memories. Along the same lines, gist memory improves when participants become more proficient, explaining why high L2 proficiency speakers show greater false memories than low L2 proficiency

speakers. On the other hand, the activation processes referred by the AMF could explain the findings reviewed above by arguing that concepts are more automatically activated by the dominant than the non-dominant language, or even by high L2-proficiency participants. This activation spread throughout a well-organized network with strong connections to associatively related words (i.e., critical lure), which in turn would produce higher false memories in the dominant language than in the non-dominant one, as well as in high rather than low L2 proficiency speakers.

Within the bilingual language processing research, the greater false memories in the dominant than non-dominant language (e.g., Sahlin et al., 2005) could be accommodated by the RHM (e.g., Kroll and Stewart, 1994) since this model proposes stronger conceptual links in L1 than in L2. Furthermore, this model also assumes that the conceptual links in the L1 and L2 would have similar strength if proficiency in both languages is similar, predicting that false memories will not differ between L1 and L2 (e.g., Cabeza and Lennartson, 2005). Finally, as L2 proficiency increases, this theory suggests that the links between L2 words and their concepts strengthened, which predicts higher false memories for higher L2 proficiency participants (e.g., Arndt and Beato, 2017). With an increase in L2 proficiency as a function of language usage (Heredia, 1997), the L2 can even come to be the dominant language, in this case expecting higher false recognition in L2 than in L1 (e.g., Anastasi et al., 2005, Experiment 2).

In conclusion, this review has demonstrated that the DRM paradigm is useful to deepen our understanding of language and memory processes in speakers with knowledge of more than one language. Moreover, this review highlights the importance of language dominance to understand the production of false memories in the L1 and the L2. Therefore, we believe that it is crucial to assess language proficiency and exhaustively report participants' language backgrounds on research that included more than one language. To do so some questionnaires have been created (e.g., Li et al., 2006; Marian et al., 2007; Luk and Bialystok, 2013; Anderson et al., 2018) that might be useful for future research. Additionally, as previous works showed that participants had far from perfect knowledge of L2 stimuli (Beato and Arndt, 2021), we encourage researchers to evaluate L2 word knowledge within future studies to assess the validity of alternative explanations for memory effects.

After reviewing the available articles investigating false memories in L1 versus L2, an issue that still seems unclear is whether participants translated L2 words during task performance (see Graves and Altarriba, 2014). Thus, further research could clarify this issue by manipulating the presentation rate of studied items or the time available during retrieval. Furthermore, researchers interested in measuring brain electrical activity need to describe the neural correlates of false recognition in the L1 and the L2 as, to our knowledge, no previous research has examined this matter. Additionally, it would be interesting to know whether false memories in the DRM paradigm differ between monolinguals and bilinguals. Although this issue is beyond the scope of this review, it would add valuable information to our understanding of language and memory processes in bilingual speakers.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Abe, N., Fujii, T., Suzuki, M., Ueno, A., Shigemune, Y., Mugikura, S., et al. (2013). Encoding- and retrieval-related brain activity underlying false recognition. *Neurosci. Res.* 76, 240–250. doi: 10.1016/j.neures.2013.05.006
- Aizpuru, A., Garcia-Bajos, E., and Migueles, M. (2009). False memories for a robbery in young and older adults. *Appl. Cogn. Psychol.* 23, 174–187. doi: 10.1002/acp
- Albuquerque, P. B. (2005). Produção de evocações e reconhecimentos falsos em 100 listas de palavras associadas portuguesas [False recall and recognition production in 100 Portuguese associate word list]. *Lab. Psicol.* 3, 3–12. doi: 10.14417/lp.766
- Anastasi, J. S., Rhodes, M. G., Marquez, S., and Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory* 13, 815–828. doi: 10.1080/09658210444000421
- Anderson, J. A. E., Mak, L., Keyvani Chahi, A., and Bialystok, E. (2018). The language and social background questionnaire: assessing degree of bilingualism in a diverse population. *Behav. Res. Methods* 50, 250–263. doi: 10.3758/s13428-017-0867-9
- Arndt, J. (2010). The role of memory activation in creating false memories of encoding context. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 66–79. doi: 10.1037/a0017394
- Arndt, J. (2012). The influence of forward and backward associative strength on false recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 747–756. doi: 10.1037/a0026375
- Arndt, J. (2015). The influence of forward and backward associative strength on false memories for encoding context. *Memory* 23, 1093–1111. doi: 10.1080/09658211.2014.959527
- Arndt, J., and Beato, M. S. (2017). The role of language proficiency in producing false memories. *J. Mem. Lang.* 95, 146–158. doi: 10.1016/j.jml.2017.03.004
- Arndt, J., and Hirshman, E. (1998). True and false recognition in MINERVA2: explanations from a global matching perspective. *J. Mem. Lang.* 39, 371–391. doi: 10.1006/jmla.1998.2581
- Beato, M. S., and Arndt, J. (2014). False recognition production indexes in forward associative strength (FAS) lists with three critical words. *Psicothema* 26, 457–463. doi: 10.7334/psicothema2014.79
- Beato, M. S., and Arndt, J. (2017). The role of backward associative strength in false recognition of DRM lists with multiple critical words. *Psicothema* 29, 358–363. doi: 10.7334/psicothema2016.248
- Beato, M. S., and Arndt, J. (2021). The effect of language proficiency and associative strength on false memory. *Psychol. Res.* [Epub ahead of print] doi: 10.1007/s00426-020-01449-3
- Beato, M. S., Boldini, A., and Cadavid, S. (2012). False memory and level of processing effect: an event-related potential study. *NeuroReport* 23, 804–808. doi: 10.1097/WNR.0b013e32835734de
- Beato, M. S., and Cadavid, S. (2016). Normative study of theme identifiability: instructions with and without explanation of the false memory effect. *Behav. Res. Methods* 48, 1252–1265. doi: 10.3758/s13428-015-0652-6
- Beato, M. S., and Díez, E. (2011). False recognition production indexes in Spanish for 60 DRM lists with three critical words. *Behav. Res. Methods* 43, 499–507. doi: 10.3758/s13428-010-0045-9
- Bhatt, R., Laws, K. R., and McKenna, P. J. (2010). False memory in schizophrenia patients with and without delusions. *Psychiatry Res.* 178, 260–265. doi: 10.1016/j.psychres.2009.02.006
- Bialystok, E., Dey, A., Sullivan, M. D., and Sommers, M. S. (2020). Using the DRM paradigm to assess language processing in monolinguals and bilinguals. *Mem. Cogn.* 48, 870–883. doi: 10.3758/s13421-020-01016-6
- Boldini, A., Beato, M. S., and Cadavid, S. (2013). Modality-match effect in false recognition: an event-related potential study. *NeuroReport* 24, 108–113. doi: 10.1097/WNR.0b013e32835c93e3
- Bookbinder, S. H., and Brainerd, C. J. (2016). Emotion and false memory: the context–content paradox. *Psychol. Bull.* 142, 1315–1351. doi: 10.1037/bul0000077
- Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Curr. Direct. Psychol. Sci.* 11, 164–169. doi: 10.1111/1467-8721.00192
- Brainerd, C. J., Reyna, V. F., and Ceci, S. J. (2008). Developmental reversals in false memory: a review of data and theory. *Psychol. Bull.* 134, 343–382. doi: 10.1037/0033-2909.134.3.343
- Brainerd, C. J., Reyna, V. F., Wright, R., and Mojardin, A. H. (2003). Recollection rejection: false-memory editing in children and adults. *Psychol. Rev.* 110, 762–784. doi: 10.1037/0033-295X.110.4.762
- Brainerd, C. J., and Wright, R. (2005). Forward association, backward association, and the false-memory illusion. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 554–567. doi: 10.1037/0278-7393.31.3.554
- Brysbaert, M., and Duyck, W. (2010). Is it time to leave behind the revised hierarchical model of bilingual language processing after fifteen years of service? *Bilingual. Lang. Cogn.* 13, 359–371. doi: 10.1017/S1366728909990344
- Cabeza, R., and Lennartson, E. R. (2005). False memory across languages: implicit associative response vs fuzzy trace views. *Memory* 13, 1–5. doi: 10.1080/09658210344000161
- Cadavid, S., and Beato, M. S. (2016). Memory distortion and its avoidance: an event-related potentials study on false recognition and correct rejection. *PLoS One* 11:e0164024. doi: 10.1371/journal.pone.0164024
- Carneiro, P., Albuquerque, P., Fernandez, A., and Esteves, F. (2007). Analyzing false memories in children with associative lists specific for their age. *Child Dev.* 78, 1171–1185. doi: 10.1111/j.1467-8624.2007.01059.x
- Carneiro, P., and Fernandez, A. (2010). Age differences in the rejection of false memories: the effects of giving warning instructions and slowing the presentation rate. *J. Exp. Child Psychol.* 105, 81–97. doi: 10.1016/j.jecp.2009.09.004
- Carneiro, P., Fernandez, A., and Dias, A. R. (2009). The influence of theme identifiability on false memories: evidence for age-dependent opposite effects. *Mem. Cogn.* 37, 115–129. doi: 10.3758/MC.37.2.115
- Carneiro, P., Garcia-Marques, L., Fernandez, A., and Albuquerque, P. (2014). Both associative activation and thematic extraction count, but thematic false memories are more easily rejected. *Memory* 22, 1024–1040. doi: 10.1080/09658211.2013.864680
- Chang, M., Brainerd, C. J., Toglia, M. P., and Schmidt, S. R. (2020). Norms for emotion–false memory lists. *Behav. Res. Methods* 53, 96–112. doi: 10.3758/s13428-020-01410-7
- Chen, J. C. W., Li, W., Westerberg, C. E., and Tzeng, O. J.-L. (2008). Test-item sequence affects false memory formation: an event-related potential study. *Neurosci. Lett.* 431, 51–56. doi: 10.1016/j.neulet.2007.11.020
- Coane, J. H., Huff, M. J., and Hutchison, K. A. (2016). The ironic effect of guessing: increased false memory for mediated lists in younger and older adults. *Aging Neuropsychol. Cogn.* 23, 282–303. doi: 10.1080/13825585.2015.1088506
- Curran, T., Schacter, D. L., Johnson, M. K., and Spinks, R. (2001). Brain potentials reflect behavioral differences in true and false recognition. *J. Cogn. Neurosci.* 13, 201–216. doi: 10.1162/089892901564261
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Devitt, A. L., and Schacter, D. L. (2016). False memories with age: neural and cognitive underpinnings. *Neuropsychologia* 91, 346–359. doi: 10.1016/j.neuropsychologia.2016.08.030
- Díez, E., Gómez-Ariza, C. J., Díez-Álamo, A. M., Alonso, M. A., and Fernandez, A. (2017). The processing of semantic relatedness in the brain: evidence from associative and categorical false recognition effects following transcranial direct current stimulation of the left anterior temporal lobe. *Cortex* 93, 133–145. doi: 10.1016/j.cortex.2017.05.004

FUNDING

This work was partially supported by grant SA052G18, Junta de Castilla y León, and by the University of Salamanca.

- Dijkstra, T., and Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingual. Lang. Cogn.* 5, 175–197. doi: 10.1017/s1366728902003012
- Dubuisson, J. B., Fiori, N., and Nicolas, S. (2012). Repetition and spacing effects on true and false recognition in the DRM paradigm. *Scand. J. Psychol.* 53, 382–389. doi: 10.1111/j.1467-9450.2012.00963.x
- Edwards, J. (2004). “Foundations of bilingualism,” in *The Handbook of Bilingualism*, eds T. K. Bhatia and W. C. Ritchie (Hoboken, NJ: Blackwell Publishing), 7–31.
- Favre, G., Horat, S. K., Herrmann, F. R., Gothuey, I., Ventura, J., Merlo, M. C. G., et al. (2020). False memory production in schizophrenia: a neurophysiological investigation. *Schizophren. Res. Cogn.* 20, 100174–100179. doi: 10.1016/j.scog.2020.100174
- Flegel, K. E., and Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Mem. Cogn.* 42, 701–711. doi: 10.3758/s13421-013-0391-9
- Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: semantic representation. *Psychol. Bull.* 125, 193–222. doi: 10.1037/0033-2909.125.2.193
- Francis, W. S. (2020). Shared core meanings and shared associations in bilingual semantic memory: evidence from research on implicit memory. *Int. J. Bilingual.* 24, 464–477. doi: 10.1177/1367006918814375
- Francis, W. S., Strobach, E. N., Penalver, R. M., Martínez, M., Gurrola, B. V., and Soltero, A. (2019). Word-context associations in episodic memory are learned at the conceptual level: word frequency, bilingual proficiency, and bilingual status effects on source memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 1852–1871. doi: 10.1037/xlm0000678
- Gallo, D. A. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. Hove: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Mem. Cogn.* 38, 833–848. doi: 10.3758/MC.38.7.833
- Giammattei, J., and Arndt, J. (2012). Hemispheric asymmetries in the activation and monitoring of memory errors. *Brain Cogn.* 80, 7–14. doi: 10.1016/j.bandc.2012.04.003
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Graves, D. F., and Altarriba, J. (2014). “False memories in bilingual speakers,” in *Foundations of Bilingual Memory*, eds R. R. Heredia and J. Altarriba (Berlin: Springer), 205–221.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingual. Lang. Cogn.* 1, 67–81. doi: 10.1017/s1366728998000133
- Havard, C., and Memon, A. (2013). The mystery man can help reduce false identification for child witnesses: evidence from video line-ups. *Appl. Cogn. Psychol.* 27, 50–59. doi: 10.1002/acp.2870
- Hellenthal, M. V., Knott, L. M., Howe, M. L., Wilkinson, S., and Shah, D. (2019). The effects of arousal and attention on emotional false memory formation. *J. Mem. Lang.* 107, 54–68. doi: 10.1016/j.jml.2019.03.010
- Heredia, R. R. (1997). Bilingual memory and hierarchical models: a case for language dominance. *Curr. Direct. Psychol. Sci.* 6, 34–39. doi: 10.1111/1467-8721.ep11512617
- Heredia, R. R., and Altarriba, J. (2001). Bilingual language mixing: why do bilinguals code-switch? *Curr. Direct. Psychol. Sci.* 10, 164–168. doi: 10.1111/1467-8721.00140
- Heredia, R. R., and Brown, J. M. (2006). “Bilingual memory,” in *The Handbook of Bilingualism*, eds T. K. Bhatia and W. C. Ritchie (Hoboken, NJ: Blackwell Publishing), 225–249.
- Howe, M. L., and Akhtar, S. (2020). Priming older adults and people with mild to moderate Alzheimer’s disease problem-solving with false memories. *Cortex* 125, 318–331. doi: 10.1016/j.cortex.2020.01.014
- Howe, M. L., Gagnon, N., and Thouas, L. (2008). Development of false memories in bilingual children and adults. *J. Mem. Lang.* 58, 669–681. doi: 10.1016/j.jml.2007.09.001
- Huff, M. J., Bodner, G. E., and Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: a meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychon. Bull. Rev.* 22, 349–365. doi: 10.3758/s13423-014-0648-8
- Huff, M. J., Bodner, G. E., and Gretz, M. R. (2020). Reducing false recognition in the Deese-Roediger/McDermott paradigm: related lures reveal how distinctive encoding improves encoding and monitoring processes. *Front. Psychol.* 11:602347. doi: 10.3389/fpsyg.2020.602347
- Kawasaki, Y., and Yama, H. (2006). The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory* 14, 68–78. doi: 10.1080/09658210444000520
- Kawasaki-Miyaji, Y., Inoue, T., and Yama, H. (2003). Cross-linguistic false recognition: how do Japanese-dominant bilinguals process two languages: Japanese and English? *Psychologia* 46, 255–267. doi: 10.2117/psysoc.2003.255
- Knott, L. M., Howe, M. L., Wimmer, M. C., and Dewhurst, S. A. (2011). The development of automatic and controlled inhibitory retrieval processes in true and false recall. *J. Exp. Child Psychol.* 109, 91–108. doi: 10.1016/j.jecp.2011.01.001
- Kolodner, J. L. (1983). Reconstructive memory: a computer model. *Cogn. Sci.* 7, 281–328. doi: 10.1016/S0364-0213(83)80002-0
- Kroll, J. F., Michael, E., Tokowicz, N., and Dufour, R. (2002). The development of lexical fluency in a second language. *Sec. Lang. Res.* 18, 137–171. doi: 10.1191/0267658302sr201oa
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1006/jmla.1994.1008
- Kroll, J. F., Van Hell, J. G., Tokowicz, N., and Green, D. W. (2010). The revised hierarchical model: a critical review and assessment. *Bilingual. Lang. Cogn.* 13, 373–381. doi: 10.1017/S136672891000009X
- Kubota, Y., Toichi, M., Shimizu, M., Mason, R. A., Findling, R. L., Yamamoto, K., et al. (2006). Prefrontal hemodynamic activity predicts false memory-A near-infrared spectroscopy study. *NeuroImage* 31, 1783–1789. doi: 10.1016/j.neuroimage.2006.02.003
- Li, P., Sepanski, S., and Zhao, X. (2006). Language history questionnaire: a Web-based interface for bilingual research. *Behav. Res. Methods* 38, 202–210. doi: 10.3758/BF03192770
- Loftus, E. F. (2018). Eyewitness science and the legal system. *Annu. Rev. Law Soc. Sci.* 14, 1–10. doi: 10.1146/annurev-lawsocsci-101317-030850
- Luk, G., and Bialystok, E. (2013). Bilingualism is not a categorical variable: interaction between language proficiency and usage. *J. Cogn. Psychol.* 25, 605–621. doi: 10.1080/20445911.2013.795574
- Malone, C., Deason, R. G., Palumbo, R., Heyworth, N., Tat, M., and Budson, A. E. (2019). False memories in patients with mild cognitive impairment and mild Alzheimer’s disease dementia: can cognitive strategies help? *J. Clin. Exp. Neuropsychol.* 41, 204–218. doi: 10.1080/13803395.2018.1513453
- Mao, W. B., Yang, Z. L., and Wang, L. S. (2010). Modality effect in false recognition: evidence from Chinese characters. *Int. J. Psychol.* 45, 4–11. doi: 10.1080/00207590902757641
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marmolejo, G., Deliberto-Macaluso, K. A., and Altarriba, J. (2009). False memory in bilinguals: does switching languages increase false memories? *Am. J. Psychol.* 122, 1–16.
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., and Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering: neuropsychological test correlates of remember-know judgments. *Neuropsychologia* 47, 2164–2173. doi: 10.1016/j.neuropsychologia.2008.11.025
- Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. *Behav. Res. Methods Instru. Comp.* 36, 402–407.
- Neuschatz, J. S., Benoit, G. E., and Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: the role of identifiability. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 35–41. doi: 10.1037/0278-7393.29.1.35
- Otgaar, H., Muris, P., Howe, M. L., and Merckelbach, H. (2017). What drives false memories in psychopathology? A case for associative activation. *Clin. Psychol. Sci.* 5, 1048–1069. doi: 10.1177/2167702617724424
- Otgaar, H., Peters, M., and Howe, M. L. (2012). Dividing attention lowers children’s but increases adults’ false memories. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 204–210. doi: 10.1037/a0025160

- Perea, M., Duñabeitia, J. A., and Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *J. Mem. Lang.* 58, 916–930. doi: 10.1016/j.jml.2008.01.003
- Pérez-Mata, M. N., Read, J. D., and Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory* 10, 161–177. doi: 10.1080/09658210143000308
- Pliatsikas, C., DeLuca, V., Moschopoulou, E., and Saddy, J. D. (2017). Immersive bilingualism reshapes the core of the brain. *Brain Struct. Funct.* 222, 1785–1795. doi: 10.1007/s00429-016-1307-9
- Reyna, V. F., and Brainerd, C. J. (1995). Fuzzy-trace theory: an interim synthesis. *Learn. Individ. Differ.* 7, 1–75. doi: 10.1016/1041-6080(95)90031-4
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, H. L., Watson, J. M., McDermott, K. B., and Gallo, D. A. (2001). Factors that determine false recall: a multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407. doi: 10.3758/BF03196177
- Sadler, D. D., Sodmont, S. M., and Keefer, L. A. (2018). Can false memory for critical lures occur without conscious awareness of list words? *Conscious. Cogn.* 58, 136–157. doi: 10.1016/j.concog.2017.10.018
- Sahlin, B. H., Harding, M. G., and Seamon, J. G. (2005). When do false memories cross language boundaries in English–Spanish bilinguals? *Mem. Cogn.* 33, 1414–1421. doi: 10.3758/BF03193374
- Schacter, D. L., Reiman, E., Curran, T., Yun, L. S., Bandy, D., McDermott, K. B., et al. (1996). Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. *Neuron* 17, 267–274. doi: 10.1016/S0896-6273(00)80158-0
- Schacter, D. L., and Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron* 44, 149–160. doi: 10.1016/j.neuron.2004.08.017
- Seamon, J. G., Luo, C. R., and Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: evidence for nonconscious processes. *Psychol. Sci.* 9, 20–26. doi: 10.1111/1467-9280.00004
- Smith, T. A., and Kimball, D. R. (2012). Revisiting the rise and fall of false recall: presentation rate effects depend on retention interval. *Memory* 20, 535–553. doi: 10.1080/09658211.2012.684881
- Stadler, M. A., Roediger, H. L., and McDermott, K. B. (1999). Norms for word lists that create false memories. *Mem. Cogn.* 27, 494–500. doi: 10.3758/BF03211543
- Surraín, S., and Luk, G. (2019). Describing bilinguals: a systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingual. Lang. Cogn.* 22, 401–415. doi: 10.1017/S1366728917000682
- Turk, K. W., Palumbo, R., Deason, R. G., Marin, A., Elshaar, A., Gosselin, E., et al. (2020). False memories: the other side of forgetting. *J. Int. Neuropsychol. Soc.* 26, 545–556. doi: 10.1017/S1355617720000016
- Ulatowska, J., and Olszewska, J. (2013). Creating associative memory distortions—a Polish adaptation of the DRM paradigm. *Polish Psychol. Bull.* 44, 449–456. doi: 10.2478/ppb-2013-0048
- Van Damme, L., and d’Ydewalle, G. (2009). Implicit false memory in the DRM paradigm: effects of amnesia, encoding instructions, and encoding duration. *Neuropsychology* 23, 635–648. doi: 10.1037/a0016017
- Van Hell, J. G., and De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. *Bilingual. Lang. Cogn.* 1, 193–211. doi: 10.1017/S1366728998000352
- Watson, J. M., McDermott, K. B., and Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: assessing the combined influence of practice and warnings in young and old adults. *Mem. Cogn.* 32, 135–141. doi: 10.3758/BF03195826
- Wojcik, D. Z., Díez, E., Alonso, M. A., Martín-Cilleros, M. V., Guisuraga-Fernández, Z., Fernández, M., et al. (2018). Diminished false memory in adults with autism spectrum disorder: evidence of identify-to-reject mechanism impairment. *Res. Aut. Spect. Disord.* 45, 51–57. doi: 10.1016/j.rasd.2017.11.001

Conflict of Interest: 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 © 2021 Suarez and Beato. 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) and the copyright owner(s) 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.



The Dynamic and Fragile Nature of Eyewitness Memory Formation: Considering Stress and Attention

Alia N. Wulff* and Ayanna K. Thomas

Cognitive Aging and Memory Laboratory, Department of Psychology, Tufts University, Medford, MA, United States

OPEN ACCESS

Edited by:

Sara Cadavid,
Rosario University, Colombia

Reviewed by:

Valerio Santangelo,
University of Perugia, Italy
Alana Christiane Krix,
Maastricht University, Netherlands

*Correspondence:

Alia N. Wulff
alia.wulff@tufts.edu

Specialty section:

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

Received: 10 February 2021

Accepted: 19 March 2021

Published: 13 April 2021

Citation:

Wulff AN and Thomas AK (2021) The
Dynamic and Fragile Nature of
Eyewitness Memory Formation:
Considering Stress and Attention.
Front. Psychol. 12:666724.
doi: 10.3389/fpsyg.2021.666724

Eyewitnesses are often susceptible to recollection failures and memory distortions. These failures and distortions are influenced by several factors. The present review will discuss two such important factors, attention failures and stress. We argue that acute stress, often experienced by eyewitnesses and victims of crimes, directly influences attentional processes, which likely has downstream consequences for memory. Attentional failures may result in individuals missing something unusual or important in a complex visual field. Amongst eyewitnesses, this can lead to individuals missing details, even unusual or important central details, regarding the crime. Surprisingly, few studies have investigated attentional failures in eyewitness scenarios, and none have investigated the relationship between stress, attention, and witness memory. This review will discuss the impact of attentional failures, mainly those resulting from inattention blindness, in applied contexts in order to bridge to eyewitness scenarios. In addition, we will integrate the applied literature on attentional failures with literature that examines the influences of arousal and stress on attention. We will conclude by presenting how future research may tease apart the independent contributions of arousal and stress on attentional failures and successes and how this research may inform understanding of eyewitness reliability.

Keywords: inattention blindness, stress, eyewitness memory, emotional arousal, memory distortions

Eyewitnesses, including those who may also be victims of crimes, are expected to remember relevant and accurate information regarding their witnessed crimes. They are questioned repeatedly and are required to remember small, potentially peripheral, details of crimes, such as the identities of other potential eyewitnesses, and more central details, such as the type of weapons present. These details are often crucial pieces of information used in investigations and in the context of legal proceedings.

It is unlikely that eyewitnesses would be able to accurately recall all details of an event. Further, repeated questioning may leave individuals susceptible to memory errors (Thomas et al., 2010; Chan and LaPaglia, 2011) and inflates confidence in repeated details (Shaw et al., 1996). Researchers have focused a great deal of their work on the impact of retrieval processes, engendered by repeated questioning, on eyewitness memory. Much of this work has demonstrated that retrieval of accurate information when questioning is interleaved with the presentation of new, sometimes incorrect, information. This phenomenon, known as Retrieval Enhanced Suggestibility (RES; cf. Thomas et al., 2010) has its roots in the well-established misinformation effect (for review see, Loftus, 2005). Researchers have found that eyewitnesses are extremely susceptible to questioner demand and are likely to report inaccurate information if asked leading questions (Loftus, 1975; Weinberg et al., 1983; Murphy and Greene, 2016). In addition to demand, researchers have demonstrated that

post-event information may disrupt access to original event details (cf. Belli, 1989), or may increase source misattributions (cf. Belli et al., 1994).

Although a large portion of researchers have focused on retrieval of witnessed or experienced events, we present a complementary, but equally important, question to consider: how do attention processes influence the encoding and later retrieval of witnessed or experienced events? Memory distortions and errors are not often investigated as a result of attention and encoding failures. In the laboratory, researchers construct memory experiments such that participants are able to attend to and successfully encode a baseline percentage of information. In a real-life scenario, it cannot be assumed that eyewitnesses encoded 60, 25, or even 1% of the information regarding a crime. We suggest that eyewitness memory and false memory researchers should consider factors that may impact attention, in order to understand the reliability of eyewitness and victim memory.

The Case for Investigating Attentional Failures and Stress in Eyewitnesses

Humans are rarely able to attend to all visual stimuli in their visual field. Furthermore, they are inundated with distractions, such as their cell phones or their conversational partners. As a result of these realities, they are often susceptible to failures of attention. One such failure of attention is inattention blindness (Mack and Rock, 1998). This is a failure to notice unusual or unexpected events in a complex visual field. There are many instances in which inattention blindness can negatively impact individuals' lives. It could be as trivial as a jogger looking at their phone to respond to a text and tripping on a curb to something as serious as a driver looking at a billboard and crashing into a telephone pole. In both of these cases, the individual may have been focusing their attention on one task and failed to notice important information in their environment.

This is directly relevant to witnesses of crimes. It is rare that individuals are ever vigilant for a crime while going about their daily lives. As such, individuals who are questioned as "eyewitnesses" may have been present at a scene but may have not noticed a crime occurring. For example, an individual may be sitting at a bus stop listening to music when two cars crash into each other less than a block away. Since they were looking in the general direction of the crash, a police officer might expect them to be able to provide an eyewitness account of the incident. However, the individual, focusing on their music, may not have been attending to the cars before the crash and only became aware of the incident when sirens alerted them. This would be an example of an eyewitness being expected to provide an account based on their memory for the event, even though they had experienced inattention blindness. While there is limited research into inattention blindness for crimes specifically, there is a plethora of research that uses paradigms which could, ostensibly, be applied to eyewitness scenarios.

Additionally, we argue that to understand and predict eyewitness memory accuracy, researchers should not only begin to directly relate attentional failures to later memory distortion, but must also begin to examine components of arousal and stress

as these physiological responses to external stimuli may have direct consequences on cognitive processes associated attention and memory formation. The acute stress response can be empirically measured in a lab setting and consists of a biphasic, two-pathway response. The first phase, an experience sometimes colloquially referred to as the "fight or flight" response, provides a sudden burst of energy while the second phase helps repair the body after the stressful experience. Marr et al. (2020) identified a fundamental difference in how different types of memory experts tend to view the impact of stress on memory. Eyewitness memory experts have suggested that stress at encoding impairs eyewitness accuracy, while basic memory experts generally argue that stress at encoding may enhance memory. This discrepancy between groups of researchers has limited the investigation of stress on eyewitness memory. For example, stress researchers often use well-validated stress induction techniques to isolate components of immediate and delayed stress responding and measure physiological changes that result from stress induction. In the past, eyewitness memory researchers have not often used these induction techniques and may have instead inferred a stress response using self-report measures. However more recently researchers have begun to use validated stress induction techniques in eyewitness studies (Krix et al., 2016; Sauerland et al., 2016). In these studies, stress was found to not impair eyewitness encoding. These researchers have even pointed out the discrepancies between the two types of memory research, such as different retention intervals and stress induction techniques, and pointed toward a need for improved methodological rigor within eyewitness research (Sauerland et al., 2016).

Furthermore, to our knowledge, while there are some studies that deal with inattention blindness under potentially emotionally arousing conditions, there have been no studies which have directly investigated the influence of an acute stress response inattention blindness. Without this critical information, not only are we failing to come to a consensus regarding the impact of stress on memory encoding, but also failing to even research the impact stress may have on parallel cognitive processes such as inattention blindness.

Therefore, the current review has two aims. The first is to broadly investigate the basic effect of inattention blindness in both lab and applied settings. The second aim is to detail the limited research regarding the impact of emotionally arousing stimuli and physiological stress on attention and the impact that stress may have on inattention blindness. We will conclude by discussing the implications of these emotional and physiological factors on eyewitness memory reliability.

WHAT IS INATTENTIONAL BLINDNESS?

The phenomenon now known as inattention blindness was first demonstrated by Neisser and Becklen (1975), who showed participants three transparent overlapping videos. Two videos depicted people passing basketballs between each other and one depicted a woman with an umbrella walking across the screen. When participants were told to count the number of times one of the teams passed the basketball, 79% of participants

missed the umbrella woman walking through the screen. Inattention blindness, as defined by Mack and Rock (1998), is the phenomenon whereby people tend to miss events that occur in their visual field, no matter how unusual or unexpected those events may be, if their attention is elsewhere. This earlier work by Neisser et al. foreshadows the now famous gorilla experiment (c.f. Simons and Chabris, 1999). In this study, participants counted basketball passes between a group of individuals while a person dressed in a gorilla costume or a woman holding an umbrella walked through the game in clear view of the camera. Participants missed the unusual events 46% of the time.

In a parallel line of research Mack and Rock (1998) had participants look at a small cross on a computer screen and report if the horizontal or vertical arm of the cross was longer. On one of the trials, a small black square appeared in a quadrant of the cross. In this study, an average of 25% of participants failed to notice the black square. However, when participants were simply told to look at the screen without additional attention instructions, all participants noticed the black square. These experiments demonstrated that when people are attending to a task, they can miss other things in their visual field. They also align with Neisser and colleagues' earlier work with more complex and ecologically relevant materials.

While researchers agree that people are susceptible to inattention blindness, there are conflicting explanations for a potential mechanism for the phenomenon. One possible explanation for inattention blindness could be that individuals who miss the unexpected stimulus simply did not look at the unexpected stimulus. However, research using eye tracking has shown that this is not the case. There have been several studies that show individuals who were placed in inattention blindness conditions and individuals who were not given inattention blindness instructions were equally likely to have eye movements near and even fixate on the stimulus (Koivisto et al., 2004; Beanland and Pammer, 2010).

Another explanation, called inattention amnesia, was proposed by Wolfe (1999). This explanation purported that the unexpected information is seen and immediately forgotten, rather than not seen at all. Supporting this inattention amnesia mechanism, Butler and Klein (2009) presented a series of overlapping pictures and words and told participants to report when one of the streams of information (words or pictures) presented the same stimulus back-to-back. They used a recognition task in which participants had lower rates of recognition for the unattended stream of information to show that participants were, in fact, not paying attention to the unattended stream. However, they also found that participants were successfully primed to report information from the unattended stream on a perceptual identification task (completing masked words). The researchers claim that while it is unlikely that participants had explicit awareness of the unattended stream of information, there is evidence that participants were able to perceive the information on some level, and perhaps simply forgot the information at the time of the recognition memory test.

Another study, completed by Ruz et al. (2005), displayed overlapping pictures and letters (which could either spell a

word, such as "CLOUD" or a non-word, such as "DLSPR") to participants. Participants were instructed to attend to either the pictures or the letters and report if stimuli repeated. The researchers used ERPs to find that words were processed differently than non-words by the brain, even when participants were attending to the picture stream of information, not the letter stream. The researchers claimed this may indicate our brains do perceive information, even when that information is not something to which we are attending. Importantly, and contrary to the claims of proponents of inattention amnesia, this may not be evidence that individuals are *forgetting* information. This is simply because it cannot show that they encoded the information in the first place.

Furthermore, there is ample evidence that runs contrary to an inattention amnesia mechanism. One such study, completed by Ward and Scholl (2015), asked participants to report on unusual stimuli mid-trial, before there was any chance for perceptual decay. The researchers found that 13% of participants were still unable to report on an unexpected stimulus even when asked immediately following the stimulus on the fourth trial in which something unexpected occurred and had been told to be vigilant for unexpected stimuli. While this rate of IB is fairly low compared to previous studies, this is still significant compared to the control condition in which the stimulus was expected.

These findings align with early work done by Becklen and Cervone (1983), who presented participants with the stimuli created by Neisser and Becklen (1975). However, unlike Neisser and Becklen, they used different versions of the video that ended at various points- at the end, when the woman with the umbrella was in full view, or directly following the woman's exit. They also told another group that the video would end suddenly, at which time the participant would have to immediately report exactly what was presented. The video for this group ended when the woman with the umbrella was on screen. If a tendency to quickly forget encoded information underlies inattention blindness, not only should the participants for whom the video ended when the woman with the umbrella was on screen have noticed her more than the other participant groups, the participants who were warned ahead of time to describe exactly what was onscreen when the video ended should be able to describe the woman. However, this was not the case. In fact, researchers found that participants who watched the full video or the video that ended once the woman left noticed her 34% of the time, while participants for whom the video ended while she was onscreen noticed her only 7% of the time, regardless of instructions. This study showed no evidence that information was encoded and then forgotten between the video ending and reporting.

An alternative explanation for inattention blindness was proposed by Mack and Rock (1998). They proposed that inattention blindness may not necessarily be a failure of seeing, but rather a lack of explicit awareness of the environment. In other words, individuals who experience inattention blindness may have some sensory awareness of the information in their environment but that sensation may not reach the threshold of awareness necessary to gain full perceptual attention. Essentially, inattention blindness may be accompanied by unconscious perception (Mack, 2003). This could be what Butler and Klein

(2009) and Ruz et al. (2005) found in their studies. There was no evidence that participants were aware of the information and then immediately forgot it. Instead, participants continued to report a lack of awareness even while their brains registered the stimuli.

To this point, research has not been able to pinpoint the specific mechanism that underlies inattention blindness; however, the proposed mechanisms both predict that people will fail to report unusual events, even when they occur in plain view. This prediction has dramatic applied consequences. As such, a large literature has been devoted to understanding the inattention blindness phenomenon in more naturalistic and applied settings.

Inattention Blindness in Realistic Scenarios

Researchers have examined inattention blindness in both the lab and in naturalistic settings in order to determine the impact that the phenomenon can have on real-world experiences. For example, Pammer et al. (2015) investigated the potentially dangerous impact that inattention blindness may have on drivers and bystanders. Participants looked at images of roads (taken from the inside of a vehicle) for 1.5 seconds and were asked to judge whether the image depicted a safe or unsafe environment in which to drive. On a critical trial that contained an unexpected object, participants were also asked if they had seen anything other than the cars, trees, and streetlight. Researchers found that 10% of participants did not report seeing a child running toward the road (the unexpected object). Additionally, over half of participants did not see either an adult or a child standing close by the road. This is an especially concerning depiction of inattention blindness as the task that participants focused on which made them blind to the road hazards was, quite ironically, looking for road hazards.

Some researchers have also used a lab setting to investigate situations in which eyewitnesses experience inattention blindness. One such study had participants watch a video of a busy shopping center and either count the number of people wearing a blue shirt, count the number of shopping bags, or just watch the video (Rivardo et al., 2011). Researchers found that 81% of participants who were counting shirts did not notice the theft of a shopping bag, while 62% of people who were counting shopping bags did not notice the theft. Participants who were told to simply watch the video failed to notice the theft only 10% of the time. Importantly, participants who had their attention specifically directed toward stimuli that were directly related to the crime were more likely to notice the crime; however, any task engagement consistently led to higher rates of inattention blindness than does no task engagement.

Other researchers have looked at rates of inattention blindness amongst people who are considered “experts” in viewing certain kinds of information. One such study investigated inattention blindness in radiologists (Drew et al., 2013). Researchers showed a series of lung scans to radiologists and non-radiologists and asked them to identify lung nodules (a common task for radiologists). A gorilla, 48 times the size of the

average lung nodule, was on the final scan. Researchers found that 83% of radiologists missed the hidden gorilla while looking for lung nodules, while 100% of non-radiologists missed the hidden gorilla while looking for lung nodules.

Other “experts” have also been found to experience inattention blindness in their job. For example, Näsholm et al. (2014) found that military personnel tasked with monitoring CCTV footage were susceptible to inattention blindness for critical information at an alarmingly high rate. Comparing novices and active-duty military personnel, the researchers found that 50% of novices missed a woman placing a suspicious package on the ground and looking into the camera before leaving frame (a task relevant stimulus) and 81% missed a woman in a pirate costume walking into frame and looking at the camera before leaving (a task irrelevant stimulus). Surprisingly, they also found that 61% of the military participants also missed the package stimulus, even though those actions could have severe consequences on a military base, while 76% of military participants missed the pirate. There was no difference in inattention blindness rates for the participants, regardless of expertise.

Using a more naturalistic approach, Hyman et al. (2009) studied rates of inattention blindness in college students walking on their campus. They had a person dressed in a clown costume unicycle in a circle near a well-traveled walking path through a campus square and surveyed individuals who walked past the clown. They found that 75% of individuals who were on their cell phones missed the unicycling clown. However, amongst individuals who were not distracted by a cell phone, only 49% of people walking alone missed the unicycling clown, while only 29% of people walking in pairs missed the clown. Common distractions, such as cell phones, led to high rates of inattention blindness, even for something as absurd, unexpected, and novel as a unicycling clown on a college campus. Interestingly, the individuals walking in pairs noticed the clown more often than the individuals walking alone, a surprising result for those who may think conversation could be a distraction. However, the rate of noticing can be explained by the fact that if one conversational partner noticed the clown, they likely told the other person. Put simply, more observers mean more opportunities for something to be noticed.

In a follow-up study done, Hyman and Wise-Swanson (2014) found that only 6% of 63 individuals who were talking or texting on a cell phone noticed money hanging from a tree branch in the middle of a walking path. Only 19% of the 333 participants who were not engaged in a cell phone saw the money. These data suggest that using a cell phone increases inattention blindness but that inattention blindness may still occur even when a cell phone is not in use. The authors suggested that the individuals who were not engaged with their cell phones may have been engaging in some form of mind-wandering. Focusing on their own thoughts may have been an engaging enough task to induce inattention blindness.

Perhaps more in line with factors that may influence our understanding of eyewitness attentional processes, a naturalistic study completed by Simons and Schlosser (2017) utilized the presence of a gun on the dashboard of a car used in a simulated

traffic stop. The simulation was completed by police trainees and experienced police officers as part of a police training exercise. Afterwards, participants were asked if they had noticed the gun plainly displayed in their field of view during the entire traffic stop. Researchers found that the experienced police officers noticed the gun more often than did trainees (67 vs. 42%). However, even amongst experienced law enforcement professionals, who are heavily trained to notice and react to potentially life-threatening objects, a full third of participants missed the unexpected stimulus.

Another study that involved eyewitness attentional processes had participants run behind a researcher while counting the number of times the researcher touched their head (Chabris et al., 2011). At a certain place beside the path were three men in a physical altercation. In this study, only 35% of participants noticed the fight when it took place at night. When it took place in the daylight, 56% of participants noticed the altercation. Noticing rates were also impacted by attentional load. When researchers took away all counting tasks, 72% of participants noticed the fight in the daylight. However, when researchers gave participants two counting tasks, only 42% of participants noticed the fight in the daylight. A full quarter of participants missed something as unusual, violent, and unexpected as a loud physical altercation in broad daylight when simply jogging at a reasonable pace behind another person. This study suggests an important role of attentional load in inattention blindness, namely that increased attentional load can lead to higher rates of inattention blindness, especially in a degraded visual field.

Inattention Blindness in Emotionally Arousing Scenarios

Inattention blindness can be experienced by everyone, even when the objects that individuals miss are glaringly obvious to an outside observer or directly relevant to tasks the individuals are trying to complete. It can even be experienced by individuals that may be put in danger by the unexpected and unnoticed object. However, there is a significant factor in many such situations that researchers have yet to investigate. Many of the scenarios that have been discussed in this paper thus far, such as traffic stops (Simons and Schlosser, 2017), CCTV surveillance (Näsholm et al., 2014), or witnessing a crime (Rivardo et al., 2011) are scenarios in which individuals are likely to experience emotional arousal and potentially even a physiological response.

Police officers are taught to be aware of life-threatening danger whenever they are on the job, including during routine traffic stops. Military CCTV operators are charged with ensuring the safety of the base and their fellow soldiers within the base. And, most relevant to the current review, eyewitnesses are often exposed to potentially violent, traumatizing, and/or stressful scenarios of many kinds. Being able to look at the direct effect of stress on inattention blindness is therefore important in understanding both how inattention blindness manifests in the context of an acute stress response, and how this interaction may influence the reliability of eyewitness memory.

Unfortunately, to our knowledge, there are no direct experimental manipulations of stress in inattention blindness

studies. However, there are studies that use negative stimuli to investigate how inattention blindness may be impacted when the unexpected event itself is something potentially emotionally arousing (see **Table 1** for a summary of these studies). These stimuli are generally items such as spiders, snakes, or guns. One caveat to these studies is that they do not induce stress prior to exposure to the unexpected stimulus. Rather, it is the unexpected stimuli themselves that are intended to induce a stress response. Although this may better align with a real-world experience of an eyewitness (i.e., it is the crime itself that would likely be the threatening stimulus, not a *prior* scenario), this methodology does not allow for direct examination of acute stress on inattention blindness. These studies are potentially good indicators of the impact that emotional arousal may have on eyewitness attention but may not provide answers as to how stress impacts eyewitness attention.

Regardless, these studies do provide some initial information regarding how inattention blindness is impacted by threatening stimuli. The first of such studies discussed here was completed by Beanland et al. (2018). The researchers had participants fixate on a cross in the middle of a blank screen, which then flicked briefly to a screen containing four pictures of animals or furniture. The task was to identify the pictures. However, on two of the nine trials the fixation cross was replaced by a threat word ("KILLER") or a non-threat word ("MERGER" or "MILLER"). Only 22% of participants were able to report one of the pictures, while only 8% were able to report both. Of main interest, however, 19% of participants were able to report the threatening word, while only 11% of participants were able to report the neutral word. This result shows that participants were more likely to report a threatening word than a non-threatening one, even when their attention was on a different task.

There are several more studies that utilize pictures, rather than words, to capture an effect of threat on inattention blindness. One such study found that 81% of participants who were under conditions that encourage inattention blindness (a line-length judgment paradigm similar to Mack and Rock, 1998) could detect a line drawing of a spider, compared to only 53% who could do the same for a line drawing of a hypodermic syringe (New and German, 2015). When participants were not under conditions of inattention blindness, 100% of them were able to detect both the spider and needle. Inattention blindness was present no matter the stimulus, but may have been reduced by the presence of such a classically negative stimulus as a spider, compared to a relatively newer and less commonly negative stimulus as a needle.

Other researchers have used modern threat objects that are more dangerous than a hypodermic needle. Gao and Jia (2017) found that participants under conditions of inattention blindness with a low perceptual load were more likely to notice a threat object (e.g. a gun; 60%) than a non-threat object (e.g., a flower; 35%). This aligns with the previous studies, in that participants are more likely to notice threatening objects than non-threatening objects. However, amongst participants who were under a high perceptual load, there was no statistical difference between identifying a threatening (35%) and nonthreatening (19%) object. The low-load task was to report the

TABLE 1 | Summary of studies using threatening stimuli in inattentional blindness paradigms.

Authors	Exp.	N	Experimental paradigm	Unexpected (IB) stimulus	Noticing rate		
					Control stimulus	Unexpected stimulus	Different from control?
Simons and Schlosser (2017)		175	Stimulated traffic stop	Gun on dashboard of car	N/A	52.6%	N/A
Näsholm et al. (2014)		171	Watch video and verbally describe events	Person leaving suspicious package or person dressed in pirate costume	<i>Pirate</i> : 21%	<i>Package</i> : 55%	Yes
Rivardo et al. (2011)		187	Count shirts or bags in a video of a theft of a bag	Individual stealing a bag	<i>Counting shirts</i> : 19%	<i>Counting bags</i> : 38%	Yes
Beanland et al. (2018)		111	Report specific items from a visual field	Threat word or not threat word	<i>Non-threat</i> : 11%	<i>Threat</i> : 19%	Yes
New and German (2015)	1	252	Line judgement task	Illustrations of spider and needle	<i>Needle</i> : 53%	<i>Spider</i> : 81%	Yes
	2	320	Line judgement task	Illustrations of spider, needle, and fly	<i>Needle</i> : 53%	<i>Spider</i> : 80%	Yes
Gao and Jia (2017)					<i>Fly</i> : 73%		No
		192	Counting number of color words	Illustrations of threat and non-threat objects	<i>Low load</i> : 35.4%	<i>Low load</i> : 60.4%	Yes
Wiemer et al. (2013)					<i>High load</i> : 19%	<i>High load</i> : 35%	No
		120	Line judgement task	Flower picture and spider picture	<i>Flower</i> : 58%	<i>Spider</i> : 52%	No
Calvillo and Hawkins (2016)	1	168	Searching for a word	Line drawings of threat and non-threat objects	<i>Non-threat</i> : 39%	<i>Threat</i> : 50%	No
	2	238	Line judgement task	Pictures of threat and non-threat objects	<i>Non-threat</i> : 53%	<i>Threat</i> : 32%	Yes, threat was noticed less
Stothart et al. (2017)	1	576	Played video game avoiding costly missiles	Square the same color as most costly missiles	<i>Unrelated color</i> : 62%	<i>Most costly</i> : 30%	Yes, cost was noticed less
	2	595	Played video game avoiding missiles	Square the same color as most costly missiles	<i>Unrelated color</i> : 70%	<i>Most costly</i> : 53%	Yes, cost was noticed less
	3	599	Played video game avoiding missiles and hitting targets	Square the same color as missiles or target	<i>Unrelated color</i> : 78%	<i>Missiles</i> : 44% <i>Target</i> : 55%	Yes, enemies and friends noticed less
Redlich et al. (2019)	1	277	Line judgement task	Colored square associated with high reward	<i>No reward associated</i> : 70.79%	62.11%	No
	2	260	Counting shape bounces	Colored shape associated with high reward	<i>No reward associated</i> : 31.03%	29.41%	No

Note. Not all stimuli from every experiment is included in this table, only those relevant to the present review.

color words (e.g., blue) from three possible words, while the high-load task was to report the color words from six possible words. Both groups of participants had one second to complete this task.

This is interesting, as it suggests that an even incremental increase in task difficulty could eliminate the effect of threatening objects on inattentional blindness. However, other studies that did not induce perceptual load found a different pattern of results. For example, Wiemer et al. (2013) found that there was no difference in noticing rates of an unexpected picture of a spider (52%) compared to an unexpected picture of a flower (58%). Importantly, these findings contrast those reported by New and German (2015), who used a similar procedure. Furthermore, on a later test of memory, participants were as likely to remember the spider as they were the flower (Wiemer et al., 2013). A reasonable explanation for the difference could be that the flower used by Wiemer et al. (2013) was simply more noticeable than the syringe

or fly used by New and German (2015), but it is unclear if this could explain the discrepancies. However, the researchers also found that pictures of spiders resulted in higher skin conductance responses and more saccadic eye movements toward them than did pictures of flowers, even amongst individuals who did not report noticing either picture. This suggests that participants may have processed the stimuli as a threat, even though the threat did not increase rates of noticing the unexpected stimulus.

Further support for the conclusion that threatening stimuli may not impact inattentional blindness comes from a study done by Calvillo and Hawkins (2016), who also used an identification task to assess rates of inattentional blindness. Participants were shown a set of four words around a screen and had one second to find the sport word (e.g., softball). On one trial, an unexpected object appeared in the middle of the screen. The researchers found that there were no differences in noticing rates

for threatening (50%) and non-threatening objects (39%). In fact, due to low identification rates of two of their stimuli in particular (a sword and a snake), threatening objects were actually identified less frequently (32%) than were non-threatening objects (53%). In addition, they found that participants were more likely to correctly identify still pictures of animate objects (e.g. a spider or bird; 54%), regardless of potential threat associated with the stimulus, compared to inanimate objects (e.g., a gun or bed; 36%). The authors concluded that it is not the threat that captures attention, but rather if the objects are animate.

In a parallel line of research, studies have shown that penalties and rewards also have little impact on inattention blindness. In one study, researchers created a computer game in which participants had to avoid enemies, as collisions with enemies would decrease their score, and hit friends, as collisions with friends would increase their score. During this game, an object that matched either the enemy color or the friend color traversed the screen. The researchers found no difference in noticing rates between objects whose colors matched the enemies, as opposed to friends. Participants did not notice unexpected objects, even when those objects were associated with a cost in their task (Stothart et al., 2017).

Similarly, when participants were given a task in which certain colors were associated with actual monetary rewards, the reward did not impact inattention blindness (Redlich et al., 2019). In contrast, although military-trained CCTV operators missed seeing a woman setting down a suspicious package and then leaving in CCTV footage, researchers found that trained and novice operators were more likely to notice the woman setting down a package than they were to notice a woman in a pirate costume staying in frame for an equivalent period of time (Näsholm et al., 2014). This is important for two reasons. First, missing an individual who has a suspicious package on CCTV footage is arguably a more relevant “cost” than missing an object similarly colored to an enemy that makes a participant lose points in a computer game or even a color that is associated with money. Second, being aware of suspicious packages is ingrained in our culture (Morewitz, 2019) and is a well-known potential threat. Because of these reasons, the point that Stothart et al. (2017) made may still stand. In a lab, participants are much less likely to recognize an object as a threat, so extra care must be taken to ensure that participants are reacting to the stimuli in ways the researchers expect. Ecologically valid paradigms, such as those using videos of real people engaging in threatening actions, are potentially the only way we can truly assess how individuals react to threatening stimuli.

The research to date does not suggest a clear picture of how individuals' rates of inattention blindness would change with the introduction of a threatening stimulus. One thing is clear, however; even in cases in which the threat is clear and present and results in *lower* rates of inattention blindness, the rates are never reduced to zero. As this relates to eyewitness reliability, the consequences may be errors of omission, memory distortion, and confabulation. Although the highest rate of noticing reached 90% (c.f., New and German, 2015), it is important to note that noticing was defined as simply acknowledging the odd occurrence without including the specifics of the occurrence. Further, this high level

of noticing was found in the context of a controlled laboratory experiment and may not represent the conditions experienced by real-world eyewitnesses.

AROUSAL, ATTENTION, AND MEMORY

The above studies may be inducing physiological or psychological arousal due to the negative emotions induced by the stimuli. Emotional arousal at the time of encoding has well-studied impacts on later memory. This is relevant to the present review as the physiological and psychological impacts of emotionally arousing stimuli and an acute stressor are similar (Lang and Bradley, 2007; Campbell and Ehler, 2012). Witnessing or being a victim to a crime may elicit a negative emotional experience. In controlled experiments, researchers have attempted to induce negative emotional arousal to understand the impact it may have on later memory. The literature on emotions and memory is vast; therefore we will focus the present discussion on studies that have examined eyewitness memory specifically or have examined the relationship between negative emotional arousal and attention.

The highly influential Easterbrook hypothesis (1959) proposed that attention narrowing occurs in the context of high emotional arousal. Whereas individuals at moderate levels of arousal are able to attend to many cues in their environment, resulting in a higher level of performance on tasks, individuals with higher levels of arousal may experience attention narrowing, resulting in salience of a subset of cues and obscurity of other cues. For example, in the context of witnessing a crime, negative emotional arousal may result in the salience of a weapon and the hand that is holding the weapon, but indistinctness or ambiguity of non-focal elements (Kocab and Sporer, 2016).

Christianson and Loftus (1991) presented participants with a narrative witnessed event across a series of ordered pictures that contained a target picture wherein a woman either rode the bicycle (neutral event) or was lying on the ground, injured from a bicycle accident (emotional event). The researchers found that participants exposed to the negative picture within the series exhibited better memory for the central detail (color of the woman's coat) but poorer memory for the peripheral detail (color of a car driving in the background) compared to those who viewed the neutral picture.

Yeghyan and Yonelinas (2011) found similar results using individual pictures, rather than a narrative. The pictures were selected from the International Affective Picture System (IAPS; Lang et al., 2008) and depicted scenarios that were designed to elicit varying levels of emotional arousal. Central details were defined as those that would change the description of the event if removed, while peripheral details were defined as those that would not change the description of the event if removed. The authors found that participants who rated the emotional content of negative pictures in the upper half a scale from one (low arousal) to nine (high arousal) exhibited poorer recognition memory for peripheral details than central details in those pictures. The authors concluded that negative emotional arousal likely led to memory narrowing at the highest levels of arousal, a conclusion that aligns with the Easterbrook hypothesis (1959).

Another study compared the memory of police officers for a domestic dispute incident that contained a handgun (the high arousal situation) and a domestic dispute incident that did not contain a handgun (the low arousal situation; Hulse and Memon, 2006). The researchers found that participants who were exposed to the high arousal situation recalled fewer details but were more accurate overall than those exposed to the low arousal situation. While this study did not differentiate between central and peripheral details, this provides further evidence that participants exposed to high levels of emotional arousal are only able to attend to a subset of information while maintaining high levels of performance.

Further, Christianson (1992) suggested that the relationship between negative emotional arousal and memory may depend on whether the emotional stimuli are related to the primary memory task or witnessed event (e.g., if the stimulus that induced emotional arousal was the crime about which the witness is then questioned). Hanoch and Vitouch (2004) considered this idea of arousal-congruent performance and concluded that, in order to see the true impact of emotional arousal on performance, researchers should induce emotional arousal using information that is relevant to the task for which researchers measure performance.

Controlling for attentional capture and fixation, Christianson et al. (1991) found that the emotional arousal-memory effect was unlikely the result of overt attentional processes. As in the earlier study, they exposed participants to a series of sequential pictures that depicted a narrative. The series contained negative and neutral pictures. When participants were restricted to a single fixation on the central object in the critical picture, participants shown the negative pictures had better recall performance for the central detail than did participants who were shown the neutral pictures. As each participant fixated on the same detail for the same length of time, this result cannot be explained by differing levels of overt attention. This same pattern occurred when participants were allowed multiple fixations. Participants who were shown negative pictures had better recall performance for central details and shorter fixation duration than did participants who were shown the neutral pictures.

Kim et al. (2013) recorded the eye movements of participants while exposing them to either negative or neutral picture stories. The authors found that participants had poorer recognition memory for peripheral details in a negative picture story than in a neutral picture story. They also found that participants in general were able to recognize central details after only a short fixation, while peripheral details were better recognized when they were fixated for a longer duration. The impact of fixation length on central details was even less relevant to memory when the pictures were negative, as central details in negative pictures were fixated for significantly less time than were central details in neutral pictures, even though the memory for those details was equal in negative and neutral conditions. However, there was no difference in duration fixation between the negative and neutral conditions for peripheral details, even though memory for those details was poorer in the negative condition than the neutral condition.

Emotional arousal at the time of encoding has an impact on subsequent retrieval. When emotional arousal is induced using negative stimuli, individuals exhibit increased memory performance for details that were central to the event and/or decreased memory performance for details that were peripheral to the event.

THE BIPHASIC ACUTE STRESS RESPONSE

Emotional arousal is only one aspect of the eyewitness experience. Another important factor that none of the previously discussed studies manipulated is the physiological response to an acute stressor. An acute stress response has the potential to occur in both eyewitnesses to and victims of crimes. This response occurs in two phases. The first phase results from the activation of the sympathetic-adreno-medullar (SAM) axis (Godoy et al., 2018). During this phase the body shuts down all unnecessary bodily functions. Adrenaline, a fast-acting and quick-burning source of energy, is released in high volume and heart rate increases.

About 20 min after the stressor has occurred, the body begins to enter the phase two stress response (Kirschbaum et al., 1993). The phase two stress response is directed by the hypothalamic-pituitary-adrenal (HPA) axis (Everly and Lating, 2013). In this phase, the body begins to restart bodily functions and heart rate begins to slow. The production of adrenaline is decreased and replaced with cortisol, a longer-lasting source of energy (Kirschbaum et al., 1993). It is possible to measure participants' response to an acute stressor during this phase by comparing their cortisol levels during a phase two stress response with their cortisol levels prior to stress induction. The peak of this stress response occurs 20 min after the introduction of the stressor (indicated by a peak in cortisol levels around this time; Kirschbaum et al., 1993). The physiological markers of a phase two stress response are generally gone within 24 h of experiencing the stressor.

There are many physiological and psychological similarities between emotional arousal and the acute stress response. Exposure to negative emotionally arousing pictures or videos has been found to increase stress hormones such as adrenaline and noradrenaline (related to the phase one stress response) and cortisol (related to the phase two stress response; Gerra et al., 1996; Codispoti et al., 2003). A well-validated set of emotionally arousing photos, the International Affective Picture System (IAPS; Lang et al., 2008) has been found to reliably induce changes in both overall heart rate and heart rate variability, as well as increase skin conductance (Appelhans and Luecken, 2006; Lang and Bradley, 2007), factors also associated with a stress response (Mackersie and Calderon-Moultrie, 2016; Kim et al., 2018). Therefore, while these responses are discussed separately in the present review, they may have similar impacts on attention and memory and should both be considered when investigating eyewitness reliability.

Although the acute stress response likely accompanies the eyewitness experience, to my knowledge only one study exists

that has manipulated the effect of arousal (associated with phase one stress responding) on inattention blindness. This study had individuals view the gorilla video created by Simons and Chabris (1999) while under varying conditions of physical activity (Hüttermann and Memmert, 2012). They found that participants who did not engage in physical activity noticed the gorilla 20% of the time, while participants under a medium physical load noticed the gorilla 40% of the time. Importantly, high physical load reduced noticing of the gorilla to zero. This pattern of results was replicated in the same paper with a different set of stimuli. However, it is possible that the physical activity itself may have served to divide attention, which had impacts on the cognitive processes independent of those associated with physiological changes.

We propose that researchers should invest efforts in additional studies that examine the acute stress response on attention and inattention blindness in eyewitness-like scenarios. As opposed to using stimuli that may increase arousal, we argue that new research employing well-established stress induction paradigms should be combined with basic and applied inattention blindness paradigms. There are many different validated protocols that have been created to induce stress in the lab, such as the Trier Social Stress Test (TSST; Birkett, 2011) or the Socially Evaluated Cold Pressor Test (SECPT; Schwabe and Schächinger, 2018). By incorporating these protocols, this area of research would foster the development of a more comprehensive and generalizable model of eyewitness reliability.

Although little is known about stress and inattention blindness, there is a small body of work examining stress and attention and a larger body of research focused on stress and memory encoding processes. A meta-analysis by Shields et al. (2016) found that individuals undergoing an acute stress response exhibit improved response inhibition (i.e., ability to withhold responses when necessary; mean effect size, $g^+ = 0.296$) but impaired cognitive inhibition (i.e., selective attention or ignoring; $g^+ = -0.208$). Furthermore, this effect is found to hold true regardless of delay between the stressor and the tasks and regardless of cortisol levels, which indicates that the impact of an acute stressor on task performance is the same whether participants were in phase one or phase two of an acute stress response.

One of the studies analyzed by Shields et al. (2016) that demonstrated decreased cognitive inhibition was completed by Sängler et al. (2014). In this study, researchers induced stress in half of participants and then gave them a task during the phase two stress response. This task required participants to report a luminance change in stimuli while ignoring more salient orientation changes in the same stimuli. They found that participants who had been exposed to the stressor made more errors (i.e., missed responding to the task entirely or responded to orientation rather than luminance) than did participants who were not stressed. Furthermore, the electrophysiological data from the same study showed that stressed participants paid less initial attention toward the luminance of the objects (i.e., the task-relevant stimulus) compared to non-stressed participants. This indicates that stressed participants had greater

difficulty inhibiting the task-irrelevant information than non-stressed participants.

However, other researchers have found results that do not support the conclusions. Booth and Sharma (2009) exposed participants to a stressor (a loud white noise) during a Stroop task. This task was completed during a phase one stress response. They found that participants who were stressed were better able to ignore irrelevant information than were those who were not stressed. This result aligns with findings presented by Chajut and Algom (2003), who found that participants who experienced a phase one stress response were better able to selectively attend to information than were participants who were not stressed.

In a related line of research, Qi et al. (2018) found that participants in a phase one stress response reported the direction an arrow was pointing faster than non-stressed participants. The authors hypothesized that stressed participants were able to focus only on the important perceptual details (i.e., the head of the arrow) rather than the entire image. All three of these studies took place during a phase one stress response, while the opposite result found by Sängler et al. (2014) occurred during a phase two stress response. This could indicate that a phase one stress response allowed the individuals to engage in more efficient perceptual processing. Rather than impairing cognitive inhibition, a phase one stress response could improve it. This conclusion, however, is muddled by the results of the meta-analysis conducted by Shields et al. (2016), which found stress impaired cognitive inhibition whether the studies tested participants during a phase one or phase two stress response.

Additionally, the impact of stress on individuals' ability to inhibit distractions is unclear. Selective attention could be improved during the phase one stress response (Chajut and Algom, 2003; Booth and Sharma, 2009; Qi et al., 2018). This could mean that participants who experienced a phase one stress response may be better able to focus on a single task and may be less likely to notice an unexpected stimulus. However, Shields et al. (2016) claim that, regardless of stress phase, participants who experienced a stress response may be more susceptible to distraction. Future research needs to specifically investigate the impact of the stress response phase on distractibility in order to make concrete claims in this area.

The impact of stress on attention is not the only factor important to consider in the case of eyewitnesses. Another equally important piece of information is how stress impacts memory. The impact of stress on memory is a well-researched field. In general, participants who experienced a stress response during an emotional event exhibit improved encoding of that event, compared to participants who did not experience a stress response (Cahill et al., 2003; Payne et al., 2007; Henckens et al., 2009). As crimes are often experienced as emotional and may lead to an acute stress response, eyewitness memory may be less susceptible to distortion than previously thought. Furthermore, a stress response at the time of encoding an emotional event has even been found to reduce the negative impact that misleading post-event information has on memory (Hoscheidt et al., 2014). This is another indication that moderate levels of stress may actually be a benefit for eyewitness memory, even downstream.

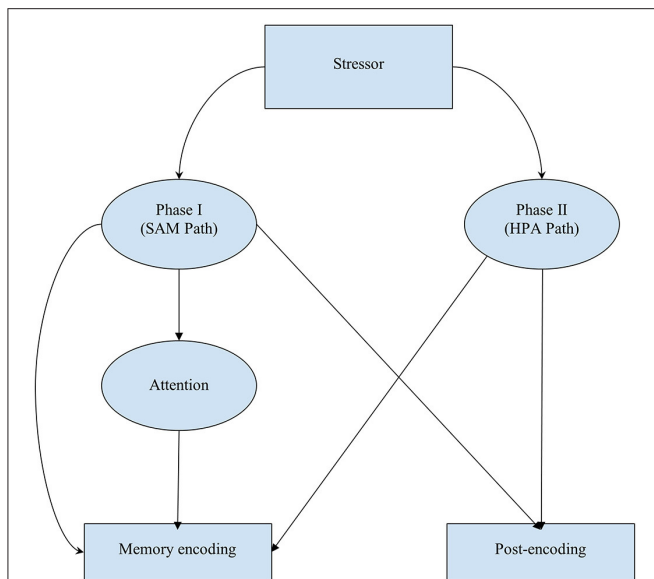


FIGURE 1 | Proposed theoretical model. The stressor results in the simultaneous activation of the SAM pathway and the HPA axis. SAM activation results in the immediate release of adrenaline and norepinephrine, impacting attention, encoding, and post-encoding processes, while the HPA axis is slower acting and results in the release of cortisol, impacting encoding and post-encoding processes. SAM activation may direct attention toward central event details, which subsequently impacts both memory encoding and post-event processes. The release of cortisol may also impact post-encoding processes.

We present a theoretical model based on the present state of the literature that has investigated the interactions between the two phases of the stress response, attention, and memory (**Figure 1**). A stressor occurs that induces a moderate level of stress, such as witnessing a theft. The stress response occurs in two phases. The phase one stress response, a product of the SAM pathway, pushes the body into a fight-or-flight response (Godoy et al., 2018). The hormones released lead to improved encoding of the event as well as improved post-encoding processes, such as consolidation and post-event retrieval (Gagnon and Wagner, 2016). In addition to encoding processes, the phase one stress response impacts attention. This most likely occurs as a reduction in cognitive inhibition, leading to poorer selective attention and less effective ignoring (Shields et al., 2016). Attention is focused on information that is central to the event, perhaps, in the case of a theft, the perpetrator themselves or the property being stolen. Information that is peripheral to the event, perhaps the identities of other witnesses, is attended to less (Christianson and Loftus, 1991; Yeghyan and Yonelinas, 2011; Kim et al., 2013).

The phase two stress response, occurring around 20 min after the initial stressor, is a product of the HPA pathway (Kirschbaum et al., 1993). At moderate levels, this response has been found to improve both encoding and post-encoding processes (Gagnon and Wagner, 2016). Therefore, a witness to a theft might have improved encoding, consolidation, and post-event retrieval due to the phase two stress response.

Overall, moderate stress would likely have a positive impact on memory for the event providing individuals are attending to details that are central to the event. If individuals are not attending to central details, it is probable that they would not exhibit the beneficial impact that moderate levels of stress can have on memory. It is important for future researchers to begin to incorporate paradigms that encourage individuals to attend to information not central to the target event, such as inattentional blindness paradigms, so we can have a better understanding of how attentional failures may interact with the present model.

Future research into inattentional blindness and stress should begin with a study that investigates the basic impact of a phase one stress response (such as that experienced by eyewitnesses) on inattentional blindness. Participant stress should be induced using a validated stress induction technique, such as the Trier Social Stress Test (TSST) or the Socially Evaluative Cold Pressor Test (SECPT). Immediately following stress induction participants should undergo a basic and well-tested inattentional blindness procedure, such as the line judgement task (Mack and Rock, 1998). The results of this experiment would provide some initial understanding of the impact of experimentally manipulated physiological stress on inattentional blindness. Once this has been determined, future experiments could begin to investigate the downstream consequences stress during attention may have on later retrieval.

CONCLUSION

Research into inattentional blindness has been ongoing for decades. There have been a multitude of studies completed in both laboratory and real-world settings that have all come to the same conclusion: humans are susceptible to missing information in our environments, even when that information is seemingly important or unique. This phenomenon may be especially relevant and unfortunate for eyewitnesses, who are already susceptible to memory failures, with this increased susceptibility a function of both attention and encoding failures. Therefore, we present an argument that research should examine the downstream consequences on eyewitness memory failures and distortions as impacted by inattentional blindness. Further, we argue that attention, encoding, and retrieval of witnessed and experienced events should also be investigated within the context of physiological and psychological reactions likely to occur when witnessing or being the survivor of a crime.

The failure to incorporate stress into inattentional blindness research has made it difficult to properly apply inattentional blindness research to eyewitness scenarios. Future research must incorporate experimentally manipulated and valid stress induction into the current inattentional blindness paradigms. Eyewitnesses most commonly experience a phase one stress response during the event, so a phase one stress response would be most applicable to real-world eyewitness scenarios.

Inattentional blindness can be an extremely dangerous attention failure. However, without further research into the direct effect of stress on inattentional blindness, researchers and

lawyers alike cannot effectively gauge the efficacy of a potential eyewitness. Researchers must work to combine our knowledge of the impact of stress on attention and inattention blindness to better serve our scientific understanding of both phenomenon and our ability to impact the legal field.

REFERENCES

- Appelhans, B. M., and Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Rev. Gen. Psychol.* 10, 229–240. doi: 10.1037/1089-2680.10.3.229
- Beanland, V., and Pammer, K. (2010). Looking without seeing or seeing without looking? Eye movements in sustained inattention blindness. *Vis. Res.* 50, 977–988. doi: 10.1016/j.visres.2010.02.024
- Beanland, V., Tan, C. H., and Christensen, B. K. (2018). The unexpected killer: effects of stimulus threat and negative affectivity on inattention blindness. *Cogn. Emot.* 32, 1374–1381. doi: 10.1080/02699931.2017.1394269
- Becklen, R., and Cervone, D. (1983). Selective looking and the noticing of unexpected events. *Mem. Cogn.* 11, 601–608. doi: 10.3758/BF03198284
- Belli, R. F. (1989). Influences of misleading postevent information: misinformation interference and acceptance. *J. Exp. Psychol. Gen.* 118, 72–85. doi: 10.1037/0096-3445.118.1.72
- Belli, R. F., Lindsay, D. S., Gales, M. S., and McCarthy, T. T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Mem. Cogn.* 22, 40–54. doi: 10.3758/BF03202760
- Birkett, M. A. (2011). The Trier Social Stress Test protocol for inducing psychological stress. *J. Visual. Exp.* 56:3238. doi: 10.3791/3238
- Booth, R., and Sharma, D. (2009). Stress reduces attention to irrelevant information: evidence from the Stroop task. *Motiv. Emot.* 33, 412–418. doi: 10.1007/s11031-009-9141-5
- Butler, B. C., and Klein, R. (2009). Inattention blindness for ignored words: comparison of explicit and implicit memory tasks. *Conscious. Cogn.* 18, 811–819. doi: 10.1016/j.concog.2009.02.009
- Cahill, L., Gorski, L., and Le, K. (2003). Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. *Learn. Mem.* 10, 270–274. doi: 10.1101/lm.62403
- Calvillo, D. P., and Hawkins, W. C. (2016). Animate objects are detected more frequently than inanimate objects in inattention blindness tasks independently of threat. *J. Gen. Psychol.* 143, 101–115. doi: 10.1080/00221309.2016.1163249
- Campbell, J., and Ehler, U. (2012). Acute psychosocial stress: does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology* 37, 1111–1134. doi: 10.1016/j.psyneuen.2011.12.010
- Chabris, C. F., Weinberger, A., Fontaine, M., and Simons, D. J. (2011). You do not talk about fight club if you do not notice fight club: Inattention blindness for a simulated real-world assault. *I-Perception* 2, 150–153. doi: 10.1068/i0436
- Chajut, E., and Algom, D. (2003). Selective attention improves under stress: implications for theories of social cognition. *J. Personal. Soc. Psychol.* 85, 231–248. doi: 10.1037/0022-3514.85.2.231
- Chan, J. C. K., and LaPaglia, J. A. (2011). The dark side of testing memory: repeated retrieval can enhance eyewitness suggestibility. *J. Exp. Psychol. Appl.* 17, 418–432. doi: 10.1037/a0025147
- Christianson, S. Å. (1992). Emotional stress and eyewitness memory: a critical review. *Psychol. Bull.* 112, 284–309. doi: 10.1037/0033-2909.112.2.284
- Christianson, S. Å., and Loftus, E. F. (1991). Remembering emotional events: the fate of detailed information. *Cogn. Emot.* 5, 81–108. doi: 10.1080/02699939108411027
- Christianson, S. Å., Loftus, E. F., Hoffman, H., and Loftus, G. R. (1991). Eye fixations and memory for emotional events. *J. Exp. Psychol.* 17, 693–701. doi: 10.1037/0278-7393.17.4.693
- Codispoti, M., Gerra, G., Montebanacci, O., Zaimovic, A., Raggi, M. A., and Baldaro, B. (2003). Emotional perception and neuroendocrine changes. *Psychophysiology* 40, 863–868. doi: 10.1111/1469-8986.00104
- Drew, T., Võ, M. L.-H., and Wolfe, J. M. (2013). The invisible gorilla strikes again: sustained inattention blindness in expert observers. *Psychol. Sci.* 24, 1848–1853. doi: 10.1177/0956797613479386
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.* 66, 183–201. doi: 10.1037/h0047707
- Everly, G. S., and Lating, J. M. (2013). *A Clinical Guide to the Treatment of the Human Stress Response*. New York, NY: Springer. doi: 10.1007/978-1-4614-5538-7
- Gagnon, S. A., and Wagner, A. D. (2016). Acute stress and episodic memory retrieval: neurobiological mechanisms and behavioral consequences: acute stress and episodic memory retrieval. *Ann. NY Acad. Sci.* 1369, 55–75. doi: 10.1111/nyas.12996
- Gao, H., and Jia, Z. (2017). Detection of threats under inattention blindness and perceptual load. *Curr. Psychol.* 36, 733–739. doi: 10.1007/s12144-016-9460-0
- Gerra, G., Fertomani, G., Zaimovic, A., Caccavari, R., Reali, N., Maestri, D., et al. (1996). Neuroendocrine responses to emotional arousal in normal women. *Neuropsychobiology* 33, 173–181. doi: 10.1159/000119273
- Godoy, L. D., Rossignoli, M. T., Delfino-Pereira, P., Garcia-Cairasco, N., and de Lima Umeoka, E. H. (2018). A comprehensive overview on stress neurobiology: basic concepts and clinical implications. *Front. Behav. Neurosci.* 12:127. doi: 10.3389/fnbeh.2018.00127
- Hanoch, Y., and Vitouch, O. (2004). When less is more: information, emotional arousal and the ecological reframing of the Yerkes-Dodson Law. *Theory Psychol.* 14, 427–452. doi: 10.1177/0959354304044918
- Henckens, M. J. A. G., Hermans, E. J., Pu, Z., Joels, M., and Fernandez, G. (2009). Stressed memories: how acute stress affects memory formation in humans. *J. Neurosci.* 29, 10111–10119. doi: 10.1523/JNEUROSCI.1184-09.2009
- Hoscheidt, S. M., LaBar, K. S., Ryan, L., Jacobs, W. J., and Nadel, L. (2014). Encoding negative events under stress: high subjective arousal is related to accurate emotional memory despite misinformation exposure. *Neurobiol. Learn. Mem.* 112, 237–247. doi: 10.1016/j.nlm.2013.09.008
- Hulse, L. M., and Memon, A. (2006). Fatal impact? The effects of emotional arousal and weapon presence on police officers' memories for a simulated crime. *Legal Criminol. Psychol.* 11, 313–325. doi: 10.1348/135532505X58062
- Hüttermann, S., and Memmert, D. (2012). Moderate movement, more vision: effects of physical exercise on inattention blindness. *Perception* 41, 963–975. doi: 10.1068/p7294
- Hyman, I. E., Boss, S. M., Wise, B. M., McKenzie, K. E., and Caggiano, J. M. (2009). Did you see the unicycling clown? Inattention blindness while walking and talking on a cell phone. *Appl. Cogn. Psychol.* 24, 597–607. doi: 10.1002/acp.1638
- Hyman, I. E. Jr, Sarb, B. A., and Wise-Swanson, B. M. (2014). Failure to see money on a tree: in attention blindness for objects that guided behavior. *Front. Psychol.* 5:356. doi: 10.3389/fpsyg.2014.00356
- Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., and Koo, B. H. (2018). Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatr. Investig.* 15, 235–245. doi: 10.30773/pi.2017.08.17
- Kim, J. S. C., Vossel, G., and Gamer, M. (2013). Effects of emotional context on memory for details: the role of attention. *PLoS ONE* 8:e77405. doi: 10.1371/journal.pone.0077405
- Kirschbaum, C., Pirke, K. M., and Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 76–81. doi: 10.1159/000119004
- Kocab, K., and Sporer, S. L. (2016). "The weapon focus effect for person identifications and descriptions: a meta-analysis," in *Advances in Psychology and Law*, vol. 1, eds M. Miller and B. Bornstein (New York, NY: Springer), 71–117. doi: 10.1007/978-3-319-29406-3_3
- Koivisto, M., Hyönä, J., and Revonsuo, A. (2004). The effects of eye movements, spatial attention, and stimulus features on inattention blindness. *Vis. Res.* 44, 3211–3221. doi: 10.1016/j.visres.2004.07.026

AUTHOR CONTRIBUTIONS

AW drafted the initial manuscript and revised subsequent drafts. AT provided critical feedback and revisions. All authors approved the final version of the manuscript.

- Krix, A. C., Sauerland, M., Raymaekers, L. H., Memon, A., Quaedflieg, C. W., and Smeets, T. (2016). Eyewitness evidence obtained with the Self-Administered Interview® is unaffected by stress. *Appl. Cogn. Psychol.* 30, 103–112. doi: 10.1002/acp.3173
- Lang, P. J., and Bradley, M. M. (2007). “The international affective picture system (IAPS) in the study of emotion and attention,” in *Handbook of Emotion Elicitation And Assessment, 1st Edn.*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press), 29–46.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*. Technical Report A-8. University of Florida, Gainesville, FL.
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cogn. Psychol.* 7, 560–572. doi: 10.1016/0010-0285(75)90023-7
- Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn. Mem.* 12, 361–366. doi: 10.1101/lm.94705
- Mack, A. (2003). Inattention blindness: looking without seeing. *Curr. Dir. Psychol. Sci.* 12, 180–184. doi: 10.1111/1467-8721.01256
- Mack, A., and Rock, I. (1998). *Inattentional Blindness*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/3707.001.0001
- Mackersie, C., and Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: heart rate variability and skin conductance. *Ear Hear.* 37(Suppl. 1), 118S–125S. doi: 10.1097/AUD.0000000000000305
- Marr, C., Otgaar, H., Sauerland, M., Quaedflieg, C. W., and Hope, L. (2020). The effects of stress on eyewitness memory: a survey of memory experts and laypeople. *Mem. Cogn.* 1–21. doi: 10.3758/s13421-020-01115-4
- Morewitz, S. J. (2019). “Suspicious objects,” in *Clinical and Psychological Perspectives on Foul Play*, ed S. J. Morewitz (New York, NY: Springer International Publishing), 225–238. doi: 10.1007/978-3-030-26840-9_9
- Murphy, G., and Greene, C. M. (2016). Perceptual load affects eyewitness accuracy and susceptibility to leading questions. *Front. Psychol.* 7:1322. doi: 10.3389/fpsyg.2016.01322
- Näsholm, E., Rohlfsing, S., and Sauer, J. D. (2014). Pirate stealth or inattentional blindness? The effects of target relevance and sustained attention on security monitoring for experienced and naïve operators. *PLoS ONE* 9:e86157. doi: 10.1371/journal.pone.0086157
- Neisser, U., and Becklen, R. (1975). Selective looking: attending to visually specified events. *Cogn. Psychol.* 7, 480–494. doi: 10.1016/0010-0285(75)90019-5
- New, J. J., and German, T. C. (2015). Spiders at the cocktail party: An ancestral threat that surmounts inattentional blindness. *Evol. Hum. Behav.* 36, 165–173. doi: 10.1016/j.evolhumbehav.2014.08.004
- Pammer, K., Bairnsfather, J., Burns, J., and Hellsing, A. (2015). Not all hazards are created equal: the significance of hazards in inattentional blindness for static driving scenes. *Appl. Cogn. Psychol.* 29, 782–788. doi: 10.1002/acp.3153
- Payne, J. D., Jackson, E. D., Hoscheidt, S., Ryan, L., Jacobs, W. J., and Nadel, L. (2007). Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories. *Learn. Mem.* 14, 861–868. doi: 10.1101/lm.743507
- Qi, M., Gao, H., and Liu, G. (2018). The effect of mild acute psychological stress on attention processing: an ERP study. *Exp. Brain Res.* 236, 2061–2071. doi: 10.1007/s00221-018-5283-6
- Redlich, D., Schnuerch, R., Memmert, D., and Kreitz, C. (2019). Dollars do not determine detection: monetary value associated with unexpected objects does not affect the likelihood of inattentional blindness. *Q. J. Exp. Psychol.* 72, 2141–2154. doi: 10.1177/1747021819835148
- Rivardo, M., Brown, K., Rodgers, A., Maurer, S., Camaione, T., Minjock, R., et al. (2011). Integrating inattentional blindness and eyewitness memory. *North Am. J. Psychol.* 13, 519–538.
- Ruz, M., Worden, M. S., Tudela, P., and McCandliss, B. D. (2005). Inattentional amnesia to words in a high attentional load task. *J. Cogn. Neurosci.* 17, 768–776. doi: 10.1162/0898929053747685
- Sänger, J., Bechtold, L., Schoofs, D., Blaszkewicz, M., and Wascher, E. (2014). The influence of acute stress on attention mechanisms and its electrophysiological correlates. *Front. Behav. Neurosci.* 8:353. doi: 10.3389/fnbeh.2014.00353
- Sauerland, M., Raymaekers, L. H., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., et al. (2016). Stress, stress-induced cortisol responses, and eyewitness identification performance. *Behav. Sci. Law* 34, 580–594. doi: 10.1002/bsl.2249
- Schwabe, L., and Schächinger, H. (2018). Ten years of research with the Socially Evaluated Cold Pressor Test: data from the past and guidelines for the future. *Psychoneuroendocrinology* 92, 155–161. doi: 10.1016/j.psyneuen.2018.03.010
- Shaw, I. I. I., J. S., and McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law Hum. Behav.* 20, 629–653. doi: 10.1007/BF01499235
- Shields, G. S., Sazma, M. A., and Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: a meta-analysis and comparison with cortisol. *Neurosci. Biobehav. Rev.* 68, 651–668. doi: 10.1016/j.neubiorev.2016.06.038
- Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception* 28, 1059–1074. doi: 10.1068/p281059
- Simons, D. J., and Schlosser, M. D. (2017). Inattentional blindness for a gun during a simulated police vehicle stop. *Cogn. Res. Princip. Implic.* 2:37. doi: 10.1186/s41235-017-0074-3
- Stothart, C. R., Wright, T. J., Simons, D. J., and Boot, W. R. (2017). The costs (or benefits) associated with attended objects do little to influence inattentional blindness. *Acta Psychol.* 173, 101–105. doi: 10.1016/j.actpsy.2016.12.012
- Thomas, A. K., Bulevich, J. B., and Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: implications for retrieval enhanced suggestibility. *J. Mem. Lang.* 63, 149–157. doi: 10.1016/j.jml.2010.04.004
- Ward, E. J., and Scholl, B. J. (2015). Inattentional blindness reflects limitations on perception, not memory: evidence from repeated failures of awareness. *Psychonom. Bull. Rev.* 22, 722–727. doi: 10.3758/s13423-014-0745-8
- Weinberg, H. I., Wadsworth, J., and Baron, R. S. (1983). Demand and the impact of leading questions on eyewitness testimony. *Mem. Cogn.* 11, 101–104. doi: 10.3758/BF03197667
- Wiemer, J., Gerdes, A. B. M., and Pauli, P. (2013). The effects of an unexpected spider stimulus on skin conductance responses and eye movements: an inattentional blindness study. *Psychol. Res.* 77, 155–166. doi: 10.1007/s00426-011-0407-7
- Wolfe, J. M. (1999). “Inattentional amnesia,” in *Fleeting Memories*, ed V. Coltheart (Cambridge, MA: MIT Press), 71–94.
- Yeghyan, N. S., and Yonelinas, A. P. (2011). Encoding details: positive emotion leads to memory broadening. *Cogn. Emot.* 25, 1255–1262. doi: 10.1080/02699931.2010.540821

Conflict of Interest: 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 © 2021 Wulff and Thomas. 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) and the copyright owner(s) 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.



Manipulations of List Type in the DRM Paradigm: A Review of How Structural and Conceptual Similarity Affect False Memory

Jennifer H. Coane^{1*}, Dawn M. McBride², Mark J. Huff³, Kai Chang¹, Elizabeth M. Marsh² and Kendal A. Smith³

¹ Colby College, Waterville, ME, United States, ² Department of Psychology, Illinois State University, Normal, IL, United States,

³ The University of Southern Mississippi, Hattiesburg, MS, United States

OPEN ACCESS

Edited by:

Maria Soledad Beato,
University of Salamanca, Spain

Reviewed by:

Marek Nieznański,
Cardinal Stefan Wyszyński
University, Poland

Carlos J. Gomez-Ariza,
University of Jaén, Spain

*Correspondence:

Jennifer H. Coane
jhcoane@colby.edu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 16 February 2021

Accepted: 29 April 2021

Published: 31 May 2021

Citation:

Coane JH, McBride DM, Huff MJ, Chang K, Marsh EM and Smith KA (2021) Manipulations of List Type in the DRM Paradigm: A Review of How Structural and Conceptual Similarity Affect False Memory. *Front. Psychol.* 12:668550. doi: 10.3389/fpsyg.2021.668550

The use of list-learning paradigms to explore false memory has revealed several critical findings about the contributions of similarity and relatedness in memory phenomena more broadly. Characterizing the nature of “similarity and relatedness” can inform researchers about factors contributing to memory distortions and about the underlying associative and semantic networks that support veridical memory. Similarity can be defined in terms of semantic properties (e.g., shared conceptual and taxonomic features), lexical/associative properties (e.g., shared connections in associative networks), or structural properties (e.g., shared orthographic or phonological features). By manipulating the type of list and its relationship to a non-studied critical item, we review the effects of these types of similarity on veridical and false memory. All forms of similarity reviewed here result in reliable error rates and the effects on veridical memory are variable. The results across a variety of paradigms and tests provide partial support for a number of theoretical explanations of false memory phenomena, but none of the theories readily account for all results.

Keywords: false memory, DRM paradigm, activation monitoring theory, fuzzy trace theory, global matching

INTRODUCTION

Over the past 25 years, the Deese-Roediger-McDermott (DRM; Deese, 1959; Roediger and McDermott, 1995) paradigm for studying experimentally induced false memories has been used in thousands of studies. To give a simple example, as of January 2021, a search on SCOPUS indicated the 1995 paper has been cited over 2,450 times. The basic findings of this corpus of research suggest that studying lists of related words (e.g., *butter, knife, slice*) elicits reliable false recall and recognition of a non-presented critical item (CI, e.g., *bread*).

The false memories obtained using this paradigm are robust across testing formats [see Gallo (2006, 2010), for reviews], emerge after encoding as few as four or five related list items (Coane et al., 2007, in preparation), and persist weeks to months following study [Seamon et al., 2002; Coane et al., manuscript in preparation; but see Colbert and McBride (2007)]. The *DRM false memory illusion* is highly replicable, both between and within participants (Zwaan et al., 2018), indicating that awareness of the paradigm does not eliminate the effect. Indeed, administering a warning to avoid recalling the CI prior to study (Gallo et al., 1997; Huff et al., 2012) or promoting more distinctive encoding processes [e.g., Israel and Schacter, 1997, see Huff et al. (2015), for a review and

meta-analysis] reduces, but does not eliminate, the illusion. When queried, participants are highly confident in their accuracy for these falsely remembered items (e.g., Roediger and McDermott, 1995) and will make confident source attributions to the non-presented lures (Payne et al., 1996). Further underscoring the strength of these false memories, when assessing the phenomenological reports given by participants, they are likely to identify the CI as *remembered* rather than *known*, suggesting that specific episodic details are associated with the retrieval of the lure (Brainerd et al., 2003; Geraci and McCabe, 2006). Furthermore, the effect is present across age groups, languages, and in individuals with dementia and other forms of neurological damage (Balota et al., 1999). In fact, normative studies have generated DRM lists across several languages, including Spanish (Anastasi et al., 2005; Beato and Díez, 2011), French (Brédart, 2000), Italian (Senese et al., 2010), Portuguese (Albuquerque, 2005), and Romanian (Horoită and Opre, 2020). DRM studies have also been conducted in Chinese (Guo et al., 2004) and Japanese (Kawasaki and Yama, 2006) languages, replicating the false memory effect in non-alphabetical languages. Despite obvious language differences across these studies, patterns found in the DRM paradigm are remarkably consistent: Manipulations that affect false memory rates using English materials show similar patterns in other languages. In sum, this work highlights the malleability of memory and the importance of examining how related words can give rise to high-confidence memory errors.

DEFINING SIMILARITY

The use of list-learning paradigms that are dependent upon similarity between list items and CIs to explore false memory has underscored several critical findings about the contributions of similarity and relatedness in memory phenomena more broadly. Such issues are at the core both of research in episodic memory and in understanding the organization of knowledge in semantic memory. In fact, these questions have been examined, in one way or another, for decades, if not since the beginning of traditional memory research. In his seminal study, Ebbinghaus (1885/1913) purposely selected meaningless syllables to avoid the potential contamination of meaning-based information in recall. The use of meaningless stimuli, which, by definition, are unrelated to one another, was a hallmark of early memory research (McGeoch, 1942) as scholars attempted to uncover memory principles and processes.

However, meaning-level information, broadly defined, exerts a powerful effect on many cognitive processes. Cognitive systems are highly adept at applying meaning to information in the environment through pattern recognition processes and the application of top-down processes (i.e., prior knowledge, context information). Organizational processes such as the Gestalt principles underscore how readily the cognitive system uses surface level features, such as proximity and similarity, to create a coherent representation of the environment. Such processes give rise to phenomena such as visual illusions and pareidolia (i.e., detecting faces in non-face stimuli; Ichikawa

et al., 2011), and are critical for how the mind organizes a complex environment. These organizational principles further extend to memory systems, where reliance on structures such as schemas and categories support encoding and retrieval processes by simplifying the amount of information to which an individual must attend. For example, categorization allows one to quickly retrieve previously known information about a novel member of a category and reduces the need for re-learning (Bruner et al., 1956). Relying on schemas and scripts similarly minimizes the amount of effort and attention necessary for navigating the world. Such reliance on prior knowledge systems and structures, however, does come at a cost: namely, the introduction of errors through a reconstructive memory process, which occurs when previous experiences are retrieved (Bartlett, 1932; Bergman and Roediger, 1999; Schacter, 2001; Roediger and DeSoto, 2015). Reconstructive processes might be more likely to occur when information is poorly encoded, due perhaps to inattention, or forgotten, due to decay or interference, leading to increased reliance on existing knowledge structures to “fill in the gaps” in a memory. This is illustrated in many types of memory errors, from schema-driven errors (e.g., falsely recalling books in an office; Brewer and Treyens, 1981) to misinformation effects (Loftus, 2005). The DRM paradigm is similarly dependent on these established knowledge systems stored in semantic memory, such that studying the list items results in the increased accessibility or familiarity of the CI due to the shared meaning between items.

Several other paradigms in memory research have underscored the powerful effects of relatedness in short- and long-term memory. For example, in short-term memory tasks, phonological similarity effects (Conrad, 1963; Baddeley, 1964; Wickelgren, 1965) refer to high error rates for phonologically and/or orthographically similar items and reflect the high reliance on surface-level features in short-term memory. In Baddeley's working memory model (Baddeley and Hitch, 1974), this is mediated via the active maintenance of verbal information in the phonological loop. When short sets of items from the same category are studied, the degree of proactive interference observed is dependent on the degree of similarity between sets (Wickens, 1970); this suggests that, in addition to phonological information, semantic information is also processed and preserved in short-term memory. Classic work examining long-term recall and recognition also highlights the powerful effects of relatedness on retention. For example, related word lists are recalled better than unrelated ones (Huff et al., 2011). Lists of categorically related words are not only recalled better than unrelated ones but show clustering effects such that the shared meaning provides organizational structure at retrieval (Bousfield, 1953). Such effects occur spontaneously or when category cues are provided at retrieval (Tulving and Pearlstone, 1966).

Clearly, as this brief (and selective) review highlights, similarity along multiple dimensions exerts powerful effects on memory. This raises the question of how similarity is defined. Characterizing the nature of “similarity and relatedness” can inform researchers about factors contributing to memory distortions and about the underlying networks that support episodic memory. An examination of the semantic memory literature reveals that operationalizing relatedness in terms of

meaning is far from straightforward. One of the fundamental debates in the literature concerns whether relatedness is driven by lexical-level associations or by semantic similarity [see Kumar (2021), for a recent review]. The former refers to the types of associations in the language that are due to co-occurrence or other types of experience. For example, *cat* and *dog* are related because they tend to be encountered in similar contexts (e.g., both are house pets, or are found in idiomatic expressions such as “it’s raining cats and dogs”). Items like *dog* and *leash* are related because of a functional association [although some researchers argue that functional relations are a form of semantic relatedness; see Lucas (2000), Wu and Barsalou (2009)]. Conversely, semantic similarity is defined as similarity in terms of overlap of primitive features or category membership. In this case, *cat* and *dog* are related because they generally share physical features (e.g., fur, claws, four legs) and belong to the same category. In an extensive review of the semantic priming literature, Hutchison (2003) concluded that automatic semantic priming, which refers to the facilitation observed when a target item (e.g., *dog*) is processed faster and/or more accurately when it is preceded by a related prime (e.g., *cat*) than an unrelated prime (e.g., *pen*), can occur following both associative relations and feature overlap. Overall, semantic priming tasks suggest that access to a target can be facilitated by associations and semantic similarity.

One of the most compelling lines of evidence in support of associative priming comes from mediated priming tasks. In such experiments, prime-target pairs are developed such that the prime and target are not directly related to each other, but indirectly related to a non-presented mediator that connects them (e.g., *lion-stripes*, in which the non-presented mediator is *tiger*; Balota and Lorch, 1986). Because mediated pairs do not share any features directly, they provide strong support for associative accounts of priming (Hutchison, 2003). Conversely, priming from synonyms and antonyms is consistent with feature overlap accounts of priming. The traditional DRM lists used in most research contain a mixture of semantic and associative relations and as such, are consistent with both semantic and associative accounts of priming. As we discuss below (see Theories of False Memory in the DRM Paradigm), whether false memories in the DRM paradigm are due to semantic and/or associative processes is at the center of theoretical debates about the mechanisms that give rise to errors. Careful manipulation of the type of relation between list items and CIs can refine these theories.

Although meaning-based similarity is a powerful determinant of memory errors, cognitive systems are also highly tuned to detecting and identifying patterns and similarity in terms of surface features. In addition to semantic priming, priming can also occur when primes and targets are related phonologically and/or orthographically (i.e., form priming). It is beyond the scope of this work to provide a full review of the literature on these forms of priming effects [for reviews see Rastle (2007), Farrell et al. (2012), Humphreys et al. (2016)]; such effects are robust in both spoken and written language. Whether priming is facilitatory or inhibitory depends on several factors, such as stimulus onset asynchrony, task demands, and masking, to name a few. However, there is consistent evidence that related primes

do affect the time it takes to retrieve a target. One important element is that such effects are due to lexical level factors and are independent of meaning. Thus, form priming appears to be distinct from semantic priming.

Similarly, memory errors can emerge due to the encoding of formally or structurally similar items. For example, Koutstaal and Schacter (1997), Koutstaal et al. (1999, 2003) have repeatedly demonstrated false memories based on perceptual information for images (both meaningful and abstract), and Zeelenberg et al. (2005) reported false memories following the study of lists of non-words. Further, variants of the DRM paradigm using items that are related phonologically and/or orthographically have produced robust false memory rates (e.g., Sommers and Lewis, 1999; Watson et al., 2003). Thus, similarity between list items and CIs—in terms of meaning or surface features—result in memory errors. In the present work, we provide an overview and review of research using the DRM paradigm to examine several manipulations of “similarity.” The findings from this work can be used in testing theories for explaining false memory in this paradigm and further our understanding of how these factors affect memory more generally. We note that most of the work using verbal materials in this area has been conducted in English; where relevant, we include evidence from other linguistic and alphabetic systems.

THEORIES OF FALSE MEMORY IN THE DRM PARADIGM

There are several current theoretical explanations of false memory. The activation-monitoring framework (AMF; Roediger et al., 2001) emphasizes the role of automatic spreading activation in lexical/semantic networks (e.g., Collins and Loftus, 1975) that increases the accessibility or familiarity of the CI through shared pathways. In support of the role of activation-based processes, the degree to which list items and CIs are associated (backward associative strength, BAS) based on free association norms (Nelson et al., 2004), is the best predictor of false recall and the second-best predictor of false recognition (Roediger et al., 2001). In addition, evidence that false memories emerge under divided attention conditions (Peters et al., 2008), occur following incidental encoding tasks (Dodd and Macleod, 2004), and via the presentation of list item distractors during a recognition test (Coane and McBride, 2006), further supports the automatic nature of this process. The second process, source monitoring (Johnson et al., 1993), can result in the misattribution of this activation to a studied event rather than to the internal generation of the CI. Source monitoring is a controlled, resource-demanding process that is necessary to avoid errors and, when it fails, an increase in false memories is observed. Evidence in support of this process is found in studies that have shown a reduction in errors when an explicit warning to avoid critical intrusions is given (Gallo et al., 1997; McCabe and Smith, 2002; Neuschatz et al., 2003), for individuals with higher working memory (e.g., Watson et al., 2005), and younger (vs. older) adults (Balota et al., 1999), where the last two groups typically possess stronger memory monitoring capacities. Thus, the AMF includes two

opponent processes: an error-increasing activation process and an error-reducing monitoring process.

Separately, fuzzy-trace theory (FTT; Brainerd and Reyna, 2002) presupposes that upon experiencing an event, two parallel traces are stored: a verbatim trace, which preserves item-specific and contextual information, and a persistent gist trace, which is based on the extraction of the general meaning of the encoded information. FTT attributes false memories to reliance on the gist trace based on the similarity between list items. CIs are consistent with the gist or thematic coherence of the list, leading to errors, whereas memory for list items can be supported by both gist and verbatim traces. Verbatim memory can serve to reduce errors through a process referred to as recollection rejection. Verbatim traces tend to decay more rapidly, whereas gist traces are persistent (e.g., Abadie and Camos, 2019). Thus, FTT incorporates an error-increasing mechanism (gist memory) and an error-reducing mechanism (verbatim memory/recollection rejection). Evidence consistent with FTT includes findings that false memories are more persistent than veridical memories [Toglia et al., 1999; Seamon et al., 2002; but see Colbert and McBride (2007)] and higher rates of false memories for lists with a stronger thematic coherence (Cann et al., 2011; Carneiro et al., 2014).

Finally, global matching models (GMM) such as MINERVA2 (Hintzman, 1986, 1988; Arndt and Hirshman, 1998) suggest that items are encoded in memory as feature vectors; related items share features, thereby leading to similar traces. The extent to which a retrieval probe matches vector traces stored in memory determines whether an item is recognized as studied or not. The feature matching process results in a familiarity echo signal that is stronger with more feature overlap between a probe and the memory traces. Feature overlap is summed for each study-test item comparison, which results in an activation value. Activation values are then summed across item comparisons to provide a familiarity level for each test probe. Because CIs, by definition, share features with all studied list items, when presented as probes they will likely elicit a strong level of familiarity due to the summed activation from the feature matches across the items, thereby leading to an incorrect old response. One advantage of these models is that they can readily account for false memories for non-words (Zeelenberg et al., 2005) or abstract images (Koutstaal et al., 2003), as they do not require pre-existing mental representations that would result in activation or gist extraction. Some also assume that when a test probe is compared with an encoded trace and overlapping features are found, all of the traces' features are activated as activation spreads from those features that overlap with the probe (Hintzman, 1986). Furthermore, these models have flexibility in defining what the features that are stored in memory traces are and include semantic features as well as surface level features.

A common factor across these theories is that they attribute an important role to similarity in veridical and false memory effects. As noted above, similarity can be defined in terms of lexical/associative properties (e.g., shared connections in associative networks), semantic properties (e.g., shared conceptual and taxonomic features), or structural properties (e.g., shared orthographic or phonological features). In the AMF,

similarity is defined in terms of connections between nodes in lexical and semantic networks; in FTT, similarity is based on gist traces that are meaning-based; and in GMM, similarity emerges through shared features that are stored with each memory trace, broadly defined. Thus, exactly how similarity or relatedness is defined varies somewhat across theories. By manipulating the type of list and its relationship to a non-studied CI, we have explored the effects of varying types of similarity on veridical and false memory in a variety of memory tasks, assessing short- and long-term memory, using recall and recognition tasks, priming tasks, and in younger and older adults. Here, we review prior research manipulating list type. To preview our conclusions, all forms of similarity we have manipulated thus far have resulted in reliable error rates and the effects on veridical memory were variable, suggesting that multiple forms of relatedness support both accurate and erroneous memory.

DECOMPOSING SEMANTIC AND ASSOCIATIVE SIMILARITY

As noted above, a core question in the field of semantic memory concerns the nature of the representations, their organization, and, by extent, how relatedness is defined. If knowledge is primarily represented and organized along shared meaning, such as category relatedness, then shared primitive features (e.g., has skin, has four legs, breathes) would be critical in determining whether two concepts, and the words that represent them, are related. Conversely, if organization relies more on shared occurrence or broader principles of association, the connections between items would not depend on shared features as much as on more broadly defined relations and on co-occurrence in similar contexts (e.g., Landauer and Dumais, 1997). Manipulations of list type in the DRM paradigm have explored the question of what sorts of items elicit greater false memory. In broad terms, researchers have distinguished between categorical lists, in which the CI is either a member of the same category as the list items (e.g., a list of fruits with *apple* as the CI) or a category superordinate (e.g., a list of fruits with *fruit* as the CI), and associative lists, in which the list items and CIs are related based on free association norms (e.g., a list of items related to the CI *fruit* includes words such as *pie*, *basket*, and *owl*). The first two types reflect semantic level relationships, whereas the latter relies more on associative and lexical relationships (although the two types of relationships are often confounded). In what follows, we use the term *categorical* to refer to relationships that depend on category co-occurrence or membership and shared features and the term *associative* to refer to relationships based on lexical co-occurrence.

Early research attempting to examine the roles of associative and categorical relations in the DRM found that associative lists resulted in higher false memory (Buchanan et al., 1999) and larger priming effects (Smith et al., 2002) than categorical lists. Other work (Dewhurst et al., 2007; Knott and Dewhurst, 2007; Knott et al., 2012) found that manipulations such as divided attention at encoding or manipulations of list presentation (e.g., blocked vs. random) exerted parallel effects on false memories

for both list types. An important factor, however, is that BAS was higher in associative than categorical lists, introducing a potential confound. When BAS was matched, however, false memories were equivalent across list types, although the lists were not “pure” in that associative lists also included some category coordinates (Knott et al., 2012). In contrast, Park et al. (2005) reported higher rates of false recall and false recognition for associative than for categorical lists, even after controlling for BAS. Because BAS is a strong driver of false memory (Roediger et al., 2001) and some types of semantic relations, specifically synonyms, situation features, and taxonomic relations, are predictors of BAS (Cann et al., 2011), it can be difficult to tease apart the effects of association from those of shared features. To address this, we (Coane et al., 2016, 2020) developed novel lists that were matched in BAS but differed in whether they shared basic features. Non-categorically associated (NCA) lists consisted of associates to a CI that did not share features or come from the same category (e.g., *dog* CI with *bone*, *bark* list items), whereas categorical plus associative lists (C+A) lists had equivalent levels of BAS as the NCA lists, but also shared features and/or came from the same category (e.g., *dog* CI with *cat*, *wolf* list items). Thus, for each CI, we had two lists: one that shared features and one that did not. Norming studies confirmed that feature similarity was greater in the C+A lists than in the NCA lists. Importantly, the lists were matched not only in BAS, but along several other dimensions [e.g., word frequency, connectivity, semantic distance according to the Latent Semantic Analysis; Landauer and Dumais, 1997; see Coane et al. (2016), for details on list development and pre-testing]. We underscore the importance of matching the lists on key dimensions that are known to influence word recognition and lexical access. Although the relationship between access and activation is not fully understood, if a given item is accessed faster, this could result in more activation than an item that requires more time to access (cf. Westbury et al., 2002).

Under these conditions, C+A lists have reliably elicited higher false memory rates than NCA lists in both recall and recognition tasks [Coane et al., 2016, 2020; see also Montefinese et al. (2015)]. This suggests that, above and beyond activation as captured by BAS alone, semantic or feature similarity results in an increase in false memory, a phenomenon we refer to as a *feature boost*. One possibility is that shared features provide additional activation beyond that which comes from associations. Semantic priming research suggests that, when primes and targets are category coordinates, such as *goat-dog*, coordinates that are associated, such as *cat-dog*, generate larger priming effects than those that are not, a phenomenon referred to as an *associative boost* (Hutchison, 2003). In other words, associations and semantic similarity might exert additive effects on target or CI accessibility.

An alternative account of the feature boost is that there are differences in the extent to which error-reducing mechanisms, such as monitoring or recollection rejection, are effective. Although warnings generally reduce false memories, their effectiveness varies with the identifiability of the CI. Specifically, when the CI is easier to identify (Neuschatz et al., 2003; Carneiro et al., 2012; Huff et al., 2012, 2018) or when CIs are strongly thematically related to lists (Carneiro et al., 2014), warnings are

more effective. To test whether the effect was due to differential CI identifiability, we (Coane et al., 2020) compared false memory for C + A and NCA lists after a warning or after participants were instructed to guess the CI. The guessing task provided an estimate of how accurately participants could identify the CI, as well as an implied warning given that participants were tasked with identifying a “missing” item. Although CIs from C + A lists were more difficult to identify, an explicit warning was equally effective at reducing errors in across list types. Importantly, when participants were able to correctly identify the CI, false alarms were equivalent across list types. Further, conditional analyses on recognition responses as a function of prior recall (one of the post-list task conditions) indicated that the feature boost only occurred when CIs had not been recalled previously. In other words, prior recall or correct guessing eliminated the feature boost effect, indicating that the feature boost in recognition only occurred when CIs were not explicitly identified either by guessing or prior recall. Thus, the feature boost does appear to be due, at least in part, to difficulties in discriminating, and thus rejecting, the CI.

In addition to our approach to separate the effects of associative and thematic similarity by holding BAS constant, other work has addressed the core question of how to tease apart semantic and associative similarity. Specifically, Brainerd et al. (2020) created a pool of 120 four-item DRM lists that varied widely in their mean BAS values and in their degree of semantic similarity between list items and CIs. All lists were normed to determine a measure of gist strength (GS) and empirically examine how BAS and GS jointly influence false recognition. They concluded that GS reliably predicted CI false recognition across levels of BAS, whereas mean BAS only predicted CI false recognition when mean GS was low.

Although in the present review we have focused on the comparison of categorical and associative lists, another manipulation is worth noting. Hutchison and Balota (2005) attempted to decouple associative strength from meaning or gist by developing lists in which the CI was a homograph (e.g., *fall*). In critical conditions, the list included items related to both meanings of the CI (e.g., *stumble*, *trip* as well as *autumn*, *leaves*). Thus, the meaning of the list items was associated to the CI at a lexical level, but the two halves of the list converged on two distinct meanings at the semantic level. False memories for these lists were compared to DRM lists in which all list items converged upon a CI with a single meaning. Critically, BAS was held constant between list types. False memories were equivalent across homograph lists and DRM lists both when homograph lists were blocked by meaning and when meanings were alternated within a list, indicating that divergent meanings exerted less of an effect than associations of any meaning, even when divergent meanings were less consistent with alternating presentations. More recently however, Huff et al. (2015) reported that blocked homograph lists could increase false recall relative to alternated lists, but only when the test was delayed. Taken together, these findings suggest that the influence of meaning and associative information on false memories is complex and variable.

Before discussing the theoretical explanations of these effects, we wanted to address some recent extensions of this work. First, we examined the effects of warnings and guessing the CI from NCA and C+A lists in a sample of older adults. Healthy aging is associated with preserved automatic processes and declines in controlled processing, and older adults have equivalent or elevated false memories compared to younger adults (e.g., Schacter et al., 1997; Balota et al., 1999; Liu and Cao, 2002; Huff and Aschenbrenner, 2018; Pansuwan et al., 2020). Older adults are also less likely than younger adults to benefit from a warning (McCabe and Smith, 2002), and there is evidence of different lifespan time courses for reliance on taxonomic (i.e., categorical) and associative or thematic information (Mirman et al., 2017; Belacchi and Artuso, 2018). Categorical information depends on more abstract and complex knowledge systems, and this organizational system emerges later in childhood compared to associative or thematic organization. As adults age, they benefit less from categorical information in recall compared to younger adults (Huff et al., 2011), whereas the benefits of associative or thematic information show less of a decline. Thus, it was possible that the feature boost would not be observed for older adults, who might rely, instead, more heavily on associative information.

In a study similar in design to Coane et al.'s (2020), a sample of 120 healthy older adults were assigned to one of the same four conditions in the study with young adults (Coane et al., in preparation): guess the CI, complete math problems (a no-retrieval control condition), complete a free recall task, or complete the free recall task with a warning. A final recognition test was then completed. The data were analyzed in conjunction with the younger adult data from Coane et al. (2020) to examine possible age differences. Older adults were less likely to correctly guess the CI for both C+A and NCA lists compared to younger adults and correct guesses were much higher for NCA lists than for C+A lists. Overall, younger and older adults' identification of the CIs was similarly affected by the different types of list relations. Free-recall performance revealed that warnings were effective for both younger and older adults in decreasing false recall of the CI, but again, there was no interaction with age or with condition. Veridical recall did not differ by age or condition, although, consistent with our earlier work, C + A list item recall was significantly greater than NCA item recall.

Older adults' performance on the final recognition test¹ also mirrored the results in younger adults: Warnings reduced false recognition overall, significantly so relative to the no-warning recall condition, indicating they were still effective on a delayed test. Importantly, C + A lists elicited higher false alarms than NCA lists, and, although older adults had higher false alarms overall vs. younger adults, age did not interact with list type or condition.

Thus, older and younger adults do not differ in the extent to which the feature boost occurs. Furthermore, the lack of age differences suggests that the effect might be driven in large part by

processes that are unaffected by age. CI identification, although lower in older than younger adults, appears to be one such process: If the CI is identified, participants appear to be able to use this information to reduce errors. However, because the CIs from C + A lists are less likely to be identified, the process is more likely to fail for these lists, resulting in higher error rates.

We have also examined the feature boost in short-term memory using a modified Sternberg task (Sternberg, 1966) with DRM lists (Xu et al., 2017). Participants studied lists of six items from a C + A or an NCA list and, after each list, responded to a single probe: the CI, a studied item from the set, or a non-studied item from the same list. False alarms to both CIs and non-studied items were higher for C+A than NCA lists; thus, a feature boost was found even under immediate test conditions when very short lists were studied. This suggests that the error-increasing effects of additional similarity in terms of features emerges rapidly and occurs when participants should be able to accurately monitor the source of an item's familiarity (because of the small set size and short delay). Overall, based on both our published and unpublished work, when CIs are both associatively related *and* share features with list items, false memories are greater than when there are no shared features present. The effect is robust across warning conditions, test type, and age.

Turning to how the three primary theoretical accounts can accommodate these findings, the AMF has difficulty explaining the difference in false memories across NCA and C + A lists given the equivalent BAS across list types. The automatic activation process, which is assumed to be predicted by BAS, should be equivalent. One possible explanation is that multiple sources of activation, lexical level associations and semantic level conceptual representations, are independent of one another and contribute separately to affect the activation or accessibility of the CI. This suggests that associations, as captured by free association norms, might be "missing" some important aspects of similarity or relatedness. In addition, the AMF cannot readily account for independent effects of forward associative strength (FAS, or the extent to which CIs elicit the list items in free association; Arndt, 2015), because the CI is not encoded, and thus, would not directly activate the list items, unless a complex process of mediated activation occurs. Conversely, this account can readily accommodate the findings from homograph lists, which appear to be driven by associative strength independent of meaning. As this review shows, similarity is a complex construct that reflects multiple layers and levels that might include several different features, including categorical, associative, lexical, orthographic, and semantic relations.

Conversely, FTT can quite readily accommodate the feature boost in long-term tests: Given that false memories are supported by gist traces, which depend heavily on thematic similarity, items sharing many features are likely to give rise to a stronger, more coherent gist. Thus, the benefit from shared features fits nicely with this theory. In addition, the evidence from Brainerd et al. (2020), that GS predicts false recognition whereas BAS only does so when GS is low, makes FTT a viable explanatory mechanism for these effects because GS is assumed to influence gist extraction and gist-sensitive retrieval processes. However, the fact that C + A CIs were harder to identify could be problematic, given

¹ Recognition analyses are reported on corrected false recognition, calculated by subtracting false alarms to CIs related to non-studied lists from false alarms to CIs related to studied lists (lists were counterbalanced across conditions). Thus, the analyses correct for differential baseline false alarm rates.

that a stronger gist should be easier to identify. In addition, the feature boost effect in short lists with immediate tests is not consistent with this description because with short lists and delays, verbatim information should be more heavily relied on than gist for recognition responses, resulting in a reduction in the feature boost effect. Further, FTT has difficulty accommodating the results of the homograph lists: Theoretically, the mixed lists include two distinct gists, which should result in the storage of two weaker gist traces than lists with a single meaning convergence. However, the increased false recall for blocked over alternated lists on delayed but not immediate tests is consistent with FTT, suggesting that the stronger gist from blocked lists persists over time.

Unlike the other two models, GMMs can account for both sets of findings if the similarity between the stored and test traces is a function of both lexical level associations and primitive features. Because these models assume that many types of features (semantic associations, categorical features, and structural/lexicographic features) are stored during encoding and then compared with test probes during retrieval, any type of overlap would presumably increase activation, and in turn familiarity, for the CIs at test. For both the feature boost effect and false memories for homograph lists, these models describe retrieval responses based on activation of all the different types of features involved. Thus, GMMs predict a feature boost effect for C+A lists due to the larger amount of feature overlap with list item traces compared with NCA lists and a similar (or larger) amount of overlap in features for homograph lists compared with standard DRM lists.

MANIPULATIONS OF ASSOCIATIVE SIMILARITY: MEDIATED LISTS

Despite its difficulty with the feature boost effect, the AMF fares better with results from studies examining mediated associations. One of the primary assumptions of the AMF is that false memories are the result of activation in lexical and semantic networks, which results in the indirect activation of the CI. Consistent with this, longer lists result in higher false alarm rates (Robinson and Roediger, 1997; Coane et al., 2007). Network models, such as Collins and Loftus's (1975), include dense networks of nodes, representing words and concepts, that are connected via pathways; the length of a pathway reflects the strength of the association between two nodes. Through a spreading activation mechanism, activation of one node results in an increase in activation of all nodes connected to it. Furthermore, this initial burst of activation continues to spread in a graded fashion throughout the network. Consistent with this assumption, mediated priming has been obtained for nodes separated by one (Balota and Lorch, 1986; Coane and Balota, 2011) or two (Chwilla and Kolk, 2002) mediators.

According to the AMF, then, studying a list of items that are directly related to an associate of the CI but are not themselves directly related to the CI, should result in increased activation of the CI. For example, for the CI *river*, directly related items

include *water*, *boat*, and *swim*. The mediated list includes *faucet* (related to *water*, but not to *river* according to the Nelson et al., 2004, free association norms), *yacht* (related to *boat*, but not *river*), and *pool* (related to *swim*, but not to *river*). To test whether mediated lists can create false memories for such CIs, Huff and Hutchison (2011) had participants study mediated lists that were immediately followed by a free-recall test or arithmetic problems (i.e., a control condition). Following completion of several study list-recall/arithmetic cycles, all participants completed a final recognition test. On the initial test, false recall of mediated CIs was not found, however, reliable false recognition of mediated CIs was found on the final test. Additionally, mediated false recognition was greater for participants who completed initial recall tests vs. arithmetic problems. Thus, consistent with an implicit AMF, false memory for CIs was found in the absence of a list theme that is directly related to the CI and this pattern was restricted to the delayed recognition test. This pattern indicates that activation processes leading to mediated false memories are likely implicit in nature, do not occur on a recollection-heavy free-recall test, and only emerge on recognition when implicit familiarity-based processes contribute to memory responses. Mediated false recognition effects were also found across different study durations (3,000 ms vs. 500 ms) and following a guessing task where participants were asked to generate the mediated CI immediately after study. Correct guessing of mediated CIs was very low, providing additional evidence regarding implicit processes that occur with mediated lists.

In a second series of experiments, Huff et al. (2012), developed new mediated lists that utilized the same CIs in the original DRM lists [Roediger and McDermott, 1995; Stadler et al., 1999; see Huff et al. (2012), for details]. Thus, for each CI, we included a direct list (i.e., the traditional list of DRM associates) and a mediated list. Participants intentionally encoded each list and performed one of four tasks that were completed immediately after study: arithmetic problems, free recall, free recall with a warning about the nature of the DRM lists, or the guessing task. Again, very few participants were able to successfully guess the CI from the mediated lists, confirming the lack of a clear theme. Warnings were effective at reducing intrusions of the CI for direct lists, and false recall of mediated CIs was very low. However, on a final recognition task, CIs from mediated lists were indeed again recognized at significant rates, consistent with spreading activation processes in AMF. More importantly, warnings and the guessing task *increased* false alarms to mediated CIs relative to the recall task, whereas they decreased false alarms to direct CIs. This *ironic effect of guessing* suggested that the additional elaborative processing participants engaged in while trying to identify the CI increased its activation. However, the difficulty in identifying the CI rendered the warnings—both the explicit warning given in the recall condition and the implied warning in the guessing condition—ineffective. A highly similar pattern of results was observed in an aging sample (Coane et al., 2016). Because older adults typically show preservation of automatic processes along with declines in more controlled processes (Balota et al., 2000), the age invariance of the effect is consistent with the involvement of automatic activation processes.

More recently, we have further evaluated implicit activation processes of mediated lists using a semantic-priming paradigm. Specifically, we (Huff et al., 2021) presented participants with mediated study lists which were immediately followed by a test list in which the CI was presented in the first, third, or eighth test positions to assess the time course of CI activation following study. Participants were tasked with responding to test items using either a semantic-classification task (concrete vs. abstract decisions), a pronunciation task (reading test items aloud), or an old/new recognition task, in which the first two tasks assessed response latencies. Mediated false recognition patterns were again in evidence, and this pattern was consistent across test positions. Importantly, CI priming was also found across test positions and this pattern was greatest in the first test position, but declined (though remained statistically reliable) across the remaining test positions. Priming was similar on both classification and pronunciation tasks. Moreover, the shape of this priming pattern is consistent with spreading-activation processes which are argued to dissipate as time and the number of intervening items between prime(s) and target increases.

Overall, the results of our work examining mediated false memories is consistent with the AMF, given its reliance on spreading-activation mechanisms in existing semantic networks. The ironic effects of guessing and warnings provide further support for the role of monitoring processes: When monitoring fails due to a failure to identify the CI, false alarms increase. Although the activation process is assumed to be automatic and thus does not require attentional processes, sustained attention, and elaboration can increase or maintain activation (Neely, 1977). In addition, the sustained focus and retrieval attempt can strengthen an episodically formed network of associated items resulting in a more persistent trace when a retrieval mode is engaged (Tulving, 1983; Meade et al., 2007). Mediated false memory is challenging for FTT: The lack of thematic consistency renders the gist extraction process difficult if not impossible, as shown by the difficulty participants have in identifying the CI. Thus, it is not clear at present, how this theoretical approach can accommodate these findings. However, GGMs might be able to account for these results due to the activation of all of a trace's features when overlap with the test probe is found (Hintzman, 1986). If one assumes that mediated items share sufficient features with the CI itself to activate all the mediated items' features (e.g., the presentation of the CI *river* activates the "water" feature of list item *faucet*, which matches that feature of the CI), then the features that do overlap between them might be similar enough that the familiarity echo would be of sufficient strength to elicit an incorrect response. However, this assumption would lead one to predict lower rates of false alarms for mediated than direct list CIs but similar patterns for these rates across task conditions, and as the results of these studies show, there is a dissociable pattern of false alarms for mediated and direct list CIs across math, recall, and guessing/warning conditions (Huff and Hutchison, 2011; Huff et al., 2012). Thus, these results present some difficulty for the GMMs.

MANIPULATIONS OF STRUCTURAL SIMILARITY: PHONOLOGICAL AND ORTHOGRAPHIC LISTS

Another form of similarity that results in significant false memories in recall and recognition is in terms of structural elements, namely shared phonemes and graphemes. Orthographically and/or phonologically similar items (hereafter, we refer to them as phonological associates for brevity) share spelling and/or pronunciation with the CI, generally in the absence of shared meaning. For example, the phonological list for the CI *sleep* includes items such as *sleet*, *keep*, and *steep*. The items often, but not necessarily, rhyme or share the first letters or letter clusters. Although the effects of phonological similarity on memory, particularly short-term memory, have been well-documented for decades, after the publication of Roediger and McDermott (1995) these effects were more systematically explored in long-term memory tasks, specifically false memory. Sommers and Lewis (1999) developed phonologically related lists using parameters from the Neighborhood Activation Model (NAM; Luce and Pisoni, 1998), which specifies that, in spoken word recognition, words are organized in similarity networks based on shared phonemes. In this model, a neighbor is an item that differs from another item based on a single phoneme; it is thus comparable to Coltheart et al.'s (1977) metric for quantifying orthographic neighbors (i.e., words that differ by a single grapheme).

Research using the DRM paradigm can inform researchers on the interactions between visual word recognition and memory processes, which are often examined separately (e.g., Westbury et al., 2002; Cortese et al., 2008; Hutchison et al., 2018). Specifically, this paradigm can be used to test predictions of word recognition models and results suggest some long-term maintenance or persistence of verbatim information. Many models of word recognition, whether spoken or written, assume that activation levels are determined by the relationship between words or word components (e.g., onset, syllable, coda). For example, Coltheart et al.'s (1993) dual process model assumes that similarity is defined within each lexicon; for example, in the orthographic lexicon, shared letters result in neighbor activation. In Plaut et al. (1996) model, which is based on parallel distributed processing, shared sublexical nodes become active and facilitate related target access.

Westbury et al. (2002) attempted to identify which specific sublexical components were most important in determining false memory rates by developing lists that shared the initial phoneme, the head (first two phonemes), or the rime (last two phonemes) with a monosyllabic CI. Head and rime related lists both elicited greater false recognition than initial phoneme lists, although all three list types elicited greater false alarms than completely unrelated controls. However, they did not find evidence suggesting the effect was driven by orthographic overlap, even though lists were presented visually. In contrast, Cortese et al. (2008) did find that false memories increased when lists had both high phonological and orthographic overlap, relative to when orthographic overlap was low. These effects

have also been tested in Chinese (a logographic language where characters that differ in written forms can be pronounced similarly, as in English, but similarly written characters can also be pronounced differently). Qu and Ding (2010) found false recognition of CIs with phonologically associated Chinese lists. Furthermore, the degree of similarity did not appear to affect false memory because list items that shared the same syllables or items that only shared onsets or rhymes with the CI produced similar false alarm rates. Using orthographically associated Chinese lists, researchers also observed false recognition of CIs (Qu et al., 2010). In contrast to what was found with phonologically similar lists, false recognition rates were positively related to similarity, as reflected by higher false recognition rates when the CIs shared a larger orthographic overlap (i.e., tonetic symbol) with the list items. This differential pattern might originate from the logographic nature of Chinese language, such that the orthographic information is more associated with semantic content than the phonetic information (Lin and Han, 1999).

In further tests in English, Hutchison et al. (2018) found that adding a single item that shared the initial phonemes with a CI to eight associates in studied lists resulted in a significant boost in false recall and specifically in false *remember* responses to recalled CIs. This is consistent with models of spoken word recognition that suggest that initial phoneme information is important in narrowing a pool of potential neighbors during word recognition, such as the cohort model (Marslen-Wilson and Tyler, 1980). Of particular interest is the finding that the addition of the phoneme overlap item only affected false memory when it was presented after the associates and not before, consistent with the cohort model's predictions that context pre-activates potential targets and phonological information provides a selection mechanism.

In sum, lists consisting of neighbors of a CI have elicited robust false memories in both recall and recognition tasks, in some cases comparable to those obtained using semantically related lists. These findings have since been replicated many times (e.g., Oliver et al., 2016; Finley et al., 2017), confirming that structural similarity between list and lure items results in elevated false memory rates. In some cases, phonological lists yield similar false memory rates to semantic lists (e.g., Sommers and Lewis, 1999), but in other cases, phonological lists elicit lower rates of false memory (e.g., Watson et al., 2003). At very rapid presentation rates (i.e., 20 ms/item), phonological lists yield higher false recall rates than semantic lists; as presentation rates increase, the opposite occurs (McDermott and Watson, 2001; Ballardini et al., 2008). This suggests that early in processing, similarity in terms of surface features is more critical than similarity in terms of semantic properties. Given that long-term memory is heavily dependent on semantic coding and that surface level information is quickly lost (Sachs, 1967), this finding is not particularly surprising.

More recently, we (McBride et al., 2019) examined phonological false memories in a short-term memory task. The phonological coding of verbal material in short-term stores is one of its key characteristics; thus, we expected to find elevated errors to phonologically related CIs. Prior work (e.g., Coane

et al., 2007; Atkins and Reuter-Lorenz, 2008; Flegal et al., 2010) had extended the DRM to short-term memory tasks, finding reliable error rates with lists as short as four or five items and test delays as short as 1 s. In our work, we directly compared semantic and phonological lists using a modified Sternberg (1966) task. Using materials from Watson et al. (2003), we created lists of six phonological and six semantic associates for each CI. After each list, a single probe was presented: the CI, a studied list item, or a non-studied associate from the same list. In addition to replicating findings of semantically driven false alarms that exceeded those for non-studied probes, false alarms to phonologically related CIs were much greater than those to semantically related CIs. Thus, in the short-term, memory errors in the DRM paradigm are more sensitive to structural/phonological similarity to the CI than to semantic similarity. Interestingly, when directly comparing short- and long-term memory for both list types, there is a reversal in false alarms to CIs: On a short-term test, phonological lists elicit more errors than semantic lists; but on a delayed test, semantic lists elicit more errors than phonological lists (Coane et al., in preparation).

There is also work examining the independent contributions of semantic and phonological similarity to CIs. Watson et al. (2003) combined lists of semantic and phonological items to create hybrid lists. These lists elicited a hyper-additive effect—false memories increased substantially when one, two, or three phonological associates were embedded in a list of semantic associates (list length was held constant across these conditions). Finley et al. (2017) further demonstrated that this hyper-additive effect was bi-directional: Including semantic associates in phonological lists or phonological associates in semantic lists resulted in parallel increases in false memories and the effect appeared to plateau after approximately three items were added. However, in a short-term task, these parallel effects were not observed: Inserting one or two semantic associates in a phonological list did not affect false memory rates, whereas inserting one or two phonological associates in a semantic list caused a dramatic increase in false memory rates (McBride et al., 2019). These studies show that replacing semantic associates in standard DRM lists with phonological associates not only increases false memory for the CI, but it also creates an over-additive effect on false memory. These findings highlight the separate contributions of semantic and phonological similarity to false memories.

Additional evidence for the role of structural overlap between list items and CIs comes from work using non-word stimuli (Zeelenberg et al., 2005). Following study of lists of similar non-words, the occurrence of false alarms to non-studied non-word CIs that shared phonemes with the list items suggested that participants were relying on surface features of the presented materials to drive a memory decision. Using lists of pseudohomophones (e.g., *dreem*, *bedd*, *awaik*), Cortese et al. (2008) found reliable false memories for semantically related items (i.e., *sleep*). The fact that pseudo-homophones, which result in reliable priming effects in word recognition tasks (Lukatela and Turvey, 1991), resulted in semantic activation of the CI underscores the importance of phonological information and

its role in activating associative information. Taken together, the results using phonologically and orthographically related items, whether words or non-words, suggest that this type of similarity does negatively affect memory accuracy. Furthermore, the evidence suggests that these effects are stronger at shorter delays, whereas more persistent false memories are observed for semantically related items. To our knowledge, phonological lists have not been used in studies assessing forgetting rates at delays of more than a few minutes.

Although we have focused on verbal materials, it is worth noting that similar effects of perceptual or structural similarity have been observed with non-verbal materials such as categorized pictures (e.g., Seamon et al., 2000; Koutstaal et al., 2001), faces (Shimane et al., 2020), and novel shapes (Koutstaal et al., 2003). However, even with these materials, the false memory effects appear to be mediated by conceptual knowledge. For example, in Koutstaal et al.'s (2003) study, older adults, who typically have higher false memory rates than younger adults, only made more errors when novel shapes were given verbal labels. Similarly, Wang et al. (2018) concluded that false memories from pictures are due to the contributions of conceptual information. Using abstract images as stimuli, Sikora-Wachowicz et al. (2019) observed false memories in younger and older adults in a short-term memory task. Older adults in particular appeared to have increased difficulty discriminating studied items from foils, although this effect was observed for both targets and lure items, suggesting a more generalized deficit in sensitivity.

Given the parallels between the NAM (Luce and Pisoni, 1998) and network models such as Collins and Loftus' model [1975; see also Anderson (1983), Steyvers and Tenenbaum (2005)], which rely on spreading activation principles, at a general level, these findings are consistent with the AMF. Specifically, the basic assumption that neighbors in a network activate one another and result in increased familiarity or accessibility can accommodate both semantic and lexical networks. Although the underlying activation mechanism is similar in accounts of phonological and semantic false memory, in some cases, the mechanisms appear different. For example, Ballou and Sommers (2008) failed to find a correlation between phonological and semantic false alarm rates. Tse et al. (2011) provided evidence that the activation mechanisms involved in semantic and phonological false memories are distinct by examining discriminability measures and observing that CIs from semantically related lists were less discriminable than associates from the same list, whereas the opposite was true for phonologically related CIs. Furthermore, when a remember/know task (Tulving, 1985) designed to assess recollection and familiarity is employed, semantic lists yield more *remember* false alarms (e.g., Gallo et al., 1997; Geraci and McCabe, 2006), whereas phonological lists yield more *know* false alarms. This suggests that participants are basing their judgments on different information in the two cases, which might, in turn, reflect the fact that semantic lists drive attention to conceptual relations and thus more vivid and detailed memories, whereas phonological lists drive attention to surface features and less vivid memories (akin to a level-of-processing effect; Craik and Lockhart, 1972). In general, however, models based on activation

processes can accommodate the results from phonologically related lists.

FTT, conversely, has more difficulty accommodating these results, given that gist traces are typically assumed to be derived from shared meaning. It is plausible that a structural similarity gist could be extracted; however, if that were the case, because gist is more persistent than verbatim traces, the reversal of false memory effects as a function of retention interval might be more difficult for this theory to accommodate. In fact, in recent work by one of the lead proponents of FTT, she claimed that "verbatim memory is memory for surface form, for example, memory representations of exact words, numbers and pictures.... Gist memory is memory for *essential meaning* [emphasis added], the 'substance' of information irrespective of exact words, numbers, or pictures" (Reyna, 2012, p. 333). Thus, lures that are similar to list items along dimensions that are distinct from the meaning of the items, should not be falsely remembered based on this definition of gist. An alternative to false memories based on gist might be through a process of trace disintegration, whereby verbatim traces decay and are recombined incorrectly during a retrieval process. However, using a modeling approach to estimate the contributions of gist and verbatim traces, Nieznański et al. (2019) concluded that the latter account is unlikely and that a more plausible account is that gist traces can emerge based on patterns of perceptual and surface similarity. However, even if one accepts that a "surface" or pattern gist is extracted, hyper-additive effects of phonological associate additions to DRM lists are also difficult to explain unless there is an assumption that two different gist traces are stronger than one.

More problematic for both AMF and FTT are non-word data. Because non-words, by definition, do not have a pre-existing memory trace and do not have meaning, it seems implausible that an activation model could readily accommodate them, given the relative dependence of this approach on pre-existing representations. GMMs, however, could readily account for these, as well as similar findings, by relying on trace similarity. Where GMMs have difficulty, though, is in explaining dissociations in false memories for phonological and semantic lists across short- and long-term delays. In GMMs, it is assumed that features of stored traces decay or are interfered with over time with no distinction made in the type of feature. However, other feature matching models, such as Nairne's (1990) feature model of short-term memory allow for weighting of different types of features over time. If such weighting is assumed, then GMMs can more readily account for the different results found at short- and long-term delays with different types of lists.

CONCLUSIONS

In this review, we have considered the different ways that "similarity" has been defined with an emphasis on how these definitions affect the creation of false memories. This has shown how complex and multi-faceted a concept "similarity" is in this area of research. Generally speaking, memory is affected by all of the definitions considered here: semantic associations, feature overlap, phonological associations, and orthographic/lexical

associations. Further, these effects are mostly age-invariant, as many studies have shown similar effects for young and older adults.

The results across a variety of paradigms and tests provide partial support for many theoretical explanations of false memory phenomena, but none of the theories readily account for all of the results. Specifically, when the word lists consisted of items that were indirectly related to the CI through a mediator but shared no direct relationship to the CI, false memories were most consistent with activation models, which assume that activation spreads through the network in a graded fashion. Because these lists do not have a common gist or many shared features, fuzzy-trace and global-matching models cannot as easily account for the results without additional assumptions. GMMs might be able to predict mediated false memories if the representations of the list items share sufficient features with the CI or with items more strongly associated to the CI. Conversely, when list items were both associatively related and shared primitive semantic features with the CI, false memories were greater than when lists were purely associatively related. Such a finding is inconsistent with activation-monitoring accounts, which prioritize associative strength as a mechanism, but consistent with fuzzy-trace and global-matching models, which emphasize similarity in terms of meaning, features, or gist. The AMF, to account for the feature boost effects, might incorporate multiple levels of activation as a core mechanism. Such an assumption was explicit in Collins and Loftus's (1975) original framework, in which semantic and lexical level information is coded. Thus, if activation spreads along pathways reflecting associative or lexical relations and along separate pathways reflecting semantic or conceptual relations, and this activation is additive in nature, the feature boost could be explained. Finally, false memories observed with lists of phonologically or orthographically related items can be accounted for by activation models, if the assumption is that the networks involved in supporting memory performance are organized in terms of structural similarity as well as semantic/associative similarity. Global-matching models can also account for such findings, provided perceptual/structural features are stored, but only if feature weight is assumed to vary in order to explain dissociations across short- and long-term delays. Finally, because fuzzy-trace models generally assume that gist is based on shared meaning, such an account cannot as readily account for these findings. By including a gist trace that is based on structural similarity it is plausible that FTT could account for these findings, although more specification would be needed. In addition, the assumption of FTT is that gist traces are more persistent than verbatim traces; however, the fact that phonological information appears to decay rapidly might be problematic, in that it would involve different forms of gist with potentially different parameters.

Clearly, the present review fails to provide unequivocal support for any of the major theoretical approaches discussed here. One potentially fruitful avenue for future work would be to critically examine the extent to which the distinction between the effects of structural, lexical, and semantic properties—which have been at the core of memory research for decades (e.g.,

Craik and Lockhart, 1972)—is as well-defined as it is typically assumed to be. For example, early conceptualizations of memory systems assumed that short-term stores primarily maintained structural properties, whereas long-term stores depended on semantic properties [see Greene (2016), for a review]. Evidence from a number of paradigms (e.g., orthographic distinctiveness effects) in addition to the phonological false memories discussed here suggest that structural information is retained in long-term memory and is involved in reconstructive processes. Furthermore, numerous effects point to the maintenance of semantic information in short-term memory (e.g., proactive interference effects, short-term false memories). Although there are clearly important distinctions between meaning-based and perceptually-based properties of stimuli, the evidence presented here does suggest that in many ways such properties exert similar effects on memory. Thus, further examining the extent to which such properties contribute to veridical and false memory and how they might interact with one another could help constrain and clarify the theories proposed here. At the moment, GMMs appear to be the most flexible in accommodating the results, given the way in which they are assumed to represent memory traces as arrays of features that are not necessarily semantic or perceptual in nature. However, such models might need to further specify the additive effects of associative and categorical effects (i.e., the feature boost) and the ways in which perceptual traces appear to have different decay rates than semantic traces.

Before closing, we acknowledge that the present review did not examine important work and converging evidence from neuroscientific and neuropsychological approaches. The question of how similarity and relatedness are represented has a long and rich history within the field, from early studies on category-specific deficits [see Lambon Ralph (2014), for a review] to work on semantic dementia, which reflects a selective loss of semantic memory in conjunction with relative preservation of episodic memory (e.g., Rogers et al., 2004; Lambon Ralph and Patterson, 2008) to more recent studies examining the role of modality-specific in semantic representations (Martin, 2007). Given that false memories driven by semantic and/or associative relations likely rely on pre-existing semantic representations, which are either activated or relied upon to extract gist traces, a more profound understanding of how these substrates contribute to episodic false memories is important. Charest et al. (2014) noted that, although there is a substantial degree of similarity across individuals in how objects are represented neutrally, idiosyncratic differences at the individual level could be predicted using imaging techniques, suggesting that the nature of similarity and relatedness might be even more complex. Whether false memories differ as a function of such individual differences might be a fruitful avenue for further work and could advance our understanding of the acquisition and malleability of shared and idiosyncratic representations of concepts. Additional work relevant to the present review suggests that the anterior temporal lobe is a central component in the maintenance and processing of conceptual and semantic information and that, interestingly, it is more strongly involved in associative processing than in categorical processing (Díez et al., 2017).

Specifically, transcranial direct current stimulation of this region caused a significant decrease in associative false memories while leaving categorical false memories unaffected. Thus, future reviews and empirical work should integrate evidence from multiple protocols.

In sum, we conclude that similarity and relatedness are critical elements in how we encode, store, and retrieve information from memory and that examining errors and distortions provides insight into the functions of memory and of the knowledge base that supports memory performance.

REFERENCES

- Abadie, M., and Camos, V. (2019). False memory at short and long term. *J. Exp. Psychol. Gen.* 148, 1312–1334. doi: 10.1037/xge0000526
- Albuquerque, P. (2005). Produção de evocações e reconhecimentos falsos em 100 listas de palavras associadas portuguesas. *Lab. Psicol.* 3, 3–12. doi: 10.14417/lp.766
- Anastasi, J. S., de Leon, A., and Rhodes, M. G. (2005). Normative data for semantically associated Spanish word lists that create false memories. *Behav. Res. Methods* 37, 631–637. doi: 10.3758/BF03192733
- Anderson, J. R. (1983). A spreading activation theory of memory. *J. Verb. Learn. Verb. Behav.* 22, 261–295. doi: 10.1016/S0022-5371(83)90201-3
- Arndt, J. (2015). The influence of forward and backward associative strength on false memories for encoding context. *Memory* 23, 1093–1111. doi: 10.1080/09658211.2014.959527
- Arndt, J., and Hirshman, E. (1998). True and false recognition in MINERVA2: explanations from a global matching perspective. *J. Mem. Lang.* 39, 371–391. doi: 10.1006/jmla.1998.2581
- Atkins, A. S., and Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Mem. Cognit.* 36, 74–81. doi: 10.3758/MC.36.1.74
- Baddeley, A. D. (1964). Immediate memory and the “perception” of letter sequences. *Q. J. Exp. Psychol.* 16, 364–367. doi: 10.1080/17470216408416394
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1
- Ballardini, N., Yamashita, J. A., and Wallace, W. P. (2008). Presentation duration and false recall for semantic and phonological associates. *Consci. Cognit.* 17, 64–71. doi: 10.1016/j.concog.2007.01.008
- Ballou, M. R., and Sommers, M. S. (2008). Similar phenomena, different mechanisms: semantic and phonological false memories are produced by independent mechanisms. *Mem. Cognit.* 36, 1450–1459. doi: 10.3758/MC.36.8.1450
- Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L. I. I., McDermott, K. B., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognit Neuropsychol.* 16, 361–384. doi: 10.1080/026432999380834
- Balota, D. A., Dolan, P. O., and Duchek, J. M. (2000). “Memory changes in healthy older adults,” in *The Oxford Handbook of Memory*, eds E. Tulving and F. I. M. Craik (Oxford: Oxford University Press), 395–409.
- Balota, D. A., and Lorch, R. F. (1986). Depth of automatic spreading activation: mediated priming effects in pronunciation but not in lexical decision. *J. Exp. Psychol. Learn. Mem. Cognit.* 12, 336–345. doi: 10.1037/0278-7393.12.3.336
- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Beato, M. S., and Diez, E. (2011). False recognition production indexes in Spanish for 60 DRM lists with three critical words. *Behav. Res. Methods* 43, 499–507. doi: 10.3758/s13428-010-0045-9
- Belacchi, C., and Artuso, C. (2018). How taxonomic and thematic associations in semantic memory modulate recall in young through old-old adults. *Psychol. Aging* 33, 1060–1069. doi: 10.1037/pag0000297
- Bergman, E. T., and Roediger, H. L. (1999). Can Bartlett's repeated reproduction experiments be replicated? *Mem. Cognit.* 27, 937–947. doi: 10.3758/BF03201224
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *J. Gen. Psychol.* 49, 229–240. doi: 10.1080/00221309.1953.9710088
- Brainerd, C. J., Chang, M., and Bialer, D. M. (2020). From association to gist. *J. Exp. Psychol. Learn. Mem. Cognit.* 46, 2106–2122. doi: 10.1037/xlm0000938
- Brainerd, C. J., Payne, D. G., Wright, R., and Reyna, V. F. (2003). Phantom recall. *J. Memory Lang.* 48, 445–467. doi: 10.1016/S0749-596X(02)00501-6
- Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Curr. Direct. Psychol. Sci.* 11, 164–169. doi: 10.1111/1467-8721.00192
- Brédart, S. (2000). When false memories do not occur: not thinking of the lure or remembering that it was not heard? *Memory* 8, 123–128. doi: 10.1080/096582100387669
- Brewer, W. F., and Treyens, J. C. (1981). Role of schemata in memory for places. *Cognit. Psychol.* 13, 207–230. doi: 10.1016/0010-0285(81)90008-6
- Bruner, J. S., Goodnow, J. J., and Austin, G. A. (1956). *A Study of Thinking*. New York, NY: John Wiley and Sons.
- Buchanan, L., Brown, N. R., Cabeza, R., and Maitson, C. (1999). False memories and semantic lexicon arrangement. *Brain Lang.* 68, 172–177. doi: 10.1006/brln.1999.2072
- Cann, D. R., McRae, K., and Katz, A. N. (2011). False recall in the Deese-Roediger-McDermott paradigm: The roles of gist and associative strength. *Q. J. Exp. Psychol.* 64, 1515–1542. doi: 10.1080/17470218.2011.560272
- Carneiro, P., Fernandez, A., Diez, E., Garcia-Marques, L., Ramos, T., and Ferreira, M. B. (2012). “Identify-to-reject”: a specific strategy to avoid false memories in the DRM paradigm. *Mem. Cognit.* 40, 252–265. doi: 10.3758/s13421-011-0152-6
- Carneiro, P., Garcia-Marques, L., Fernandez, A., and Albuquerque, P. (2014). Both associative activation and thematic extraction count, but thematic false memories are more easily rejected. *Memory* 22, 1024–1040. doi: 10.1080/09658211.2013.864680
- Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., Kriegeskorte, N., and Ungerleider, L. G. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14565–14570. doi: 10.1073/pnas.1402594111
- Chwilla, D. J., and Kolk, H. H. J. (2002). Three-step priming in lexical decision. *Mem. Cognit.* 30, 217–225. doi: 10.3758/BF03195282
- Coane, J. H., and Balota, D. A. (2011). Face (and nose) priming for book: the malleability of semantic memory. *Exp. Psychol.* 58, 62–70. doi: 10.1027/1618-3169/a000068
- Coane, J. H., Huff, M. J., and Hutchison, K. A. (2016). The ironic effect of guessing: Increased false memory for mediated lists in younger and older adults. *Aging Neuropsychol. Cognit.* 23, 282–303. doi: 10.1080/13825585.2015.1088506
- Coane, J. H., McBride, D. M., Cam, Y., Chang, K., and Marsh, E. (in preparation). *Reversal of list Type Effects in False Memories in Short- and Long-Term Tests*.
- Coane, J. H., and McBride, D. M. (2006). The role of test structure in creating false memories. *Mem. Cognit.* 34, 1026–1036. doi: 10.3758/BF03193249
- Coane, J. H., McBride, D. M., Raulerson, B. A. III., and Jordan, J. S. (2007). False memory in a short-term memory task. *J. Exp. Psychol.* 54, 62–70. doi: 10.1027/1618-3169.54.1.62

AUTHOR CONTRIBUTIONS

All authors contributed to the writing and editing of the manuscript.

FUNDING

This work was supported by funding from a James McDonnell Foundation *Understanding Human Cognition* Grant awarded to JC (#220020426). The funding agency had no input on the organization or content of the review or on the writing.

- Coane, J. H., McBride, D. M., and Xu, S. (2020). The feature boost in false memory: the roles of monitoring and critical item identifiability. *Memory* 28, 481–493. doi: 10.1080/09658211.2020.1735445
- Coane, J. H., Yu, Z., Xu, S., Walsh, C., Feng, Y., and Gray, S. (manuscript in preparation). False memory for associatively and categorically related lists in younger and older adults.
- Colbert, J. M., and McBride, D. M. (2007). Comparing decay rates for accurate and false memories in the DRM paradigm. *Mem. Cognit.* 35, 1600–1609. doi: 10.3758/BF03193494
- Collins, A. M., and Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychol. Rev.* 100, 589–608. doi: 10.1037/0033-295X.100.4.589
- Coltheart, M., Jonasson, J. T., Davelaar, E., and Besner, D. (1977). “Access to the internal lexicon,” in *Attention and Performance VI*, ed S. Dornic (Hillsdale, NJ: Academic Press), 535–555.
- Conrad, R. (1963). Acoustic confusions and memory span for words. *Nature* 197, 1029–1030. doi: 10.1038/1971029a0
- Cortese, M. J., Khanna, M. M., White, K., Veljkovic, I., and Drumm, G. (2008). The activation and monitoring of memories produced by words and pseudohomophones. *J. Mem. Lang.* 58, 393–414. doi: 10.1016/j.jml.2007.11.003
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Dewhurst, S. A., Barry, C., Swannell, E. R., Holmes, S. J., and Barhurst, G. L. (2007). The effect of divided attention on false memory depends on how memory is tested. *Mem. Cognit.* 35, 660–667. doi: 10.3758/BF03193304
- Díez, E., Gómez-Ariza, C. J., Díez-Álamo, A. M., Alonso, M. A., and Fernandez, A. (2017). The processing of semantic relatedness in the brain: Evidence from associative and categorical false recognition effects following transcranial direct current stimulation of the left anterior temporal lobe. *Cortex* 93, 133–145. doi: 10.1016/j.cortex.2017.05.004
- Dodd, M. D., and Macleod, C. M. (2004). False recognition without intentional learning. *Psycho. Bull. Rev.* 11, 137–142. doi: 10.3758/BF03206473
- Ebbinghaus, H. (1885/1913). *Memory: A Contribution to Experimental Psychology*, eds H. A. Ruger and C. E. Bussenius (New York, NY: Teachers College). doi: 10.1037/10011-000
- Farrell, M. T., Abrams, L., and White, K. K. (2012). “The role of priming in lexical access and speech production,” in *Perspectives on Cognitive Psychology: Psychology of Priming*, eds N. Hsu and Z. Schütt (New York, NY: Nova Science Publishers), 205–244.
- Finley, J. R., Sungkhasettee, V. W., Roediger, H. L. III., and Balota, D. A. (2017). Relative contributions of semantic and phonological associates to over-additive false recall in hybrid DRM lists. *J. Mem. Lang.* 93, 154–168. doi: 10.1016/j.jml.2016.07.006
- Flegel, K. E., Atkins, A. S., and Reuter-Lorenz, P. A. (2010). False memories seconds later: the rapid and compelling onset of illusory recognition. *J. Exp. Psychol. Learn. Mem. Cognit.* 36, 1331–1338. doi: 10.1037/a0019903
- Gallo, D. A. (2006). *Associative Illusions of Memory: False Memory Research in DRM and Related Tasks*. New York, NY: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Mem. Cognit.* 38, 833–848. doi: 10.3758/MC.38.7.833
- Gallo, D. A., Roberts, M. J., and Seamon, J. G. (1997). Remembering words not presented in lists: can we avoid creating false memories? *Psycho. Bull. Rev.* 4, 271–276. doi: 10.3758/BF03209405
- Geraci, L., and McCabe, D. P. (2006). Examining the basis for illusory recollection: the role of remember/know instructions. *Psycho. Bull. Rev.* 13, 466–473. doi: 10.3758/BF03193871
- Greene, R. L. (2016). The role of stimulus structure in semantic memory. *Psychol. Learn. Motiv.* 64, 133–159. doi: 10.1016/bs.plm.2015.09.004
- Guo, X. Y., Zhou, C., and Zhou, M. H. (2004). A study on the factors influencing false memory. *Chin. J. Appl. Psychol.* 10, 3–8. Available online at: https://caod.oriprobe.com/articles/6921961/A_Study_on_the_Factors_Influencing_False_Memory.htm
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428. doi: 10.1037/0033-295X.93.4.411
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol. Rev.* 95, 528–551. doi: 10.1037/0033-295X.95.4.528
- Horoiță, A., and Opre, A. (2020). False memories: Romanian Deese-Roediger-McDermott lists of words. *Cognit. Brain Behav. Interdiscipl. J.* 24, 163–186. doi: 10.24193/cbb.2020.24.10
- Huff, M. J., and Aschenbrenner, A. J. (2018). Item-specific processing reduces false recognition in older and younger adults: separating encoding and retrieval using signal detection and the diffusion model. *Mem. Cognit.* 46, 1287–1301. doi: 10.3758/s13421-018-0837-1
- Huff, M. J., Bodner, G. E., and Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: a meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psycho. Bull. Rev.* 22, 349–365. doi: 10.3758/s13423-014-0648-8
- Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., and Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: effects of initial recall, recognition, and the ironic effect of guessing. *J. Exp. Psychol. Learn. Mem. Cognit.* 38, 1720–1730. doi: 10.1037/a0028476
- Huff, M. J., Di Mauro, A., Coane, J. H., and O'Brien, L. M. (2021). Mapping the time course of semantic activation in mediated false memory: immediate classification, naming, and recognition. *Q. J. Exp. Psychol.* 74, 483–469. doi: 10.1177/1747021820965061
- Huff, M. J., and Hutchison, K. A. (2011). The effect of mediated word lists on false recall and recognition. *Mem. Cognit.* 39, 941–953. doi: 10.3758/s13421-011-0077-0
- Huff, M. J., Meade, M. L., and Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory* 19, 317–330. doi: 10.1080/09658211.2011.568494
- Huff, M. J., Yates, T. J., and Balota, D. A. (2018). Evaluating the contributions of task expectancy in the testing and guessing benefits on recognition memory. *Memory* 26, 1065–1083. doi: 10.1080/09658211.2018.1467929
- Humphreys, G. W., Evett, L. J., Quinlan, P. T., and Besner, D. (2016). “Orthographic priming: Qualitative differences between priming from identified and unidentified primes,” in *Attention and Performance XII: The Psychology of Reading*, ed M. Coltheart (London: Lawrence Erlbaum Associates, Inc), 105–126.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psycho. Bull. Rev.* 10, 785–813. doi: 10.3758/BF03196544
- Hutchison, K. A., and Balota, D. A. (2005). Decoupling semantic and associative information in false memories: explorations with semantically ambiguous and unambiguous critical lures. *J. Mem. Lang.* 52, 1–28. doi: 10.1016/j.jml.2004.08.003
- Hutchison, K. A., Meade, M. L., Williams, N. S., Manley, K. D., and McNabb, J. C. (2018). How do associative and phonemic overlap interact to boost illusory recollection? *Memory* 26, 664–671. doi: 10.1080/09658211.2017.1393091
- Ichikawa, H., Kanazawa, S., and Yamaguchi, M. K. (2011). Finding a face in a face-like object. *Perception* 40, 500–502. doi: 10.1068/p6926
- Israel, L., and Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psycho. Bull. Rev.* 4, 557–581. doi: 10.3758/BF03214352
- Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3
- Kawasaki, Y., and Yama, H. (2006). The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory* 14, 68–78. doi: 10.1080/09658210444000520
- Knott, L. M., and Dewhurst, S. A. (2007). The effects of divided attention at study and test on false recognition: a comparison of DRM and categorized lists. *Mem. Cognit.* 35, 1954–1965. doi: 10.3758/BF03192928
- Knott, L. M., Dewhurst, S. A., and Howe, M. L. (2012). What factors underlie associative and categorical memory illusions? The roles of backward associative strength and interitem connectivity. *J. Exp. Psychol. Learn. Mem. Cognit.* 38, 229–239. doi: 10.1037/a0025201
- Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., and Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: testing the semantic categorization account. *J. Exp. Psychol. Learn. Mem. Cognit.* 29, 499–510. doi: 10.1037/0278-7393.29.4.499

- Koutstaal, W., and Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *J. Mem. Lang.* 37, 555–583. doi: 10.1006/jmla.1997.2529
- Koutstaal, W., Schacter, D. L., and Brenner, C. (2001). Dual task demands and gist-based false recognition of pictures in younger and older adults. *J. Mem. Lang.* 44, 399–426. doi: 10.1006/jmla.2000.2734
- Koutstaal, W., Schacter, D. L., Brenner, C., and Jackson, E. M. (1999). Perceptually based false recognition of novel objects in amnesia: effects of category size and similarity to category. *Cognit. Neuropsychol.* 16, 317–341. doi: 10.1080/026432999380816
- Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psycho. Bull. Rev.* 28, 40–80. doi: 10.3758/s13423-020-01792-x
- Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philos. Trans. R. Soc. B* 369:20120392. doi: 10.1098/rstb.2012.0392
- Lambon Ralph, M. A., and Patterson, K. (2008). Generalization and differentiation in semantic memory: insights from semantic dementia. *Ann. N. Y. Acad. Sci.* 1124, 61–76. doi: 10.1196/annals.1440.006
- Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240. doi: 10.1037/0033-295X.104.2.211
- Lin, Z. X., and Han, B. X. (1999). The orthographic, phonetic and semantic coding of Chinese words. *Chin. J. Psychol. Sci.* 22, 1–4. doi: 10.16719/j.cnki.1671-6981.1999.01.001
- Liu, A. L., and Cao, S. Y. (2002). Research on false memory to different age and visual channel. *Chin. J. Appl. Psychol.* 8, 28–34. Available online at: https://caod.orioprobe.com/articles/4424583/Research_on_False_Memory_to_Different_Age_and_Visu.htm
- Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn. Mem.* 12, 361–366. doi: 10.1101/lm.94705
- Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psycho. Bull. Rev.* 7, 618–630. doi: 10.3758/BF03212999
- Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001
- Lukatela, G., and Turvey, M. T. (1991). Phonological access of the lexicon: evidence from associative priming with pseudohomophones. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 951–966. doi: 10.1037/0096-1523.17.4.951
- Marslen-Wilson, W., and Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition* 8, 1–71. doi: 10.1016/0010-0277(80)90015-3
- Martin, A. (2007). The representation of object concepts in the brain. *Ann. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143
- McBride, D. M., Coane, J. H., Xu, S., Feng, Y., and Yu, Z. (2019). Short-term false memories vary as a function of list type. *Q. J. Exp. Psychol.* 72, 2726–2741. doi: 10.1177/1747021819859880
- McCabe, D. P., and Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Mem. Cognit.* 30, 1065–1077. doi: 10.3758/BF03194324
- McDermott, K. B., and Watson, J. M. (2001). The rise and fall of false recall: the impact of presentation duration. *J. Mem. Lang.* 45, 160–176. doi: 10.1006/jmla.2000.2771
- McGeoch, J. A. (1942). *The Psychology of Human Learning*. Longmans: B Green and Co., Inc.
- Meade, M. L., Watson, J. M., Balota, D. A., and Roediger, H. L. III. (2007). The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm. *J. Mem. Lang.* 56, 305–320. doi: 10.1016/j.jml.2006.07.007
- Mirman, D., Landrigan, J. F., and Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychol. Bull.* 143, 499–520. doi: 10.1037/bul0000092
- Montefinese, M., Zannino, G. D., and Ambrosini, E. (2015). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychol. Res.* 79, 785–794. doi: 10.1007/s00426-014-0615-z
- Nairne, J. S. (1990). A feature model of immediate memory. *Mem. Cognit.* 18, 251–269. doi: 10.3758/BF03213879
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–254. doi: 10.1037/0096-3445.106.3.226
- Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behav. Res. Methods Instrum. Comput.* 36, 402–407. doi: 10.3758/BF03195588
- Neuschatz, J. S., Benoit, G. E., and Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: The role of identifiability. *J. Exp. Psychol. Learn. Mem. Cognit.* 29, 35–41. doi: 10.1037/0278-7393.29.1.35
- Nieznanski, M., Obidziński, M., Niedziałkowska, D., and Zyskowska, E. (2019). False memory for orthographically related words: research in the simplified conjoint recognition Paradigm. *Am. J. Psychol.* 132, 57–69. doi: 10.5406/amerjpsyc.132.1.0057
- Oliver, M. C., Bays, R. B., and Zabrocky, K. M. (2016). False memories and the DRM paradigm: effects of imagery, list, and test type. *J. Gen. Psychol.* 143, 33–48. doi: 10.1080/00221309.2015.1110558
- Pansuwan, T., Breuer, F., Gazder, T., Lau, Z., Cueva, S., Swanson, L. et al. (2020). Evidence for adult age-invariance in associative false recognition. *Memory* 28, 172–186. doi: 10.1080/09658211.2019.1705351
- Park, L., Shobe, K. K., and Kihlstrom, J. F. (2005). Associative and categorical relations in the associative memory illusion. *Psychol. Sci.* 16, 792–797. doi: 10.1111/j.1467-9280.2005.01616.x
- Payne, D. G., Elie, C. J., Blackwell, J. M., and Neuschatz, J. S. (1996). Memory illusions: recalling, recognizing, and recollecting events that never occurred. *J. Mem. Lang.* 35, 261–285. doi: 10.1006/jmla.1996.0015
- Peters, M. J. V., Jellic, M., Gorski, B., Sijstermans, K., Giesbrecht, T., and Merckelback, H. (2008). The corrective effects of warning on false memories in the DRM paradigm are limited to full attention conditions. *Acta Psychol.* 129, 308–314. doi: 10.1016/j.actpsy.2008.08.007
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115. doi: 10.1037/0033-295X.103.1.56
- Qu, Z., and Ding, Y. L. (2010). The effect of Chinese phonological association on false memory. *Acta Psychol. Sin.* 42, 193–199. doi: 10.3724/SP.J.1041.2010.00193
- Qu, Z., Liu, Y., and Bi, Y. H. (2010). The effect of Chinese orthographic association on false memory. *Chin. J. Appl. Psychol.* 16, 146–153.
- Rastle, K. (2007). “Visual word recognition,” in *The Oxford Handbook of Psycholinguistics*, ed M. G. Gaskell (Oxford: Oxford University Press), 71–87. doi: 10.1093/oxfordhb/9780198568971.013.0005
- Reyna, V. F. (2012). A new intuitionism: meaning, memory, and development in fuzzy-trace theory. *Judg. Deci. Mak.* 7, 332–359. Available online at: <http://journal.sjdm.org/11/111031/jdm111031.pdf>
- Robinson, K. J., and Roediger, H. L. III. (1997). Associative processes in false recall and false recognition. *Psychol. Sci.* 8, 231–237. doi: 10.1111/j.1467-9280.1997.tb00417.x
- Roediger, H. L., and DeSoto, K. A. (2015). “The psychology of reconstructive memory,” in *International Encyclopedia of the Social and Behavioral Sciences*, ed J. Wright (Washington, DC: Elsevier).
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cognit.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, H. L. III, Balota, D. A., and Watson, J. M. (2001). “Spreading activation and arousal of false memories,” in *Science Conference Series. The Nature of Remembering: Essays in Honor of Robert G. Crowder*, eds H. L. Roediger III, J. S. Nairne, I. Neath, and A. M. Surprenant (Oxford: American Psychological Association), 95–115. doi: 10.1037/10394-006
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psycho. Rev.* 111, 205–235. doi: 10.1037/0033-295X.111.1.205
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Percept. Psychophys.* 2, 437–442. doi: 10.3758/BF03208784
- Schacter, D. L. (2001). *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Houghton, Mifflin and Company.
- Schacter, D. L., Koutstaal, W., and Norman, K. A. (1997). False memories and aging. *Trends Cognit. Sci.* 1, 229–236. doi: 10.1016/S1364-6613(97)01068-1
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., et al. (2002). Are false memories more difficult to forget than accurate

- memories? The effect of retention interval on recall and recognition. *Mem. Cognit.* 30, 1054–1064. doi: 10.3758/BF03194323
- Seamon, J. G., Luo, C. R., Schlegel, S. E., Greene, S. E., and Goldenberg, A. B. (2000). False memory for categorized pictures and words: the category associates procedure for studying memory errors in children and adults. *J. Mem. Lang.* 42, 120–146. doi: 10.1006/jmla.1999.2676
- Senese, V. P., Sergi, I., and Iachini, T. (2010). Comparison of activation level between true and false items in the DRM paradigm. *Cognit. Process.* 11, 213–217. doi: 10.1007/s10339-009-0271-1
- Shimane, D., Matsui, H., and Itoh, Y. (2020). False memory for faces is produced by the Deese-Roediger-McDermott paradigm based on the morphological characteristics of a list. *Curr. Psychol.* Available online at: <https://link.springer.com/article/10.1007/s12144-020-00830-0#citeas>. doi: 10.1007/s12144-020-00830-0
- Sikora-Wachowicz, B., Lewandowska, K., Keresztes, A., Werkle-Bergner, M., Marek, T., and Fafrowicz, M. (2019). False recognition in short-term memory – age-differences in confidence. *Front. Psychol.* 10:2785. doi: 10.3389/fpsyg.2019.02785
- Smith, S. M., Gerkens, D. R., Pierce, B. H., and Choi, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: the Kirkpatrick and Deese hypotheses. *J. Mem. Lang.* 47, 436–447. doi: 10.1016/S0749-596X(02)00012-8
- Sommers, M. S., and Lewis, B. P. (1999). Who really lives next door: creating false memories with phonological neighbors. *J. Mem. Lang.* 40, 83–108. doi: 10.1006/jmla.1998.2614
- Stadler, M. A., Roediger, H. L. III, and McDermott, K. B. (1999). Norms for word lists that create false memories. *Mem. Cognit.* 27, 494–500. doi: 10.3758/BF03211543
- Sternberg, S. (1966). High-speed scanning in human memory. *Science* 153, 652–654. doi: 10.1126/science.153.3736.652
- Steyvers, M., and Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognit. Sci.* 29, 41–78. doi: 10.1207/s15516709cog2901_3
- Toglia, M. P., Neuschatz, J. S., and Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory* 7, 233–256. doi: 10.1080/741944069
- Tse, C. S., Li, Y., and Neill, W. T. (2011). Dissociative effects of phonological vs. semantic associates on recognition memory in the Deese/Roediger-McDermott paradigm. *Acta Psychol.* 137, 269–279. doi: 10.1016/j.actpsy.2011.01.013
- Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Clarendon Press.
- Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12. doi: 10.1037/h0080017
- Tulving, E., and Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *J. Verb. Learn. Verb. Behav.* 5, 381–391. doi: 10.1016/S0022-5371(66)80048-8
- Wang, J., Otgaar, H., Howe, M. L., Lippe, F., and Smeets, T. (2018). The nature and consequences of false memories for visual stimuli. *J. Mem. Lang.* 101, 124–135. doi: 10.1016/j.jml.2018.04.007
- Watson, J. M., Balota, D. A., and Roediger, H. L. III. (2003). Creating false memories with hybrid lists of semantic and phonological associates: over-additive false memories produced by converging associative networks. *J. Mem. Lang.* 49, 95–118. doi: 10.1016/S0749-596X(03)00019-6
- Watson, J. M., Bunting, M. F., Poole, B. J., and Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese–Roediger–McDermott paradigm. *J. Exp. Psychol. Learn. Mem. Cognit.* 31, 76–85. doi: 10.1037/0278-7393.31.1.76
- Westbury, C., Buchanan, L., and Brown, N. R. (2002). Sounds of the neighborhood: False memories and the structure of the phonological lexicon. *J. Mem. Lang.* 46, 622–651. doi: 10.1006/jmla.2001.2821
- Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *Am. J. Psychol.* 78, 567–574. doi: 10.2307/1420917
- Wickens, D. D. (1970). Encoding categories of words: an empirical approach to meaning. *Psychol. Rev.* 77, 1–15. doi: 10.1037/h0028569
- Wu, L. L., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol.* 132, 173–189. doi: 10.1016/j.actpsy.2009.02.002
- Xu, S., Yu, Z., McBride, D. M., and Coane, J. H. (2017). “Short-term false memories vary as a function of list type,” in *58th Annual Meeting of the Psychonomic Society* (Vancouver, BC).
- Zeelenberg, R., Boot, I., and Pecher, D. (2005). Activating the critical lure during study is unnecessary for false recognition. *Consci. Cognit.* 14, 316–326. doi: 10.1016/j.concog.2004.08.004
- Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., et al. (2018). Participant nonnaïveté and the reproducibility of cognitive psychology. *Psycho. Bull. Rev.* 25, 1968–1972. doi: 10.3758/s13423-017-1348-y

Conflict of Interest: 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 © 2021 Coane, McBride, Huff, Chang, Marsh and Smith. 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) and the copyright owner(s) 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.



Does Alexithymia Affect Memory for a Crime? The Relationship Between Alexithymia, Executive Functions, and Memories

Fabiana Battista^{1,2*†}, Tiziana Lanciano^{1†} and Antonietta Curci^{1†}

¹ Department of Education, Psychology, Communication, University of Bari "Aldo Moro", Bari, Italy, ² Leuven Institute of Criminology, Catholic University of Leuven, Leuven, Belgium

OPEN ACCESS

Edited by:

Beatriz Martín-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

Gianluca Serafini,
San Martino Hospital (IRCCS), Italy
Hein De Haan,
Tactus Verslavingszorg, Netherlands
Clare Margaret Eddy,
Birmingham and Solihull Mental
Health NHS Foundation Trust,
United Kingdom

*Correspondence:

Fabiana Battista
fabiana.battista@uniba.it

†ORCID:

Fabiana Battista
orcid.org/0000-0003-4086-739X
Tiziana Lanciano
orcid.org/0000-0001-6533-1052
Antonietta Curci
orcid.org/0000-0002-0932-7152

Specialty section:

This article was submitted to
Forensic and Legal Psychology,
a section of the journal
Frontiers in Psychology

Received: 19 February 2021

Accepted: 19 May 2021

Published: 30 June 2021

Citation:

Battista F, Lanciano T and Curci A
(2021) Does Alexithymia Affect
Memory for a Crime? The Relationship
Between Alexithymia, Executive
Functions, and Memories.
Front. Psychol. 12:669778.
doi: 10.3389/fpsyg.2021.669778

Prior studies on alexithymia and memory have found a negative association between the two constructs, especially when emotional memories are considered. Moreover, it has been demonstrated that also the executive functioning (EF) of the individuals influences this relationship. Thus, the goal of this study is to verify whether alexithymia can influence the memory accuracy for a violent crime in people with different levels of EF resources in terms of both correct details and memory distortions (i.e., omissions and commissions) reported. We assessed the alexithymia and EF resources of individuals and showed participants a video of a violent crime (i.e., murder). We then asked participants to testify about the content of the video by imagining to be witnesses of the crime. A memory test was run on two moments in time: immediately after the video presentation and after 10 days. Findings demonstrated that alexithymia influences the recall of the event both in proneness to report correct details and memory distortions of the participants (i.e., omissions and commissions). Additionally, we found a contribution of EF resources in this relationship. The findings provide new information for legal professionals on memory functioning.

Keywords: eyewitness, alexythimia, memory distortions, forgetting, executive functions

INTRODUCTION

Alexithymia has been defined as a personality construct specifically characterized by difficulties in identifying feelings (DIFs), difficulties in describing feelings (DDFs), and externally oriented thinking (EOT) (e.g., Sifneos, 1973; Nemiah, 1977; Bagby et al., 1994a,b; Taylor, 2000; Kooiman et al., 2002; Vermeulen and Luminet, 2009). In the 1960's, alexithymia was conceived as a personality disorder (e.g., Kooiman et al., 2002) due to a strong link to psychosomatic dysfunction, such as anxiety, depression, posttraumatic stress disorder, abuse of alcohol and drugs, and eating disorders (e.g., Zeitlan and McNally, 1993; Petterson, 2004; Li et al., 2015; Brady et al., 2016). Some studies have also found that alexithymia might mediate the relationship between maltreatment during childhood and psychopathologies (e.g., Paivio and McCulloch, 2004; Serafini et al., 2017). However, many recent studies have shown that alexithymia generally occurs also in individuals with no declared diseases (e.g., Franz et al., 2008; Donges and Suslow, 2015); thus, it is to be considered as a personality variable rather than a disorder (e.g., Donges and Suslow, 2015). In other words, alexithymia is a personality characteristic that is present in all people to a different degree and involving not only feelings and the ability to describe them, but also the orientation of thinking of individuals (e.g., external, internal; Taylor, 1984; da Silva et al., 2017).

Indeed, studies so far have shown that individuals with a high level of alexithymia have problems in emotional responses both in terms of emotional regulation and cognitive processing (e.g., Lane et al., 1996; Taylor et al., 1997; Swart et al., 2009). Individuals with a high level of alexithymia have difficulty (i) in perceiving, identifying their feelings and emotions and, in turn, describing them to others; and (ii) incorrectly using emotions either to lead their actions and thinking or to learn from emotional experience (e.g., Luminet et al., 2004, 2006). Recent studies have explained that such difficulties may be due to problems in sensory perception processing (e.g., Serafini et al., 2017). Scholars showed that the inability of alexithymics to identify and describe feelings is associated with a lower response to sensory stimulation and inputs (e.g., Grynberg and Pollatos, 2015; Serafini et al., 2016). Moreover, other studies collected further evidence showing that the lower capacity to recognize emotions can be related to a lower ability in attentional processing (e.g., Vermeulen et al., 2006; Borhani et al., 2016; Correro et al., 2019). In particular, these studies have shown a difficulty of high alexithymia people in the early automatic attentional processing that does not permit to correctly employ their attentional resources for emotional and relevant information in the context (e.g., Aftanas et al., 2003; Swart et al., 2009; Mather and Sutherland, 2011; van der Velde et al., 2015; Nielson and Correro, 2017). Furthermore, a plethora of electrophysiological studies have underlined that the deficit in processing emotional information might be caused by an alteration in the activation of the brain regions (i.e., amygdala, dorsal anterior, and middle cingulate cortex) normally employed by the emotional attention system (e.g., Wager et al., 2008; Diekhof et al., 2011; Pollatos and Gramann, 2011; Van der Velde et al., 2013). Collectively, these difficulties seem to affect, in turn, the memory for emotional information in people with high alexithymia level (Jacob and Hautekeete, 1998; Suslow et al., 2003; Luminet et al., 2006; Vermeulen and Luminet, 2009; Meltzer and Nielson, 2010; Vermeulen et al., 2010, 2018; Donges and Suslow, 2015; Dressaire et al., 2015; Correro et al., 2019; Ridout et al., 2020).

Studies investigating the relationship between alexithymia and memory are increasing in the last years (Jacob and Hautekeete, 1998; Lundh et al., 2002; Luminet et al., 2006; Vermeulen and Luminet, 2009; Correro et al., 2019; Ridout et al., 2020) and are particularly relevant in light of the practical implications that can have, for instance, in the legal arena. Indeed, these studies provided further information to understand whether alexithymia can be one of the several factors that lead to memory distortions and thus make the statements of the people unreliable. However, the results of these studies are contrasting. On the one hand, a few studies have demonstrated that alexithymia does not affect the retrieval of information (e.g., Jacob and Hautekeete, 1998; Lundh et al., 2002). On the other hand, a large number of studies have found a negative impact of alexithymia on memory (e.g., Luminet et al., 2006; Vermeulen and Luminet, 2009; Correro et al., 2019; Ridout et al., 2020). For example, a study by Luminet et al. (2006) showed that people with high alexithymia have reduced recall of emotional information. Specifically, the authors tested the alexithymia of the participants by adopting the typical scale employed for assessing the construct, i.e., the

Toronto Alexithymia Scale-20 (TAS-20; Bagby et al., 1994a) consisting of three subscales corresponding to the main features of the construct: (i) DIFs, (ii) DDFs, and (iii) EOT. Then, participants watched a list of emotional (i.e., negative and positive) and neutral words with different instructions to process the information (i.e., perceptual or semantic processing). Finally, the memory of the participants was tested with a remember/know task. Results indicated that high alexithymia participants had a worse recall for the emotional words than low alexithymia participants with regard to the processing adopted during the first phase (i.e., perceptual or semantic) of the study and the type of memory measure (i.e., “remember” or “know”).

In the following study, by adopting a similar procedure, Vermeulen and Luminet (2009) replicated the abovementioned findings. However, the authors added that the impairment in the recollection of the positive and negative words depends on specific subscales of the TAS-20. Indeed, they found that the DIF subscale was the most involved in the lower recall of emotional words in highly alexithymic individuals, while both the DIF and the other subscale EOT were related to the detrimental recall of the positive and negative words in terms of “remember” response.

Further evidence has been provided from another study by Vermeulen et al. (2010) showing that the memory impairment for emotional words in highly alexithymic individuals also depends on the congruency of the emotional valence between the context during the encoding and the studied information. The authors found that the subscales DIF and DDF are associated with the effects on memory when there is congruency between the context and information, while DIF and EOT are related to the recall of the studies information when there is no congruency.

Collectively, all the described studies have provided support to the idea that alexithymia negatively affects memory performance. However, all these studies have demonstrated this issue by using verbal stimuli (i.e., lists of words) instead of more complex stimuli (i.e., pictures, videos). Indeed, only a few studies have tried to fill this gap by adopting pictures of emotional faces or scenes of social interaction (e.g., DiStefano and Koven, 2012; Takahashi et al., 2015). Such studies showed that memory undermining in highly alexithymic individuals also occurs for other types of stimuli. DiStefano and Koven (2012) demonstrated that people with a higher level of alexithymia had a worse recall of the social scenes than people with a lower level of alexithymia. Still, Takahashi et al. (2015) found an association between alexithymia and memory for positive faces but not for the negative ones. Based on these contrasting results as compared to the significant association between alexithymia and negative words, two recent studies by Ridout et al. (2020) further investigated the possibility that alexithymia can also affect the correct recall of complex negative information (i.e., faces and video clips). Results revealed that alexithymia also contributes to the recall of negative faces: The higher the DDF score, the lower the correct recall of negative faces. By contrast, evidence from the second study using emotional video clips demonstrated that alexithymia interferes with the recall of the negative video clip information with regard to the other two subscales of DIF and EOT.

Moreover, recent studies have underlined that memory deficits in highly alexithymic individuals are due to reduced cognitive capacities in terms of executive functioning (EF) (Diamond, 2013). Indeed, high alexithymia people have dysfunction in problem-solving, fluency, shifting, inhibition, and self-report (Henry et al., 2006; Wood and Williams, 2007; Onor et al., 2010; Zhang et al., 2012; Santorelli and Ready, 2015). In particular, some studies have shown that the subscales particularly associated with executive dysfunction are the DIF and DDF (Henry et al., 2006; Koven and Thomas, 2010; Santorelli and Ready, 2015), while other evidence supports the idea that the EOT subscale is the most implicated in the effects of alexithymia on EFs (e.g., Correro et al., 2019).

The idea that the influence of alexithymia on memory is moderated by the EFs abilities can also be supported by studies showing that EFs are essential for cognitive tasks and processes, such as memory recall (Carpenter et al., 2000; Miyake et al., 2000; Duff et al., 2005; Carretti et al., 2010; Hill et al., 2012). In the last years, a plethora of studies have demonstrated that EFs are involved in memory accuracy in terms of both correct information and memory distortions (i.e., omission and commission errors) reported during the retrieval (e.g., Gonsalves and Paller, 2002; Jaschinski and Wentura, 2002; Schacter and Slotnick, 2004; Marsh et al., 2005; Gerrie and Garry, 2007; Peters et al., 2007; Zhu et al., 2010; Leding, 2012; Mirandola et al., 2015; Kersten et al., 2018; Battista et al., 2020). Specifically, it has been shown that the abilities of EF of individuals, in terms of working memory as digit span, shifting as reaction times to different mathematical operations, inhibition as the number of correct responses to visual stimuli, and updating as phonemic fluency, impact on the likelihood to report correct details and errors as follows: Low EF people report less correct details and a higher number of omissions and commissions than high EF people. These recurrent findings have been displayed both for simple stimuli (i.e., words) (e.g., Peters et al., 2007) and for complex stimuli (i.e., video) (e.g., Battista et al., 2020) as well as for neutral stimuli (e.g., Gerrie and Garry, 2007) and emotional stimuli (e.g., Mirandola et al., 2015).

Despite the growing number of studies on alexithymia and memory, research on this topic is still quite limited and findings are often in contrast to each other. Moreover, no studies have verified whether alexithymia is also responsible for a higher or lower susceptibility to report memory distortions (i.e., omissions and commissions). With a forensic perspective in mind, studies investigating a possible relationship among EF resources of individuals, alexithymia, and memory for a complex emotional event, such as a crime scene, seem to be particularly necessary. Thus, this study will try to address this lack of evidence by testing whether the recall of a violent crime (i.e., murder) can be influenced by the three scores of alexithymia (i.e., DIF, DDF, and EOT) in a sample of individuals having high or low EF resources. By doing so, in accordance with previous studies on memory and EFs (Kersten et al., 2018; Battista et al., 2020) as well as Correro et al. (2019) on EF, memory, and alexithymia, we followed a comprehensive approach to measure the EF availability of individuals. In particular, based on studies showing that the three functions of Updating, Shifting, and Inhibition work together while performing cognitive tasks (e.g., recall) (e.g.,

Kimberg and Farah, 1993; Duncan Roger Johnson Michaela Swales Charles Freer, 1997; Dempster and Corkill, 1999; Miyake et al., 2000), we collapsed in an aggregate measure the individuals' EFs measures of the three EFs. Specifically, Updating refers to the capacity (i) to monitor and code the incoming and relevant information for the execution of a task and (ii) to change the old and irrelevant information into newer and more relevant information (Morris and Jones, 1990; Lehto, 1996). Shifting is the ability to switch among different tasks, operations, or mental sets (Monsell, 1996). Inhibition refers to the capacity to intentionally suppress dominant and automatic responses when necessary (Miyake et al., 2000). Note that the three EFs (i.e., Updating, Shifting, and Inhibition) were chosen also in accordance with recent studies underlining that they are the most implicated in superior cognitive performance, in particular, memory recall (e.g., Espy, 2004; Burgess and Simons, 2005; Diamond, 2013; Kersten et al., 2018; Battista et al., 2020, 2021).

THE PRESENT STUDY

Aims and Hypotheses

The current study aimed to verify whether the degree of alexithymia of individuals can influence memory accuracy for a crime experience. More specifically, this study is intended to verify whether the likelihood to recall correct details, memory distortions (i.e., omissions and commissions) for an emotional event is related to the ability to discriminate and describe feelings and the tendency to have EOT. We tested this issue by focusing on memory recall immediately after a target experience and after 10 days. Based on previous evidence showing that alexithymia is related to EF (e.g., Correro et al., 2019) and on studies demonstrating the positive association between EFs and memory (e.g., Peters et al., 2007; Zhu et al., 2010; Leding, 2012; Mirandola et al., 2015; Kersten et al., 2018; Battista et al., 2020), we preliminarily assessed the degree of EF resources of participants (i.e., low vs. high) and collapsed the indices of the three EFs of Updating, Shifting, and Inhibition into an aggregate measure. Then, participants filled in some questionnaires measuring their affective state and degree of alexithymia. Immediately after, they watched a mock crime video and answered some questions about it. Ten days later, in a second session, the memory of the participants was assessed for the second occasion. Immediately after, EF resources of the participants were evaluated through three neuropsychological tasks.

In line with previous studies (e.g., Correro et al., 2019), we expected that higher alexithymia scores would be predictive of a lower memory recall, in terms of the amount of correct details (Hypothesis 1). Moreover, in accordance with studies showing that the likelihood to report correct details goes hand in hand with the likelihood to report memory distortions (e.g., Battista et al., 2020), higher alexithymia scores would also be predictive of higher memory distortions (i.e., omissions and commissions) (Hypothesis 2). Finally, based on studies showing that EF influences the effects of alexithymia on memory (e.g., Henry et al., 2006; Koven and Thomas, 2010; Santorelli and Ready, 2015; Correro et al., 2019), we further expected that the negative relationship between alexithymia and memory would be moderated by EF scores of individuals. That is, such a

relationship would be stronger in individuals with lower EF resources than in those with higher EF resources (Hypothesis 3).

METHOD

Participants and Design

Using G*power (Faul et al., 2007), on a priori power analysis for regression analysis with a power of 0.80, a medium effect size ($f = 0.15$), and four predictors (i.e., TAS-DIF, TAS-DDF, TAS-EOT, and EF) indicated that a sample of 85 participants was required. A sample of 110 undergraduate students (women = 86.4%, $M_{\text{age}} = 21.47$, $SD = 3.11$) was thus recruited at the Department of Education, Psychology, Communication of the University of Bari “Aldo Moro.” No specific inclusion and exclusion criteria have been adopted. To achieve the main goal, we created aggregate scores of EF resources (as shown in the section on Results). Therefore, the study was a correlational study with the three scores of alexithymia (i.e., TAS-DIF, TAS-DDF, and TAS-EOT) and EF resources of individuals as predicted variables. The predicted variables were memory scores in terms of correct details and memory distortions (i.e., omissions and commissions scores) in cued recall tests performed after the event (T1) and 10 days later (T2). The ethical committee of the Department of Education, Psychology, Communication of the University of Bari “Aldo Moro” approved the study (No. ET-19-11). Participants did not receive any compensation for taking part in the study and were tested individually in a laboratory. All materials of the study are available on Open Science Framework (OSF): <https://osf.io/bwfek/>.

Measures and Procedure

This study consisted of two different sessions. In the first session, participants answered some questionnaires, then watched a video, and answered a cued recall task about the video. The second session was performed after 10 days and required participants to perform the same cued recall and three neuropsychological tasks useful to assess their EF resources. Before starting the experiment, each participant filled in an informed consent form.

Session 1

Participants were first invited to complete the TAS-20 (Bagby et al., 1994a,b) and the Positive and Negative Affect Schedule-State (PANAS-S; Watson et al., 1988; Terraciano et al., 2003). Thus, participants watched the crime video and finally filled in again the PANAS-S questionnaire to check whether the video emotionally impacted upon them.

Toronto Alexithymia Scale-20 (Bagby et al., 1994a; Bressi et al., 1996). The TAS-20 consists of 20 5-point items (1 = strongly disagree, 5 = strongly agree). The questionnaire evaluates the level of alexithymia of the participant through three scales: DIF, DDF, and EOT. Specifically, the DIF scale (composed of seven items) measures (Cronbach's α in the present study = 0.85) the proneness of an individual to recognize feelings and emotions and to differentiate them from somatic sensations. The DDF scale (Cronbach's α in the present study = 0.79) assesses the ability of an individual to describe feelings

to other people through five items. Finally, the EOT scale (Cronbach's α in the present study = 0.72) consists of eight items measuring the degree of external orientation in the thinking of an individual. Following the scoring system (Bagby et al., 1994a; Caretti et al., 2011), the three subscores are calculated by summing up the answers reported by each participant to the corresponding items of the three scales. Before adding up scores, items 4, 5, 10, 18, and 19 were reversed. The higher the scores, the higher the level of the alexithymia feature of the participant. The score reported by the participants at all the 20 items have been calculated ($M = 41.3$, $SD = 11.8$, range 22–77). The total score can vary from 20 to 100 with a cutoff of 61. In this sample, only four people reported a total score higher than 61.

The PANAS-S (Watson et al., 1988; Terraciano et al., 2003). The questionnaire assesses the emotional state of participants across two scales: The Positive Affect (PA) and the Negative Affect (NA), composed, in turn, of 10 items. PANAS-S includes 20 5-point items (0 = not at all, 4 = completely). The PA-S (Cronbach's α in the present study = 0.88) and NA-S (Cronbach's α in the present study = 0.89) evaluate the positive and negative affective and emotional state while filling it. The scores of the two dimensions are summed up in accordance with the scoring system (Terraciano et al., 2003). The higher the scores, the higher the affective and emotional state of the participants. Thus, the higher the PA score, the higher the positive emotional state; in addition, the higher the NA score, the higher the negative emotional state of participants.

Mock Crime Video

After the completion of the battery of questionnaires, participants were invited to carefully watch a mock crime video. The video (lasting ~2 min 30 s) has been used in other memory studies (Mangiulli et al., 2019; Battista et al., 2020) and shows a debate between two men. This discussion continues with a fight between the two men and the murder of one by the other. Having seen the video, participants filled the PANAS-S (Watson et al., 1988; Terraciano et al., 2003).

Memory Test Phase (T1)

Participants completed a memory test by imagining to be an eyewitness of the crime and to provide testimony to a police officer. The test was a cued recall task; hence, participants answered 14 open questions (i.e., *What was the victim wearing?*) based on what they remember having seen during the video and without guessing. The questions asked participants to recall several details of the crime scene in the video (i.e., the weapon of the murder, the wearing of the murderer and victim, etc.) and are reported in **Appendix A (Supplementary Material)**. All the questions were proposed in the same order to all participants. This memory test has been already adopted in prior studies on memory (e.g., Battista et al., 2020).

Session 2

Memory Test Phase (T2)

After 10 days, participants came back to the laboratory and answered the same questions of the cued recall performed at

Session 1. The questions were proposed in the same order of the cued recall administered at T1 to all participants.

Executive Functioning Assessment

After the memory test, participants completed the three neuropsychological tasks assessing their EF resources. The first two were paper and pencil tasks, while the third one was a computerized task of the Psychology Experiment Building Language 2.0 battery (PEBL 2.0; Mueller and Piper, 2014). At the end of the three tasks, participants were thanked and debriefed.

The phonemic fluency (Novelli et al., 1986). The task measures the ability of Updating (Miyake et al., 2000). The experimenter gives a letter and participants have to refer to as many words as possible in 60 s, beginning with such a letter and excluding names of people and cities. The experiment provides three different letters (C, P, and S). The score of this task is the average number of words produced for each letter. The score is computed by excluding repetitions (i.e., words repeated more times) and intrusions (i.e., words not conforming with the instructions).

Plus-minus task (Jersild, 1927; Spector and Biederman, 1976). This task assesses the capacity of Shifting (Miyake et al., 2000). It is composed of three trials of mathematical operations (i.e., additions and subtractions) on three different sets of random numbers. Specifically, in the first trial, participants have to add the value of three to each number, while in the second trial, they have to subtract three from each number. In the final trial, participants have to alternate among additions and subtractions, that is, participants start with adding three to the first number, continue with subtracting three to the second number, and proceed by switching between the two operations. The experimenter records the time to perform the three trials. The score is calculated by subtracting the average time of the first two trials from the time of the final trial.

Go-NoGo task (Bezdjian et al., 2009). The task measures the ability of Inhibition (Miyake et al., 2000). The task is carried out by using a computer, where participants watch a square divided, in turn, into four squares where the letter P or R appears in a random fashion. Participants are instructed to respond to some stimuli (Go stimuli) and give no response to others (NoGo stimuli) based on the block they are performing. During the first block, the P-Go block, participants have to press the mouse every time the P appears on the screen and not press when the letter R appears. During the second block, the R-Go block, participants have the opposite instructions: Press the mouse when the letter R appears and not press when the letter P appears on the screen. Both blocks are comprised of 160 stimuli and the letters appear on the screen for 500 ms. The score is the commission errors reported at the two blocks.

Cued Recall Scoring

The scoring of both the cued recall tasks (i.e., T1 and T2) was independently performed by the first author and a student assistant, blind to the experimental design. The scoring followed a scoring system employed in a prior study (e.g., Battista et al., 2020). Thus, the three scores of correct details, omissions, and commissions were computed as follows. With regard to the

correct details score, one point was assigned for each completely correct answer (e.g., “*What was the victim wearing?*” “*He was wearing a black jacket, white skirt, and black shirt*”), half a point was assigned for a partially correct answer (e.g., “*He was wearing a black jacket*”), and zero was given for no answer (e.g., “*I do not remember*”) or a completely wrong answer (e.g., “*He was wearing a green suit*”). Regarding the omissions score, one point was given when no answer was provided (e.g., “*I do not remember*”). Concerning the score of the commission, one point was assigned when new and wrong (i.e., not present in the video) information was reported (e.g., “*He was wearing a green suit*”) and half a point when a partially distorted answer was provided (e.g., “*He was wearing a blue jacket, blue skirt, and black shirt*”). The maximum score obtainable for the three indices was 14. In addition, proportions were calculated by dividing the score obtained by each participant by the maximum score obtainable (i.e., 14). The ICC average measure for the number of correct details at T1 and T2 was good: 0.77 and 0.68 (both $ps < 0.001$), respectively. The ICC average measure of omissions and commissions was also good: 0.79 and 0.78 (both $ps < 0.001$) for T1 and 0.78 and 0.76 (both $ps < 0.001$) for T2.

Data Analysis

First, we calculated an aggregate measure of EF for each participant. To order and compare the resources of individuals for the three EFs based on the performance reported at the tasks, we first calculated the rank distributions for each score and then calculated the average of the rank scores. We followed this approach in accordance with previous studies on memory and EFs (e.g., Kersten et al., 2018; Battista et al., 2020). Then, to check the emotional impact of the mock crime video on the affective state of the participants, two paired samples *t*-tests were run on the scores of the two subscales of PANAS-S (i.e., PA and NA). Finally, to verify the Hypothesis 1, 2, and 3 and after checking the associations among the memory scores, the three TAS scores, and EF score (i.e., Pearson’s correlation), a series of general linear model (GLM) analyses were run on the memory scores for correct details (Hypotheses 1), omissions, and commissions (Hypothesis 2) both at T1 and T2. In such analyses, we tested the main effects of the alexithymia scores (i.e., TAS-DIS, TAS-DDF, and TAS-EOT) and the EF score as well as their interaction effects. Moreover, when we found a significance for the interaction effects, we carried out moderation analyses considering the EF score as a moderator (Hypothesis 3). All the analyses have been computed by using *Jamovi* 1.1.8 (Love et al., 2021) that adopts the Preacher and Hayes approach (Preacher and Hayes, 2004).

RESULTS

Analyses on the Aggregate Measure of EF

We created the aggregate measure of EF as follows. First, we checked the rank distributions for each neuropsychological task (i.e., Phonemic Fluency, Plus-Minus Task, and Go/NoGo). Then, we calculated the aggregate score for each participant by averaging the rank score across the three tasks (Table 1).

TABLE 1 | Table shows the mean reported by participants at the three EF (i.e., Shifting, Inhibition, and Updating) tasks and the aggregate score of EFs.

EFs scores	
Shifting	85.98 (32.44) 95% CI [79.85 – 92.11]
Inhibition	309.38 (32.44) 95% CI [396.38 – 312.39]
Updating	14.80 (3.35) 95% CI [14.16 – 14.85]
Aggregate score	19.58 (13.07) 95% CI [17.11 – 22.05]

In particular, the Shifting score refers to the time employed in seconds, the Inhibition score is the average of correct responses, and the Updating score is the average of words reported. Standard deviations and 95% CI are shown between parentheses.

Impact of the Crime Video

In order to assess the emotional impact of the mock crime on participants, a paired samples *t*-test with mock crime (pre vs. post) as a *within-factor* was run on PA-S and NA-S scores. The analysis showed a statistically significant difference on PA-S, $t_{(109)} = 3.62$, $p < 0.001$, $d = 0.35$. In particular, the positive state was significantly lower after the video than before it, $M_{\text{pre-mock crime}} = 25.69$, 95% CI [24.39, 26.99] vs. $M_{\text{post-mock crime}} = 23.58$, 95% CI [22.15, 25.02]. By contrast, no statistically significant difference was found on NA-S, $t_{(109)} = -0.59$, $p = 0.56$, $d = -0.06$. Although no statistical significance was reached, the negative state appeared to increase after the video presentation ($M_{\text{pre-mock crime}} = 6.48$, 95% CI [5.00, 7.97] vs. $M_{\text{post-mock crime}} = 6.77$, 95% CI [5.35, 8.19]).

Moreover, with the aim to verify whether the affective (i.e., positive and negative) states of participants were related to their alexithymia level, Pearson's correlations were carried out among the PA-S and NA-S scores (pre- and post-mock crime) and the three TAS scores (i.e., TAS-DIF, TAS-DDF, and TAS-EOT). We found that TAS-DIF was positively associated with NA-S scores, both pre- and post-mock crime, $r = 0.44$, $p < 0.001$ and $r = 0.36$, $p < 0.001$, respectively. Moreover, TAS-DDF negatively correlated with PA-S scores, both pre- and post-mock crime, $r = -0.24$, $p = 0.01$ and $r = 0.20$, $p = 0.04$, respectively, as well as positively correlated with NA-S scores, both pre- and post-mock crime, $r = 0.38$, $p < 0.001$ and $r = 0.27$, $p = 0.004$, respectively. Finally, TAS-EOT was only negatively associated with the PA-S score pre-mock crime, $r = -0.25$, $p = 0.008$. No other statistically significant correlations were found ($p > 0.05$).

Executive Functioning and Alexithymia

To verify whether the EF resources of individuals are related to alexithymia, Pearson's correlations were run between the three TAS-20 scores (i.e., TAS-DIF, TAS-DDF, and TAS-EOT) and the EF score. Only TAS-DIF correlated significantly with EF, $r = 0.27$, $p = 0.005$. By contrast, no statistically significant correlation was found between TAS-DDT and TAS-EOT with EF, $r = 0.17$, $p = 0.07$. and $r = 0.05$, $p = 0.64$, respectively.

Cued Recall Scores¹

¹We also run exploratory analyses considering the three measures of EFs separately. The analyses showed a statistically significant interaction effect of

TABLE 2 | Table shows Pearson's correlation scores among TAS scores, the EF score and memory scores (i.e., correct details, omissions, and commissions) both at T1 and T2.

	TAS-DIF	TAS-DDF	TAS-EOT	EF score
Correct details T1	0.04	-0.03	-0.10	0.55*
Correct details T2	0.12	-0.08	-0.01	0.04
Omissions T1	-0.06	0.01	0.13	-0.30*
Omissions T2	0.06	0.02	-0.32**	0.03
Commissions T1	0.03	-0.04	0.07	-0.28*
Commissions T2	-0.08	0.04	0.08	-0.04

* $p < 0.01$, ** $p < 0.001$.

Pearson's correlations were run among the three TAS-20 scores (i.e., TAS-DIF, TAS-DDF, and TAS-EOT), the EF score, and memory scores (i.e., correct details, omissions, and commissions at T1 and T2). All the correlation indices are shown in **Table 2**. Regarding alexithymia, only TAS-EOT correlated significantly with omissions score at T2, $r = -0.32$, $p < 0.001$. No other statistically significant correlation was found between alexithymia and memory scores (i.e., correct details, omissions, and commissions at T1 and T2) ($p > 0.05$). Moreover, the EF score was statistically significant correlated with correct details, omissions, and commissions scores at T1, $r = 0.55$, $p = 0.05$, $r = -0.30$, $p = 0.002$, and $r = -0.28$, $p = 0.004$, respectively.

A set of GLM with the three TAS scores (i.e., TAS-DIF, TAS-DDF, and TAS-EOT) and EF scores as predictors was carried out on the cued recall scores of correct details, omissions, and commissions both at T1 and at T2 (i.e., predicted variables). All the direct and the interaction effects are reported in **Table 3**.

Correct Details

With regard to the correct details score, none of the TAS scores predicted the likelihood to report correct details at T1 as well as no interaction effects among the three TAS scores and EF was significant ($p > 0.05$). Overall, the model fit indices indicated an acceptable fit of the model (Burnham and Anderson, 2002), $R^2 = 0.05$, RMSE = 0.067, $F_{(2, 102)} = 0.83$, $p = 0.56$, $\eta^2 = 0.05$.

Regarding the correct details score at T2, all the main effects of TAS scores were not statistically significant. However, the interaction effect of EOT with EF reached the significance level, $\beta = -0.47$, $t = -2.11$, $p = 0.04$. Simple slopes analyses (i.e., moderation analyses) showed that, as long as the EF aggregate score decreases, EOT scores are negatively associated with the proportion of correct details reported at T2, $\beta = -0.004$, $t = -1.65$, $p = 0.04$. No other interaction effects were statistically significant ($p > 0.05$). The model fit indices were $R^2 = 0.07$, RMSE = 0.076, $F_{(2, 102)} = 1.07$, $p = 0.39$, $\eta^2 = 0.07$, showing that the model had an acceptable fit (Burnham and Anderson, 2002).

Shifting*DIF on the amount of correct details and commissions reported at T2. In addition, also the interaction effect of Shifting*EOT on the amount of commissions reported at T2 was statistically significant. Specifically, the results demonstrated that as long as the Shifting score decreases, DIF is positively associated with correct details score and negatively associated with the amount of commissions at T2. Similarly, as long as the Shifting score decreases, EOT is negatively associated with the amount of commissions at T2. These analyses can be found on: <https://osf.io/bwfek/>.

TABLE 3 | Table shows the direct and interaction of effects tested in our General Linear Models on the memory scores.

	β, t, p					
	TAS-DIF	TAS-DDF	TAS_EOT	TAS-DIF*EF	TAS-DDF*EF	TAS-EOT*EF
Correct details T1	0.11, 0.80, 0.42	-0.09, -0.61, 0.54	-0.18, -1.55, 0.12	-0.08, -0.28, 0.78	0.02, 0.07, 0.94	-0.44, -1.93, 0.56
Correct details T2	0.15, 1.06, 0.29	-0.04, -0.26, 0.80	-0.03, -0.27, 0.79	0.16, 0.57, 0.57	0.03, 0.10, 0.92	-0.47, -2.11, 0.04
Omissions T1	-0.09, -0.66, 0.51	0.16, 1.13, 0.26	-1.17, -1.54, 0.13	-0.07, -0.25, 0.81	0.11, 0.39, 0.70	0.15, 0.68, 0.50
Omissions T2	0.03, 0.22, 0.83	0.10, 0.70, 0.49	-0.39, -3.66, <0.001	0.29, 1.49, 0.30	-0.29, -1.00, 0.32	0.16, 0.75, 0.46
Commissions T1	0.01, 0.07, 0.94	-0.11, -0.75, 0.46	0.05, 0.46, 0.65	0.14, 0.50, 0.62	-0.14, -0.49, 0.63	0.18, 0.81, 0.42
Commissions T2	-0.13, -0.91, 0.37	0.05, 0.35, 0.72	0.02, 0.19, 0.85	-0.12, -0.43, 0.67	<0.001, 0.003, 0.99	-0.57, -2.56, 0.01

Omissions

Concerning the omissions score at T1, none of the three TAS scores were found associated with the likelihood to report omissions at T1 ($p > 0.05$). The interaction effects between EF and the three TAS scores were not statistically significant ($p > 0.05$). The model reached the following fit indices: $R^2 = 0.08$, $RMSE = 0.092$, $F_{(2, 102)} = 1.24$, $p = 0.29$, $\eta^2 = 0.08$, suggesting that the model did not have a good fit (Burnham and Anderson, 2002).

With regard to the omissions score at T2, the main effect of TAS-EOT was statistically significant, $\beta = -0.39$, $t = -3.66$, $p < 0.001$. That is, the higher the EOT score, the lower the omissions at T2. In contrast, the direct effects of TAS-DIF and TAS-DDF did not reach statistical significance ($p > 0.05$). In addition, the interaction effects between EF and the three TAS scores were not statistically significant ($p > 0.05$). The model fit indices were $R^2 = 0.13$, $RMSE = 0.104$, $F_{(2, 102)} = 2.16$, $p = 0.04$, $\eta^2 = 0.13$, indicating that the model did not have a good fit (Burnham and Anderson, 2002).

Commissions

Regarding the commissions score at T1, none of the three TAS scores predicted the likelihood to report commissions at T1 as well as the interaction effects ($p > 0.05$). The model fit indices were acceptable, $R^2 = 0.07$, $RMSE = 0.082$, $F_{(2, 102)} = 1.20$, $p = 0.31$, $\eta^2 = 0.08$.

Finally, concerning the commissions score at T2, none of the main effects of TAS scores was statistically significant ($p > 0.05$). The interaction effect EF*EOT was statistically significant, $\beta = -0.57$, $t = -2.56$, $p = 0.01$. Simple slopes analyses (i.e., moderation analyses) showed that as EF score decreases, a negative relationship between EOT and commissions at T2 becomes significant, $\beta = -0.01$, $t = -2.46$, $p = 0.02$. The other interaction effects of EF and TAS subscales were not statistically significant ($p > 0.05$). The model reached good fit indices $R^2 = 0.09$, $RMSE = 0.077$, $F_{(2, 102)} = 1.49$, $p = 0.18$, and $\eta^2 = 0.09$ (Tables 3, 4).

DISCUSSION

The current study aimed to verify the role of alexithymia on memory accuracy for a violent crime. Collectively and in accordance with prior studies (e.g., Vermeulen et al., 2010; Corroero et al., 2019; Ridout et al., 2020), we found that a higher

TABLE 4 | Mean proportions of the memory scores (i.e., correct details, omissions, commissions) reported during the first (T1) and the second (T2) memory test by participants.

Memory scores	T1	T2
Correct details	0.37 (0.07) 95% CI [0.36, 0.38]	0.41 (0.08) 95% CI [0.40, 0.43]
Omissions	0.40 (0.10) 95% CI [0.38, 0.41]	0.34 (0.11) 95% CI [0.32, 0.36]
Commissions	0.23 (0.09) 95% CI [0.22, 0.25]	0.58 (0.08) 95% CI [0.57, 0.60]

Standard deviations and 95% CI are shown between parentheses.

alexithymia level (only in terms of EOT score) is associated with lower correct details and higher memory distortions reported by participants during the recall of the event, especially with regard to the EF resources of individuals. That is, the lower the EF, the higher the negative effect of TAS on memory performance. By contrast, we did not find a similar pattern of results for the DIF and DDF scores.

Alexithymia, Correct Details, and EF

To begin with, we demonstrated that the EF resources of individuals are positively associated with the proportion of correct details at T1. The higher the EF resources, the higher the proportion of correct details reported by participants at T1. This is completely in accordance with studies conducted so far showing that having high EF resources facilitates the encoding of an event by resulting in a better retrieval (e.g., Battista et al., 2020). Due to their availability of EF resources, high EF people are more able to shift among the details of an event (i.e., Shifting), suppress interferences and irrelevant information (i.e., Inhibition), and monitor relevant information (i.e., Updating) at the encoding than low EF people. Having a better encoding of original material allows them to better retrieve that material. In contrast with the expectation (Hypothesis 1), the analysis showed no main effect of the three scores of alexithymia (i.e., TAS-DIF, TAS-DDF, and TAS-EOT) on the recollection of correct details at both T1 and T2. This is also in contrast with the results achieved by previous studies (e.g., Vermeulen et al., 2010; DiStefano and Koven, 2012; Takahashi et al., 2015; Corroero et al., 2019; Ridout et al., 2020). For example, Ridout et al. (2020) found that at least the ability to identify and interpret feelings

and emotions (i.e., DIF) is negatively associated with the correct recall of emotional information for complex stimuli, such as short video clips. The authors stressed the idea that the inability to correctly process emotional information during the encoding results in a worse retrieval of such information. Probably, the lack of evidence can be explained by considering the type of stimuli adopted. Indeed, in this study, we used a video of a mock crime showing the murder following a quarrel between two people. Although the analyses confirmed that the video emotionally affected the emotional states of the participants by decreasing the initial positive state, we did not find an increase in the negative states of the participants after watching the video. Therefore, it can be that we failed to replicate a direct effect of alexithymia scores on memory recall because the video partially impacted the emotional states of the participants. Although there was no main effect of the three alexithymia scores, the results demonstrated that the interaction between EOT (i.e., the characteristic of directing thoughts at external reality and hardly or not at inner experience) and EFs affects the correct recall of the details of the event at T2. Indeed, we found that the negative effect of EOT on the proportion of correct details recalled at T2 increased with a decrease in the availability of EF resources of an individual. Although some scholars have pointed out that EOT is not a proper representative of the characteristic of alexithymia (e.g., Kooiman et al., 2002), the results underline the idea that EOT, because it is the characteristic of alexithymia that impedes people to have internal monitoring (e.g., Correro et al., 2019), does not allow the use of internal cognitive control that promotes memory processes (Dressaire et al., 2015). Moreover, in line with the Hypothesis 3 for which we expected that alexithymia would have affected the recall by interacting with the availability of EF resources of an individual, we found that the negative association between having an externally oriented style of thinking and the amount of correct details reported is moderated by the availability of EF resources of an individual. These findings are in accordance with previous findings (e.g., Henry et al., 2006; Correro et al., 2019) for two reasons. First, we confirm that EOT seems to be the main characteristic of alexithymia that influences memory performance instead of the abilities to identify and describe feelings and emotions (i.e., DIF and DDF). Indeed, in a recent study and across three experiments, Correro et al. (2019) have shown that the negative effects of alexithymia on the encoding and, in turn, on the retrieval of information can depend on EOT. Second, we provided further support to the idea that such a negative influence of alexithymia on memory is strictly linked to the EFs of the individuals (e.g., Henry et al., 2006; Correro et al., 2019; Ridout et al., 2020). Alexithymia does not directly affect memory performance but only when interacts with the EF resources of individuals. This is also in line with studies showing that alexithymia is associated with executive dysfunction (e.g., Henry et al., 2006; Correro et al., 2019; Ridout et al., 2020). As a matter of fact, in the screening analyses, we also found that a higher amount of alexithymia is associated with lower EF resources. We found this negative association with regard to the DIF score rather than the EOT score. Note that although we found a statistically significant interaction effect between EF and EOT on correct details score, the findings of the screening analysis are fully concordant with prior studies underlying the

negative association between EF and alexithymia (i.e., DIF) (e.g., Henry et al., 2006; Correro et al., 2019).

Alexithymia, Memory Distortions, and EF

The findings on memory distortions (i.e., omissions and commissions) stressed an interesting pattern of results. We first found that the likeliness to report omissions (i.e., details encoded but not recalled) at T1 is negatively associated with the EF score. Thus, the lower the EF score, the higher the omissions at T1. This finding confirmed previous research showing that low EF individuals are more prone to report memory distortions than high EF individuals (e.g., Battista et al., 2020). However, we did not find a similar pattern of results with regard to the omissions reported at T2. This seems to suggest that EF abilities influence the proneness to report this kind of distortions only immediately after the event. A possible explanation could be that the differences in cognition of people, that typically inform on the formation of memory distortions, are downsized by the delay in the recollection, and this leads to similar memory performance. In addition, the data showed that the likelihood to report this type of distortion is related to the main effect of the EOT score. Indeed, we also found a significant direct effect of EOT on omissions scores at T2: The higher EOT, the lower the omissions regardless of the EF availability of the individuals. This evidence is in contrast with Hypothesis 2: We found an effect of alexithymia score (i.e., EOT) but in a contrasting direction. The results suggest that a higher tendency to EOT leads to report fewer omissions at T2. That is, participants with a high level of EOT are able to recall more event-related details than participants with a low level of EOT. However, so far, scholars have argued that EOT interferes with the encoding of information by resulting in a lower recall of information (e.g., Correro et al., 2019). This makes it reasonable to believe that this inability to correctly encode information would lead to reporting lower information during the retrieval. However, the finding seems to provide two important insights: (i) The influence of the EOT level of individuals depend on when people recall the event-related information and (ii) people with high EOT level report more information (i.e., fewer omissions) during a late recall than people with low EOT level. A possible explanation can be that EOT negatively impacts the encoding and, in turn, the retrieval in terms of the correctness (i.e., true vs. distorted or false) of the information reported rather than in the amount of information recalled. The findings seem to support the idea that having EOT does not lead to forgetting of the experienced information (e.g., Vermeulen and Luminet, 2009; Ridout et al., 2020).

The pattern of results concerning the commission score can be explained as follows. We found that EF resources are also associated with the amount of commissions (i.e., details distorted or false with respect to the details encoded) at T1. Again, we demonstrated that the lower the EF score, the higher the commissions reported at T1. This is in accordance with prior studies demonstrating that a high amount of distorted information or information never encoded is due to lower EF resources (i.e., Mirandola et al., 2015; Battista et al., 2020). In addition, this result perfectly fits with the data of the current study on the correct details and with prior evidence showing that the likelihood to report more commissions goes hand in

hand with the likelihood to report lower correct details (e.g., Battista et al., 2020). Contrary to our expectations, no statistically significant main effect of alexithymia (i.e., TAS-DIF, TAS-DDF, and TAS-EOT) was found for the commissions reported at both T1 and T2. However, we showed that EOT score significantly interacted with EF resources on the score of the commission at T2. Specifically, we demonstrated that the EF resources of individuals moderate the effect of EOT on commissions recalled at T2: As EF decreases, the negative association between EOT and commissions becomes statistically significant. Again, this is in line with the findings on the correct details score at T2 and Hypothesis 3. Combining the explanations provided by the line of research on EF and memory distortions (e.g., Gerrie and Garry, 2007; Peters et al., 2007; Unsworth and Brewer, 2010; Leding, 2012; Battista et al., 2020) and the second line on alexithymia and memory (e.g., Correro et al., 2019; Ridout et al., 2020), we can explain this latter finding as follows. In general, having lower EF resources makes people more susceptible to encode interferences and less focused on the relevant event-related information during the encoding. This, in turn, makes it difficult to distinguish the original information from the one imagined during the retrieval resulting in more commissions than high EF individuals (e.g., Roediger and McDermott, 1995). Moreover, scholars have been argued that a higher external-oriented style of thinking (i.e., EOT) does not permit to employ the necessary cognitive resources to correctly encode incoming information (e.g., Correro et al., 2019). Hence, the EF resources of individuals further affect the normal bad encoding of participants with high EOT by making people more likely to report commissions.

Limitations

Although the study has the merit to provide new evidence on unexplored questions, such as the role of alexithymia on the formation of memory distortions, some limitations need to be considered. For instance, the study was correlational and, thus, although we have provided information on the relationship between alexithymia and memory, we were not able to provide a cause-and-effect direction between the variables of interest. Furthermore, the composition of the sample could have underestimated the results. Indeed, we tested students with a small range of age and, thus, similar EF. The findings could be not representative of what occurs in the normal population. Moreover, related to this first limitation, it could be the case that the investigation of the relationship between alexithymia and EF might be different in a sample composed of people categorized as alexithymic (i.e., individuals reporting a level of alexithymia higher than the cutoff). Another limitation is due to the emotional stimulus adopted. We indeed wanted to test how alexithymia interferes in the recall of an emotional and complex stimulus. We thus administered to participants a video of mock violent crime that showed in previous research an emotional impact on the state of the participants. However, in the present experiment, the video partially impacted the state of the participants. Hence, it is possible that different findings could occur for a stronger emotional stimulus. Moreover, because the primary aim of our study was to replicate a real forensic situation as much as possible, we did not include a non-emotional stimulus. Nevertheless, including a control

situation (i.e., non-emotional video) could further inform on the relationship between alexithymia and the ability to recall an experience. Further studies are necessary to gap these caveats.

GENERAL CONCLUSION AND PRACTICAL IMPLICATIONS

In conclusion, the present experiment provided further information on the relationship between alexithymia, EF, and memory. A high degree of alexithymia (i.e., EOT score) leads to negative mnemonic effects with regard to both correct details and memory distortions, and this depends also on the availability of EF of an individual. Our findings are particularly helpful for legal professionals that deal with the memory-related statements of individuals in legal proceedings. Collectively, the results suggest that alexithymia can contribute to the formation of memory distortions; hence, it is an important factor to take into consideration for the assessment of the reliability of witnesses, suspects, and the statements of victims. Legal professionals have to keep in mind that participants with a high level of alexithymia or people with a diagnosis of alexithymia might unintentionally provide unreliable information. This is particularly relevant within the legal arena as statements represent one of the main sources for legal practitioners (e.g., judges) to correctly perform their job and that can lead to accurate or wrong legal decisions (e.g., conviction, absolution). In addition, on a more general note, the findings of the current study can be relevant also for clinical psychologists that work with alexithymic patients. Indeed, such findings can inform them of the (un)ability of patients to correctly retrieve emotional experiences useful to structure a clinical intervention.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Education, Psychology, Communication of the University of Bari Aldo Moro. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FB conceived the study, conducted data collection, analyze the data, and wrote the manuscript. TL and AC critically revised the manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.669778/full#supplementary-material>

REFERENCES

- Aftanas, L. I., Varlamov, A. A., Reva, N. V., and Pavlov, S. V. (2003). Disruption of early event-related theta synchronization of human EEG in alexithymics viewing affective pictures. *Neurosci. Lett.* 340, 57–60. doi: 10.1016/S0304-3940(03)00070-3
- Bagby, R. M., Parker, J. D., and Taylor, G. J. (1994a). The twenty-item toronto alexithymia scale -I. Item selection and cross-validation of the factor structure. *J. Psychosomat. Res.* 38, 23–32. doi: 10.1016/0022-3999(94)90005-1
- Bagby, R. M., Taylor, G. J., and Parker, J. D. (1994b). The twenty-item Toronto Alexithymia Scale—II. Convergent, discriminant, and concurrent validity. *J. Psychosomat. Res.* 38, 33–40. doi: 10.1016/0022-3999(94)90006-X
- Battista, F., Otgaar, H., Lanciano, T., and Curci, A. (2020). Individual differences impact memory for a crime: a study on executive functions resources. *Consciousn. Cogn.* 84:103000. doi: 10.1016/j.concog.2020.103000
- Battista, F., Otgaar, H., Mangiulli, I., and Curci, A. (2021). The role of executive functions in the effects of lying on memory. *Acta Psychol.* 215:103295. doi: 10.1016/j.actpsy.2021.103295
- Bezdjian, S., Baker, L. A., Lozano, D. I., and Raine, A. (2009). Assessing inattention and impulsivity in children during the Go/NoGo task. *Br. J. Dev. Psychol.* 27, 365–383. doi: 10.1348/026151008X314919
- Borhani, K., Borgomaneri, S., Ládavas, E., and Bertini, C. (2016). The effect of alexithymia on early visual processing of emotional body postures. *Biol. Psychol.* 115, 1–8. doi: 10.1016/j.biopsycho.2015.12.010
- Brady, R. E., Bujarski, S. J., Feldner, M. T., and Pyne, J. M. (2016). Examining the effects of alexithymia on the relation between posttraumatic stress disorder and over-reporting. *Psychol. Trauma Theory Res. Practice Policy* 9:80. doi: 10.1037/tra0000155
- Bressi, C., Taylor, G., Parker, J., Bressi, S., Brambilla, V., Aguglia, E., et al. (1996). Crossvalidation of the factor structure of the 20-item Toronto Alexithymia Scale: an Italian multicenter study. *J. Psychosomat. Res.* 4, 551–559. doi: 10.1016/S0022-3999(96)00228-0
- Burgess, P. W., and Simons, J. S. (2005). “Theories of frontal lobe executive function: clinical applications,” in *Effectiveness of Rehabilitation for Cognitive Deficits*, ed P. W. Halligan, D. T. Wade (New York, NY: Oxford Univ. Press), 211–31. doi: 10.1093/acprof:oso/9780198526544.003.0018
- Burnham, K. P., and Anderson, D. R. (2002). *A Practical Information-Theoretic Approach. Model Selection and Multimodel Inference, 2nd edn.* New York, NY: Springer.
- Caretti, V., Porcelli, P., Solano, L., Schimmenti, A., Bagby, R. M., and Taylor, G. J. (2011). Reliability and validity of the Toronto Structured Interview for Alexithymia in a mixed clinical and nonclinical sample from Italy. *Psychiatry Res.* 187, 432–436. doi: 10.1016/j.psychres.2011.02.015
- Carpenter, P. A., Just, M. A., and Reichle, E. D. (2000). Working memory and executive function: evidence from neuroimaging. *Curr. Opin. Neurobiol.* 10, 195–199. doi: 10.1016/S0959-4388(00)00074-X
- Carretti, B., Belacchi, C., and Cornoldi, C. (2010). Difficulties in working memory updating in individuals with intellectual disability. *J. Intellect. Disabil. Res.* 54, 337–345. doi: 10.1111/j.1365-2788.2010.01267.x
- Correro, A. N., Paitel, E. R., Byers, S. J., Nielson, K. A. (2019). The role of alexithymia in memory and executive functioning across the lifespan. *Cogn. Emot.* 2019, 1–16. doi: 10.1080/02699931.2019.1659232
- da Silva, A. N., Vasco, A. B., and Watson, J. C. (2017). Alexithymia and emotional processing: A mediation model. *J. Clin. Psychol.* 73, 1196–1205. doi: 10.1002/jclp.22422
- Dempster, F. N., and Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educ. Psychol. Rev.* 11, 1–88. doi: 10.1023/A:1021992632168
- Diamond, A. (2013). Executive functions. *Ann. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Diekhof, E., Geier, K., Falkai, P., and Gruber, O. (2011). Fear is only as deep as the mind allows: a coordinate-based meta-analysis of neuroimaging studies on the regulation of negative affect. *Neuroimage* 58, 275–285. doi: 10.1016/j.neuroimage.2011.05.073
- DiStefano, R. A., and Koven, N. S. (2012). Dysfunctional emotion processing may explain visual memory deficits in alexithymia. *Personal. Individ. Differ.* 52, 611–615. doi: 10.1016/j.paid.2011.12.007
- Donges, U. S., and Suslow, T. (2015). Alexithymia and memory for facial emotions. *Universitas Psychol.* 14, 103–110. doi: 10.11144/javeriana.upsy14-1.amfe
- Dressaire, D., Stone, C. B., Nielson, K. A., Guerdoux, E., Martin, S., Brouillet, D., et al. (2015). Alexithymia impairs the cognitive control of negative material while facilitating the recall of neutral material in both younger and older adults. *Cogn. Emot.* 29, 442–459. doi: 10.1080/02699931.2014.919898
- Duff, K., Schoenberg, M. R., Scott, J. G., and Adams, R. L. (2005). The relationship between executive functioning and verbal and visual learning and memory. *Archiv. Clin. Neuropsychol.* 20, 111–122. doi: 10.1016/j.acn.2004.03.003
- Duncan Roger Johnson Michaela Swales Charles Freer, J. (1997). Frontal lobe deficits after head injury: unity and diversity of function. *Cogn. Neuropsychol.* 14, 713–741. doi: 10.1080/026432997381420
- Espy, K. A. (2004). Using developmental, cognitive, and neuroscience approaches to understand executive control in young children. *Dev. Neuropsychol.* 26, 379–384. doi: 10.1207/s15326942dn2601_1
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Franz, M., Popp, K., Schaefer, R., Sitte, W., Schneider, C., Hardt, J., et al. (2008). Alexithymia in the German general population. *Soc. Psychiatry Psychiatric Epidemiol.* 43, 54–62. doi: 10.1007/s00127-007-0265-1
- Gerrie, M. P., and Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory* 15, 561–571. doi: 10.1080/09658210701391634
- Gonsalves, B., and Paller, K. A. (2002). Mistaken memories: remembering events that never happened. *Neuroscientist* 8, 391–395. doi: 10.1177/107385802236964
- Grynberg, D., and Pollatos, O. (2015). Alexithymia modulates the experience of the rubber hand illusion. *Front. Hum. Neurosci.* 9:357. doi: 10.3389/fnhum.2015.00357
- Henry, J. D., Phillips, L. H., Crawford, J. R., Theodorou, G., and Summers, F. (2006). Cognitive and psychosocial correlates of alexithymia following traumatic brain injury. *Neuropsychologia* 44, 62–72. doi: 10.1016/j.neuropsychologia.2005.04.011
- Hill, B. D., Alosco, M., Bauer, L., and Tremont, G. (2012). The relation of executive functioning to CVLT-II learning, memory, and process indexes. *Appl. Neuropsychol. Adult* 19, 198–206. doi: 10.1080/09084282.2011.643960
- Jacob, S., and Hautekeete, M. (1998). Alexithymia and memory: a more rigorous criterion for acceptance in recognition tasks? *Encephale-Revue De Psychiatrie Clinique Biologique Et Therapeutique* 24, 199–204.
- Jaschinski, U., and Wentura, D. (2002). Misleading postevent information and working memory capacity: an individual differences approach to eyewitness memory. *Appl. Cogn. Psychol.* 16, 223–231. doi: 10.1002/acp.783
- Jersild, A. T. (1927). Mental set and shift. *Archiv. Psychol.* 14:81.
- Kersten, A. W., Earles, J. L., Auccello, K., Tautiva, E., McRostie, N., Brydon, C., et al. (2018). Influences of executive and memory functioning on memory for the sources of actions. *Psychol. Aging* 33, 1115–1133. doi: 10.1037/pag0000312
- Kimberg, D. Y., and Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: the role of working memory in complex, organized behavior. *J. Exp. Psychol.* 122:411. doi: 10.1037/0096-3445.122.4.411
- Kooiman, C. G., Spinhoven, P., and Trijsburg, R. W. (2002). The assessment of alexithymia: a critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *J. Psychosomat. Res.* 53, 1083–1090. doi: 10.1016/S0022-3999(02)00348-3
- Koven, N. S., and Thomas, W. (2010). Mapping facets of alexithymia to executive dysfunction in daily life. *Personal. Individ. Differ.* 49, 24–28. doi: 10.1016/j.paid.2010.02.034
- Lane, R. D., Sechrest, L., Reidel, R., Weldon, V., Kaszniak, A., and Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomat. Med.* 58, 203–210. doi: 10.1097/00006842-199605000-00002
- Leding, J. K. (2012). Working memory predicts the rejection of false memories. *Memory* 20, 217–223. doi: 10.1080/09658211.2011.653373

- Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quart. J. Exp. Psychol. A* 49, 29–50. doi: 10.1080/027249896392793
- Li, S., Zhang, B., Guo, Y., and Zhang, J. (2015). The association between alexithymia as assessed by the 20-item Toronto Alexithymia Scale and depression: a meta-analysis. *Psychiatry Res.* 227, 1–9. doi: 10.1016/j.psychres.2015.02.006
- Love, J., Dropmann, D., Selker, R., Gallucci, M., Jentschke, S., and Balci, S. (2021). *The jamovi Project*. jamovi [Computer Software]. Retrieved from: <https://www.jamovi.org>
- Luminet, O., Rimé, B., Bagby, R. M., and Taylor, T. (2004). A multimodal investigation of emotional responding in alexithymia. *Cogn. Emot.* 18, 741–766. doi: 10.1080/02699930341000275
- Luminet, O., Vermeulen, N., Demaret, C., Taylor, G. J., and Bagby, R. M. (2006). Alexithymia and levels of processing: evidence for an overall deficit in remembering emotion words. *J. Res. Personal.* 40, 713–733. doi: 10.1016/j.jrp.2005.09.001
- Lundh, L. G., Johnsson, A., Sundqvist, K., and Olsson, H. (2002). Alexithymia, memory of emotion, emotional awareness, and perfectionism. *Emotion* 2, 361–379. doi: 10.1037/1528-3542.2.4.361
- Mangiulli, I., Lanciano, T., van Oorsouw, K., Jelicic, M., and Curci, A. (2019). Do reminders of the crime reverse the memory-undermining effect of simulating amnesia? *Memory Cogn.* 19:939. doi: 10.3758/s13421-019-00939-z
- Marsh, E. J., Balota, D. A., and Roediger, H. L. (2005). Learning facts from fiction: effects of healthy aging and early-stage dementia of the Alzheimer type. *Neuropsychology* 19, 115–129. doi: 10.1037/0894-4105.19.1.115
- Mather, M., and Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspect. Psychol. Sci.* 6, 114–133. doi: 10.1177/1745691611400234
- Meltzer, M. A., and Nielson, K. A. (2010). Memory for emotionally provocative words in alexithymia: a role for stimulus relevance. *Consciousn. Cogn.* 19, 1062–1068. doi: 10.1016/j.concog.2010.05.008
- Mirandola, C., Toffalini, E., Ciriello, A., and Cornoldi, C. (2015). Working memory affects false memory production for emotional events. *Cogn. Emot.* 31, 33–46. doi: 10.1080/02699931.2015.1075379
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Monzell, S. (1996). “Control of mental processes,” in *Unsolved Mysteries of the Mind: Tutorial Essays in Cognition*, ed V. Bruce (Hove: Erlbaum), 93–148.
- Morris, N., and Jones, D. M. (1990). Memory updating in working memory: the role of the central executive. *Br. J. Psychol.* 81, 111–121. doi: 10.1111/j.2044-8295.1990.tb02349.x
- Mueller, S. T., and Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *J. Neurosci. Methods* 222, 250–259. doi: 10.1016/j.jneumeth.2013.10.024
- Nemiah, J. C. (1977). Alexithymia. Theoretical considerations. *Psychotherapy Psychosomat.* 28, 199–206. doi: 10.1159/000287064
- Nielson, K. A., and Corroero, A. N. 2nd (2017). Post-learning arousal enhances veridical memory and reduces false memory in the Deese-Roediger-McDermott paradigm. *Neurobiol. Learn. Memory* 144, 198–207. doi: 10.1016/j.nlm.2017.07.009
- Novelli, G., Papagno, C., Capitani, E., and Laiacina, M. (1986). Tre test clinici di ricerca e produzione lessicale. Taratura su soggetti normali [Three clinical tests to research and rate the lexical performance of normal subjects]. *Arch. Psicol. Neurol. Psichiatria* 47, 477–506.
- Onor, M., Trevisiol, M., Spano, M., Aguglia, E., and Paradiso, S. (2010). Alexithymia and aging: a neuropsychological perspective. *J. Nervous Mental Dis.* 198, 891–895. doi: 10.1097/NMD.0b013e3181fe743e
- Paivio, S. C., and McCulloch, C. R. (2004). Alexithymia as a mediator between childhood trauma and self-injurious behaviors. *Child Abuse Neglect* 28, 339–354. doi: 10.1016/j.chiabu.2003.11.018
- Peters, M. J. V., Jelicic, M., Verbeek, H., and Merckelbach, H. (2007). Poor working memory predicts false memories. *Eur. J. Cogn. Psychol.* 19, 213–232. doi: 10.1080/09541440600760396
- Peterson, R. (2004). Nameless desire: Alexithymia and the anorectic patient. *Am. J. Psychoanal.* 64, 77–90. doi: 10.1023/B:TAJP.0000017992.93812.c6
- Pollatos, O., and Gramann, K. (2011). Electrophysiological evidence of early processing deficits in alexithymia. *Biol. Psychol.* 87, 113–121. doi: 10.1016/j.biopsycho.2011.02.016
- Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instr. Comput.* 36, 717–731. doi: 10.3758/BF03206553
- Ridout, N., Smith, J., and Hawkins, H. (2020). The influence of alexithymia on memory for emotional faces and realistic social interactions. *Cogn. Emot.* 35, 540–558. doi: 10.1080/02699931.2020.1747991
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Memory Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Santorelli, G. D., and Ready, R. E. (2015). Alexithymia and executive function in younger and older adults. *Clin. Neuropsychol.* 29, 938–955. doi: 10.1080/13854046.2015.1123296
- Schacter, D. L., and Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron* 44, 149–160. doi: 10.1016/j.neuron.2004.08.017
- Serafini, G., Gonda, X., Canepa, G., Pompili, M., Rihmer, Z., Amore, M., et al. (2017). Extreme sensory processing patterns show a complex association with depression, and impulsivity, alexithymia, and hopelessness. *J. Affect. Disord.* 210, 249–257. doi: 10.1016/j.jad.2016.12.019
- Serafini, G., Gonda, X., Pompili, M., Rihmer, Z., Amore, M., and Engel-Yeger, B. (2016). The relationship between sensory processing patterns, alexithymia, traumatic childhood experiences, and quality of life among patients with unipolar and bipolar disorders. *Child Abuse Neglect* 62, 39–50. doi: 10.1016/j.chiabu.2016.09.013
- Sifneos, P. E. (1973). The prevalence of “alexithymic” characteristics in psychosomatic patients. *Psychotherapy Psychosomat.* 22, 255–262. doi: 10.1159/000286529
- Spector, A., and Biederman, I. (1976). Mental set and mental shift revisited. *Am. J. Psychol.* 89:669. doi: 10.2307/1421465
- Suslow, T., Kersting, A., and Arolt, V. (2003). Alexithymia and incidental learning of emotional words. *Psychol. Rep.* 93, 1003–1012. doi: 10.2466/pr0.2003.93.3f.1003
- Swart, M., Kortekaas, R., and Aleman, A. (2009). Dealing with feelings: characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. *PLoS ONE* 4:e5751. doi: 10.1371/journal.pone.0005751
- Takahashi, J., Hirano, T., and Gyoba, J. (2015). Effects of facial expressions on visual short-term memory in relation to alexithymia traits. *Personal. Individ. Differ.* 83, 128–135. doi: 10.1016/j.paid.2015.04.010
- Taylor, G. J. (1984). Alexithymia: concept, measurement, and implications for treatment. *Am. J. Psychiatry* 141, 725–732. doi: 10.1176/ajp.141.6.725
- Taylor, G. J. (2000). Recent developments in alexithymia theory and research. *Can. J. Psychiatry* 45, 134–142. doi: 10.1177/070674370004500203
- Taylor, G. J., Bagby, R. M., and Parker, J. D. A. (1997). *Disorders of Affect Regulation: Alexithymia in Medical and Psychiatric Illness*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511526831
- Terraciano, A., McCrae, R. R., and Costa, Jr. P. T. (2003). Factorial and construct validity of the Italian positive and negative affect schedule (PANAS). *Eur. J. Psychol. Assess.* 19, 131–141. doi: 10.1027//1015-5759.19.2.131
- Unsworth, N., and Brewer, G. A. (2010). Individual differences in false recall: a latent variable analysis. *J. Mem. Lang.* 62, 19–34. doi: 10.1016/j.jml.2009.08.002
- van der Velde, J., Gromann, P. M., Swart, M., Wiersma, D., de Haan, L., Bruggeman, A. (2015). Alexithymia influences brain activation during emotion perception but not regulation. *Soc. Cogn. Affect. Neurosci.* 10, 285–293. doi: 10.1093/scan/nsu056
- Van der Velde, J., Servaas, M. N., Goerlich, K. S., Bruggeman, R., Horton, P., Costafreda, S. G., et al. (2013). Neural correlates of alexithymia: a meta-analysis of emotion processing studies. *Neurosci. Biobehav. Rev.* 37, 1774–1785. doi: 10.1016/j.neubiorev.2013.07.008
- Vermeulen, N., Domachowska, I., and Nielson, K. A. (2018). “Memory and executive functions in alexithymia,” in *Alexithymia: Advances in Research, Theory, and Clinical Practice*, eds O. Luminet, R. M. Bagby, and G. J. Taylor (Cambridge: Cambridge University Press), 78–89. doi: 10.1017/9781108241595.007

- Vermeulen, N., and Luminet, O. (2009). Alexithymia factors and memory performances for neutral and emotional words. *Personal. Individ. Differ.* 47, 305–309. doi: 10.1016/j.paid.2009.03.018
- Vermeulen, N., Luminet, O., and Corneille, O. (2006). Alexithymia and the automatic processing of affective information: evidence from the affective priming paradigm. *Cogn. Emot.* 20, 64–91. doi: 10.1080/02699930500304654
- Vermeulen, N., Toussaint, J., and Luminet, O. (2010). The influence of alexithymia and music on the incidental memory for emotion words. *Eur. J. Personality* 24, 551–568. doi: 10.1002/per.758
- Wager, T., Davidson, M., Hughes, B., Lindquist, M., and Ochsner, K. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron* 59, 1037–1050. doi: 10.1016/j.neuron.2008.09.006
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Personal. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- Wood, R. L., and Williams, C. (2007). Neuropsychological correlates of organic alexithymia. *J. Int. Neuropsychol. Soc.* 13, 471–479. doi: 10.1017/S1355617707070518
- Zeitlan, S. B., and McNally, R. J. (1993). Alexithymia and anxiety sensitivity in panic disorder and obsessive compulsive disorder. *Am. J. Psychiatry* 150, 658–660. doi: 10.1176/ajp.150.4.658
- Zhang, L., Ye, R., Yu, F., Cao, Z., Zhu, C., Cai, Z., et al. (2012). How does emotional context modulate response inhibition in alexithymia: electrophysiological evidence from an ERP study. *PLoS ONE* 7:e51110. doi: 10.1371/journal.pone.0051110
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., et al. (2010). Individual differences in false memory from misinformation: cognitive factors. *Memory* 18, 543–555. doi: 10.1080/09658211.2010.487051

Conflict of Interest: 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 © 2021 Battista, Lanciano and Curci. 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) and the copyright owner(s) 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.



Prospective Memory and Positivity Bias in the COVID-19 Health Crisis: The Effects of Aging

Alaitz Aizpurua*, Malen Migueles and Ainara Aranberri

Faculty of Psychology, University of the Basque Country UPV/EHU, San Sebastián, Spain

OPEN ACCESS

Edited by:

Maria Soledad Beato,
University of Salamanca, Spain

Reviewed by:

Pedro B. Albuquerque,
University of Minho, Portugal
Sandra Düzel,
Max Planck Institute for Human
Development, Germany

*Correspondence:

Alaitz Aizpurua
alaitz.aizpurua@ehu.eus

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 11 February 2021

Accepted: 02 June 2021

Published: 21 July 2021

Citation:

Aizpurua A, Migueles M and
Aranberri A (2021) Prospective
Memory and Positivity Bias
in the COVID-19 Health Crisis:
The Effects of Aging.
Front. Psychol. 12:666977.
doi: 10.3389/fpsyg.2021.666977

This study aimed to determine whether the observed tendency to remember more positive than negative past events (positivity phenomena) also appears when recalling hypothetical events about the future. In this study, young, middle-aged, and older adults were presented with 28 statements about the future associated with the COVID-19 pandemic, half positive and half negative. In addition, half of these statements were endowed with personal implications while the other half had a more social connotations. Participants rated their agreement/disagreement with each statement and, after a distraction task, they recalled as many statements as possible. There was no difference in the agreement ratings between the three age groups, but the participants agreed with positive statements more than with negative ones and they identified more with statements of social content than of personal content. The younger and older individuals recalled more statements than the middle-aged people. More importantly, older participants recalled more positive than negative statements (positivity effect), and showed a greater tendency to turn negative statements into more positive or neutral ones (positivity bias). These findings showed that the positivity effect occurs in even such complex and situations as the present pandemic, especially in older adults. The results are discussed by reference to the notion of commission errors and false memories resulting from the activation of cognitive biases.

Keywords: positivity effect, COVID-19, aging, future events, false memories, positivity bias, personal and social contents

INTRODUCTION

The main research aim of this study is to analyze recall accuracy and transformations in different age groups when recalling hypothetical positive and negative future events linked to the COVID-19 global pandemic. The study was carried out in a state of alarm due to the health crisis of COVID-19, when all the inhabitants of Spain were in lockdown. Being confined at home involves a significant change of routines, especially those linked to work, studies, and leisure (Benke et al., 2020; Vindegaard and Benros, 2020). It implies a loss of freedom and separation of friends and family. The Spanish culture is highly focused on family life and leisure with friends in open spaces, and confinement represents a novel situation that requires an important adaptation process (for a review on the impact of COVID-19 in Spain, see Balluerka et al., 2020; Odriozola-González et al., 2020a,b; Rodríguez-Rey et al., 2020). At the same time, the actual context offers an opportunity to analyze the cognitive processes involved in this emotionally exceptional situation.

Recent studies show that the pandemic is causing feelings of isolation and economic uncertainty in the general population, which are generating higher levels of anxiety and depression and a reduction in the feeling of well-being compared to pre-health-crisis states (Carstensen et al., 2020; Killgore et al., 2020; Odriozola-González et al., 2020b; Vindegaard and Benros, 2020). The numbers are shocking. Results obtained from surveys in China, Spain, Italy, Iran, Turkey, Nepal, and Denmark (Xiong et al., 2020) show that the situation has altered people's lives, affecting multiple dimensions and generating dramatic increases in stress (8.1–81.9%), anxiety (6.33–50.9%), or depression (14.6–48.3%). The number of suicides associated with joblessness and hopelessness due to an uncertain future in the adult population has increased (Griffiths and Mamun, 2020; Thakur and Jain, 2020). The pandemic has also generated great concern in the university population about the well-being of their family and friends, a negative view of the evolution of their training process, and its impact in the future (Araújo et al., 2020; Odriozola-González et al., 2020a; Zhai and Du, 2020).

All of these aspects may be increased in older people because they are an at-risk population, where contagions leading to death are higher than in the younger population. In addition, the media (TV, press) and social networks at that stage of confinement were filled with news of deaths linked to COVID-19 in nursing homes and hospitals of people who were unaccompanied and without family support in their last moments. While there was still a lack of medical resources and medical instruments for patients with severe symptomatology (e.g., mechanical ventilators), the debate arose as to whether older people should receive such treatments when younger people were in the same situation. The context was significantly more unfavorable for the older population, which could lead to worsening mental health (Armitage and Nellums, 2020; García-Portilla et al., 2020) or suffering from anxiety and depression (e.g., Santini et al., 2020). Cognitive theories of depression indicate that thoughts, inferences, interpretations, and how people attend to and recall fear-related information can be relevant factors to increase depression and anxiety (Mathews et al., 1997; Booth and Sharma, 2020). Taking into account that good emotion regulation requires adequate functioning of the working memory and the inhibitory processes that block access to negative information (Gotlib and Joormann, 2010), older people may be especially vulnerable to mental health problems arising from the pandemic.

Although the global COVID-19 pandemic has paralyzed the world, people's brain activity has not ceased and continues inexorably to recall their experiences and activities before the pandemic, and to imagine and think about the future. The ability to imagine and plan for the future is a crucial mental process in adaptation, which has been studied in different areas of Cognitive Psychology, especially in episodic future thinking, prospective memory, and mind-wandering (for a recent review, see Kvavilashvili and Rummel, 2020). It is well known that not only does memory recall past experiences, it is also the vehicle that allows us to travel mentally through time to the future (Tulving, 1985, 2005). The projection and mental journey into the future to imagine specific events that may occur, is as frequent as remembering experiences from our

past (Finnbogadottir and Berntsen, 2013). Also, thinking about the future activates the same brain areas as remembering past experiences (Addis et al., 2007; Botzung et al., 2008), and both types of episodic thinking have similar characteristics, including sensory and spatial information, and emotion and knowledge about the world (D'Argembeau and Van der Linden, 2004; for a review, see Schacter et al., 2007; Szpunar, 2010). However, there are important differences between thinking about the past and imagining the future. Although both situations involve the recreation and enjoyment of pleasant thoughts and the uncomfortable anticipation of fears and concerns, it has been observed that thoughts about future experiences are more positive than past events (Berntsen and Rubin, 2002; Berntsen and Jacobsen, 2008; García-Bajos et al., 2017; Zaragoza Scherman et al., 2020). Interestingly, according to aging literature, older people tend to remember their past more positively (Kennedy et al., 2004; Schryer and Ross, 2014) and to perceive their future as more idyllic and positive compared to young people (e.g., Berntsen and Rubin, 2002; Burr et al., 2020).

The central concept of this research is individuals' positivity or our preference for positive information as opposed to negative information when performing attention and memory tasks; this preference is enhanced in older adults, a phenomenon known as the positivity effect (Charles et al., 2003; Carstensen and Mikels, 2005; Mather and Carstensen, 2005; Reed and Carstensen, 2012; Schryer and Ross, 2014; García-Bajos et al., 2017). Thus, positivity effect means that, compared to young people, older people react less to negative situations and preferentially attend to and recall emotionally meaningful and positive stimuli (Reed et al., 2014). This positivity effect has been observed with a multitude of materials such as scenes, drawings, and faces (e.g., Charles et al., 2003; Mather and Carstensen, 2005; Reed et al., 2014; Mammarella et al., 2016), words (Kensinger, 2008; Hamilton and Allard, 2020), or autobiographical experiences of the past and recreations of the future (Berntsen and Jacobsen, 2008; Gallo et al., 2011; Cole et al., 2016; García-Bajos et al., 2017). Older people not only show a greater preference for the positive than young people, but they also generate false memories or transform and modify negative content to make it more positive and, thus, achieve greater consistency with their emotional goals and motivations, and higher emotion regulation and well-being (Charles et al., 2003; Carstensen et al., 2020; Zaragoza Scherman et al., 2020). Although the positivity effect is robust and consistent, as shown by the meta-analysis of 100 studies on the subject by Reed et al. (2014), some contrasting results have shown practically no differences between young and older adults (e.g., Kensinger et al., 2002; Grühn et al., 2005). A less-studied aspect is whether middle-aged adults liken their performance to that of young participants or are closer to that of older people (Carstensen and DeLiema, 2018).

Various theories have been proposed to explain the positivity effect. Some theories posit that age-related advantages reflect the avoidance of stressors (Charles, 2010), whereas others maintain that the advantages of age are driven by motivational shifts that direct cognitive and behavioral resources toward positive and meaningful aspects of life (Carstensen et al., 2020). The

Socio-emotional Selectivity Theory (SST; Carstensen et al., 1999; Carstensen, 2006) emphasizes an increase with age to the accessibility of positive information. A person's priorities and motivations change with age. The fragility of life and the reduction of life expectancy lead them to prioritize objectives, ideas, and content that afford them general satisfaction and that are pleasurable and rewarding. Other theories underscore older people's difficulty to recreate and imagine the future and argue that generating and processing positive future events requires less cognitive effort and less time than negative events (Newby-Clark and Ross, 2003; Schacter et al., 2008; Berntsen and Bohn, 2010), mainly because negative content is more complex to process than positive content (Labouvie-Vief et al., 2010). Finally, it is also proposed that older people focus on emotion regulation by implementing their cognitive control resources, such as activating inhibitory resource to block access to negative information (García-Bajos and Migueles, 2017; Giebl et al., 2016; Marsh et al., 2019). That is, cognitive abilities and motivation contribute to the positivity effect.

This study has three priority objectives. First, to analyze in three age groups possible differences between the recall of hypothetical future negative thoughts related to the threats and repercussions of COVID-19 and positive thoughts for the future, desires, and plans after the pandemic. It could be considered that the current situation leads to focusing on COVID-19-related sources of fear (Mathews et al., 1997; Booth and Sharma, 2020) and that this, in turn, can lead to a state of mood-congruent retrieval, focusing recall on negative content (for a review, see: Blaney, 1986). However, the literature also indicates that to increase the sense of well-being and reduce stress and anxiety, people prefer to codify and remember positive aspects, showing a positivity phenomenon. Thus, we expect that participants in our study will show a tendency to process and remember positive statements better than negative ones; and we anticipate that this trend will be more pronounced in older people than in young adults, that is, a positivity effect. Second, we shall examine the transformations, biases, errors, and false memories that emerge to face adversity in individuals of different age groups, and we expect to see a greater positive bias in older than in young people. This finding was hypothesized because, as with other types of false memories, older people tend to use the cognitive and attentional resources available to them to adjust their thoughts to their previous knowledge (e.g., Schacter et al., 1997), and to regulate their emotional state, in this case, by imagining the events of their own future, adapting it to their desires and personal expectations. Thirdly and lastly, we analyze the effects on recall of the social or personal nature of the thoughts about the hypothetical future. It has been observed that the recreation of the future is more likely to be performed in the third person rather than from a first-person perspective (D'Argembeau and Van der Linden, 2004, 2006). In addition, in the current context of pandemic, social content can be perceived as high-value information or more important than personal content (Hargis and Castel, 2017). Therefore, we expect a better recall of social than of personal imagined future events. However, little is known about the effect of the social or personal perception of future experiences, and whether one's perspective interacts with the positive or negative

valence of thoughts and/or with the age of the individuals who recall those thoughts.

In regard to the emotional regulation, although age-related biological, psychological, and socio-economic factors are not favorable for older people, the literature provides us with abundant data indicating that older people's emotional experiences are more stable and positive than younger people's, and they also show a greater sense of well-being (Carstensen et al., 2000; Carstensen, 2006; Stone et al., 2010; Burr et al., 2020). This may be due, at least in part, to the fact that older people have more pronounced mechanisms than young people to adapt to adversity, sources of stress, and emotionally negative events. For example, Carstensen et al. (2020) interviewed people between the age of 18 and 76 in the middle of the COVID-19 pandemic, assessing the frequency and intensity of a range of positive and negative emotions, and the subjectively perceived risk of contagion and complications from the virus. They observed that older people showed relatively greater emotional well-being than young people. In other words, older people appear to have greater resilience and mechanisms to regulate their emotions and deal with adverse situations positively (Fontes and Neri, 2015; Silva Junior et al., 2019). One of these mechanisms is the positivity bias.

Positivity bias in autobiographical memory and episodic future thinking is considered important in mental wellbeing, as a cognitive strategy to reduce stress and depression. When attempting to remember positive and negative thoughts linked to the pandemic, one's memory works to find a way out of the situation. Memory uses adaptive cognitive processes, which reconstruct reality using preexisting knowledge, beliefs, expectations, and desires, and generate errors and distortions (Schacter, 1999, 2021; Schacter et al., 2011); one of these distortions is the positivity bias, which make one less vulnerable to emotional disorders (such as depression and anxiety), and help improve mood.

To summarize, the main objective of this study is to examine in three age groups (young, middle-aged and older adults) how the situation of pandemic and confinement affects the memory of positive and negative thoughts of the future linked to COVID-19. We are interested in determining if the situation of stress, fear and worry that we are experiencing induces us to remember more content congruent with that depressed mood or if, on the contrary, a positive effect appears and we remember more positive content. This tendency to remember more positive than negative content may be accentuated in older people (positivity effect), and possibly so does the tendency to turn negative statements into more positive or neutral ones (positivity bias), both understood as cognitive strategies to achieve emotional regulation and feelings of well-being; *a priority* as we get older. Analyzing these aspects and knowing how middle-aged people behave are priority objectives of this study.

MATERIALS AND METHODS

Participants

Participants were 33 young adults (M age = 20.33, SD = 1.93; range: 19–25 years), 23 middle-aged adults (M age = 42.48,

$SD = 7.29$; range: 28–54 years), and 23 older adults (M age = 64.27, $SD = 5.81$; range 55–77 years). The young participants were students of different degrees at the University of the Basque Country (UPV/EHU), and the older participants came from cultural groups or were undergraduate students at the University of the Basque Country (UPV/EHU) who pursued a humanities career for older people called Experience Classrooms. An *a priori* power analysis was conducted with G*Power (Faul et al., 2009) to determine the sample size required to achieve a medium effect size of $F = 0.25$, with a significance level of $\alpha = 0.05$, and statistical power of 0.80. A 3 (Group: Young vs. Middle-aged vs. Older) $\times 2$ (Valence: Positive vs. Negative) $\times 2$ (Nature: Personal vs. Social) mixed factorial design with group as a between-participants variable, and the emotional valence and nature of the statements as within-participants variables require a minimum sample of 17 for each age group. In this study, at least 22 participants were included in each age group.

Materials and Procedure

This study was carried out following the American Psychological Association standards for the ethical treatment of participants, the Declaration of Helsinki, and was approved by the Ethics Committee of the University of the Basque Country (UPV/EHU). Participants were first informed that the experiment dealt with the positive and negative nature of thoughts about the future linked to the COVID-19 pandemic. Because the alarm status had been decreed in Spain, and the entire population was confined without the possibility of leaving their homes, this experiment was conducted online through Google forms. The survey was disseminated through the student council, coordinators, and undergraduate delegates, and the university's website, which opened a space for studies linked to COVID-19. In the form, participants were asked: "In the stage of lockdown and pandemic due to COVID-19, one thinks about the future and imagines positive experiences and facts, but negative fears and experiences also come to mind about what could happen to us in the near or distant future." They were informed that they would be presented with a total of 32 statements or thoughts about the future and they should rate their level of agreement with each of those thoughts. Of these statements, 28 were experimental, 14 positive and 14 negative, and the other 4 were used to control the primacy and recency effects and were not included in subsequent analyses.

Each participant received positive and negative thoughts randomly (with no more than two positive or negative statements in a row), but they were not instructed that they would subsequently be requested to perform a recall task. The statements were drawn from future estimates from the news, newspapers, and social media's concerns and aspirations. To select the statements, 21 adults who did not participate in the study evaluated the statements in two dimensions: valence (positive or negative) and nature (personal or social), and we chose those that obtained clearly defined scores (more than 70% agreement)¹. Half of the statements were positive (e.g., "I think we will be strengthened by this pandemic"), and the rest were

negative (e.g., "I think this virus will mutate and we won't be able to beat it"). Also, half of the positive and negative ideas were personal (e.g., "This pandemic helps me know myself better") or had a more social connotation (e.g., "Popular concerts and festivals won't come back"). After the participants had received the instructions and agreed to participate in the study, a statement and a scale ranging from 1 (*totally disagree*) to 7 (*totally agree*) appeared on the screen to rate their level of agreement.

The participants worked at their own pace and after rating all the sentences, a distracting task consisting of writing words that started with S for 3 min appeared. After the distracting task, a free recall task was administered. Participants were encouraged to write as many of the previously presented thoughts about the future as possible, in any order. This free recall task has the additional advantage of revealing participants' strategies to organize the material. The experimental phase lasted approximately 10–15 min.

Design

The present study employed a 3 (Group: Young vs. Middle-aged vs. Older) $\times 2$ (Valence: Positive vs. Negative) $\times 2$ (Nature: Personal vs. Social) mixed factorial design with group as a between-participants variable, and the emotional valence and nature of the statements as within-participants variables. Correct recall and errors were measured for the positive and negative contents produced by each participant.

RESULTS

Rating Thoughts About the Future

Participants were asked to rate their level of agreement with each of the statements about the near future. The agreement level was higher than the average value of 3.5, both in general and in all the age groups (all $ps < 0.001$). In order to estimate the internal consistency of the material employed, Cronbach's Alpha was calculated for the agreement with the positive and negative items in the sample studied. These two values indicate a good level of reliability for both the positive ($\alpha = 0.77$) and negative ($\alpha = 0.73$) statements.

In order to analyze the ratings given by the participants, a 3 (Group: Young vs. Middle-aged vs. Older) $\times 2$ (Valence: Positive vs. Negative) $\times 2$ (Nature: Personal vs. Social) ANOVA was conducted. The analyses indicated that the Group factor was non-significant in the ratings, $F < 1$. That is, there were no differences between young, middle-aged, and older adults in their level of agreement with the statements (see Table 1).

The Valence variable was significant, $F(1, 76) = 65.33$, $p < 0.001$, $\eta_p^2 = 0.462$, indicating a higher level of general agreement with positive than with negative thoughts ($M = 4.58$, $SD = 0.82$, vs. $M = 3.61$, $SD = 0.77$). The effects of the variable Nature, $F(2, 76) = 26.63$, $p < 0.001$, $\eta_p^2 = 0.260$, were also significant, as higher agreement was observed for the social statements than for the personal ones ($M = 4.23$, $SD = 0.69$,

¹ The participants who valued the sentences were representative of the three age groups and similar to the total sample examined in the study; specifically, they

were 6 young adults (M age = 20.33, $SD = 1.93$), 8 middle-aged (M age = 42.48, $SD = 7.29$), and 7 old adults (M age = 63.87, $SD = 5.99$), and 13 women and 8 men.

TABLE 1 | Mean proportion and standard deviations (in parentheses) in total agreement with the statements depending on their valence (positive, negative) and nature (personal, social), in the different age groups.

Group	Agreement	Positive	Negative	Personal	Social
Young	4.18 (0.43)	4.65 (0.66)	3.64 (0.75)	4.14 (0.33)	4.13 (0.64)
Middle-aged	4.01 (0.65)	4.55 (0.88)	3.37 (0.67)	3.64 (0.66)	4.27 (0.69)
Older	4.20 (0.78)	4.51 (0.98)	3.81 (0.86)	4.01 (0.86)	4.32 (0.77)
Total	4.14 (0.61)	4.58 (0.82)	3.61 (0.77)	3.96 (0.65)	4.23 (0.69)

vs. $M = 3.96$, $SD = 0.65$). The significant Group $\times$ Nature interaction, $F(1, 76) = 9.98$, $p < 0.001$, $\eta_p^2 = 0.208$, showed that the level of agreement was higher for social statements than for personal ones in middle-aged adults, $t(22) = 6.21$, $p < 0.001$, and older adults, $t(22) = 3.59$, $p = 0.002$, but not in young adults, who rated the dimensions of both statements equally. The Valence $\times$ Nature interaction was also significant, $F(1, 76) = 50.76$, $p < 0.001$, $\eta_p^2 = 0.400$, indicating a greater level of agreement with positive statements of a social nature ($M = 4.50$, $SD = 0.93$), followed by personal positive statements ($M = 4.66$, $SD = 0.86$), negative social statements ($M = 3.96$, $SD = 0.92$), and finally, negative statements of a personal nature ($M = 3.26$, $SD = 0.84$), with all the differences between them statistically significant, $p < 0.05$. To some extent, these ratings of the level of agreement show a preference for positive aspects when individuals imagine their future, a consistent result with previous outcomes related to positivity.

Recall of Thoughts About the Future

In the recall task, those participants who recalled two or less correct statements were discarded (this happened with two participants in each age group). The rating criteria to correct free-recall task were strict. To consider a sentence as correct, a literal replication of the original statements was not required, but the preservation of the gist (i.e., the defining content of the sentence) or relevant details were needed, because it is known that memory is of a reconstructive nature (for example, “This coronavirus is a bioweapon created in the laboratory” was considered correct as the recall of the sentence “I believe that the coronavirus is part of a biological war”). Free recall was scored by two judges assigning one point for every correctly recalled sentence. The very few discrepancies were resolved by a third independent judge blind to the experimental conditions.

In order to analyze the thoughts recalled by the participants, a 3 (Group: Young vs. Middle-aged vs. Older) $\times$ 2 (Valence: Positive vs. Negative) $\times$ 2 (Nature: Personal vs. Social) ANOVA was conducted (see **Table 2**). The effects of the Group factor were significant, $F(2, 76) = 5.50$, $p = 0.006$, $\eta_p^2 = 0.013$. Although there were no differences between young and old adults, young people remembered a higher proportion of thoughts about the future than middle-aged adults, $t(55) = 3.21$, $p = 0.002$, $d = 0.43$.

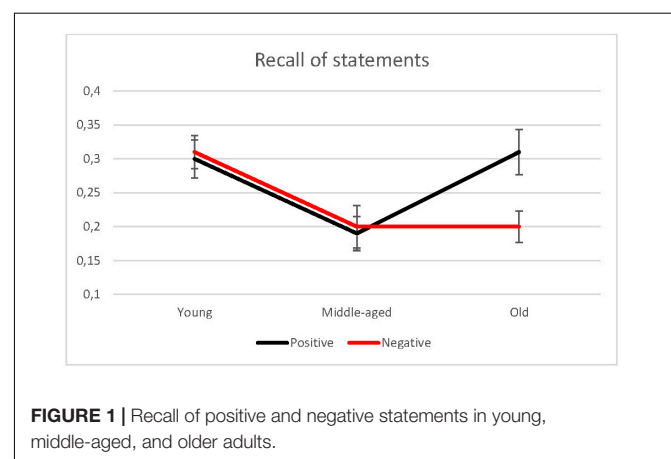
The variable Nature was non-significant, $F < 1$. There were no differences between the recall proportions of personal and social thoughts ($M = 0.24$, $SD = 0.15$ vs. $M = 0.27$, $SD = 0.16$) in the total sample.

TABLE 2 | Mean proportion (SD) of correct recall depending on their valence (positive and negative) and nature (personal, social), in the different age groups.

Group	Recall	Positive	Negative	Personal	Social
Young	0.30 (0.14)	0.30 (0.16)	0.31 (0.14)	0.29 (0.17)	0.31 (0.14)
Middle-aged	0.19 (0.11)	0.19 (0.12)	0.20 (0.15)	0.19 (0.14)	0.20 (0.13)
Older	0.25 (0.12)	0.31 (0.16)	0.20 (0.11)	0.23 (0.13)	0.28 (0.18)
Total	0.26 (0.13)	0.27 (0.16)	0.25 (0.15)	0.24 (0.15)	0.27 (0.16)

The Valence variable was also non-significant, $F(1, 76) = 3.51$, $p = 0.065$, $\eta_p^2 = 0.044$, revealing no statistically significant differences in the total sample between the recall rates of positive and negative thoughts ($M = 0.27$, $SD = 0.16$ vs. $M = 0.25$, $SD = 0.15$). However, the effects of the Group $\times$ Valence interaction were significant, $F(2, 76) = 6.11$, $p = 0.003$, $\eta_p^2 = 0.014$, indicating that, unlike the other two age groups, older people were influenced by the valence of thoughts, with a greater recall of positive thoughts ($M = 0.31$, $SD = 0.16$) than of negative ones ($M = 0.20$, $SD = 0.11$), $t(22) = 3.50$, $p = 0.002$ (see **Figure 1**). That is, as predicted by the positivity effect, older people showed a preference for positive content, recalling the same number or more of positive thoughts and the same number or fewer negative thoughts than young and middle-aged adults. In other words, for positive thoughts, middle-aged adults recalled a lower proportion than young adults, $t(54) = 2.79$, $p = 0.007$, $d = 0.38$, and older adults, $t(44) = 2.85$, $p = 0.007$, whereas, in the case of negative thoughts, young adults recalled significantly more thoughts than older people, $t(54) = 3.13$, $p = 0.003$, $d = 0.43$, and middle-aged adults, $t(54) = 2.83$, $p = 0.006$. There were no more significant interactions between the variables. In addition, correlational analyses between rating and recall of thoughts were conducted but the results were not significant, $r(79) = 0.04$, $p = 0.718$.

We also examined the clustering of positive and negative statements of the to be remembered material. Clustering refers to the tendency for items to take place next to one another in time. To quantify clustering we applied the Adjusted Ratio of Clustering (ARC; Roenker et al., 1971; Senkova and Otani, 2012), in which chance clustering is set at 0, perfect clustering

**FIGURE 1** | Recall of positive and negative statements in young, middle-aged, and older adults.

at 1, and negative scores indicate clustering below chance. We calculated this measure for both, the material administered in the encoding phase, and the final free recall performance. The ARC value was negative (-0.62) for the positive and negative statements presented to the participants, showing that statements were interleaved during the encoding phase. For the recall performance, ARC value in the total sample was also negative (-0.14), without statistically significant differences between young (-0.12), middle-aged (-0.10) and older adults (-0.25), $F(2, 65) = 0.60$, $p < 0.55$, $\eta_p^2 = 0.014$. Therefore, the participants in the 3 age-groups interleaved positive and negative thoughts in the final recall task much the same way as they were clustered within the material presented in the encoding phase. In addition, clustering of the positive or negative statements in final recall cannot explain the positivity bias, because the ARC scores were below zero for positive and negative statements. We also analyzed the number of repetitions of the positive and negative statements in the total sample in the final recall task. The only interesting result was that young participants ($M = 1.48$, $SD = 1.4$) were more likely to repeat negative statements than older adults ($M = 0.89$, $SD = 0.89$), although this was only a tendency, $t(46) = 1.77$, $p = 0.084$. The correlations between ARC and number of repetitions (both positive and negative statements) were not statistically significant.

Transformations and Biases

When recalling thoughts about the future, the participants sometimes modified their positive or negative valence. Analyzing these changes of valence or transformations allowed us to examine the memory biases during the personal elaboration and recovery of the previously presented material. Four types of transformations were classified: (1) initially negative statements transformed into neutral (2) or positive statements, (3) originally positive statements modified to neutral (4) or negative statements.

In order to analyze the transformations made by the participants, a 3 (Group: Young vs. Middle-aged vs. Older) $\times$ 2 (Nature: Personal vs. Social) ANOVA was conducted. The first two types of transformations (see Table 3) were considered a consequence of the positivity bias, because the participant positivized the statements by removing the negative connotation (e.g., *I talk a lot about this subject, I am obsessed and I do not cease looking for information* $\rightarrow$ *I have tried to obtain information about it. This will lead to a negative change at the social and political level* $\rightarrow$ *There will be political-social change*), or by transforming

the idea into something positive (e.g., *Mass concerts will not come back, nor will parties nor the great stadiums* $\rightarrow$ *I think partying is something that will soon come back. I think that the coronavirus is part of a biological warfare* $\rightarrow$ *I don't think the virus is an invention*).

In transformations from negative to neutral, the Group factor was significant, $F(1, 76) = 6.38$, $p = 0.003$; $\eta_p^2 = 0.144$. Without any differences between them, both older adults ($M = 0.15$, $SD = 0.24$) and middle-aged adults ($M = 0.07$, $SD = 0.17$) had a greater tendency to positivize initially negative phrases than did young adults, who did not produce any examples. The Nature variable also had significant effects, $F(1, 76) = 4.79$, $p = 0.032$; $\eta_p^2 = 0.059$, indicating that social statements ($M = 0.10$, $SD = 0.30$) were generally more positivized than personal ones ($M = 0.03$, $SD = 0.16$). The Group $\times$ Nature interaction was non-significant. Only the variable Nature, $F(1, 76) = 4.62$, $p = 0.035$; $\eta_p^2 = 0.057$, was significant in the negative to positive transformations because these changes were observed only for social statements (i.e., no examples of negative-to-positive transformations were observed for statements with a personal connotation). The Group $\times$ Nature interaction was non-significant.

The other two types of transformations, that is, initially positive thoughts that were transformed into neutral or negative ones, were very scarce (see Table 4). Participants transformed initially positive to neutral phrases (e.g., *I have good prospects for the future* $\rightarrow$ *This will influence my future. Although we will need time, we will travel again* $\rightarrow$ *The way we travel will change*) and they also negativized originally positive thoughts to negative ones (e.g., *The crisis is bringing out our best* $\rightarrow$ *I don't think we will be better people after the crisis. This confinement has allowed me to meet my neighbors and have new friends* $\rightarrow$ *I think this has not brought me any closer to my family or neighbors*).

In the first type of negativizations, there were no significant effects of the variables, whereas in the case of positive statements transformed into negative ones, only the variable nature was significant, $F(1, 76) = 3.84$, $p = 0.054$; $\eta_p^2 = 0.048$, revealing that these changes were observed to a greater extent for social statements than for statements with a personal connotation.

DISCUSSION

The central objective of this study was to examine the recall of positive and negative thoughts about the future linked to the global COVID-19 pandemic, and to examine the biases,

TABLE 3 | Mean proportion (SD) of valence changes from initially negative to neutral or positive statements, for personal and social claims in young, middle-aged, and older adults.

Group	From negative to neutral			From negative to positive		
	Change	Personal	Social	Change	Personal	Social
Young	0 (0)	0 (0)	0 (0)	0.09 (0.29)	0 (0)	0.18 (0.58)
Middle-aged	0.07 (0.17)	0 (0)	0.13 (0.34)	0.02 (0.10)	0 (0)	0.04 (0.21)
Older	0.15 (0.24)	0.09 (0.29)	0.22 (0.42)	0.04 (0.14)	0 (0)	0.09 (0.29)
Total	0.06 (0.17)	0.03 (0.16)	0.10 (0.30)	0.06 (0.21)	0 (0)	0.11 (0.42)

TABLE 4 | Mean proportion (SD) of changes in valence from initially positive to neutral or negative statements, for personal and social statements in young, middle-aged, and older adults.

Group	From positive to neutral			From positive to negative		
	Change	Personal	Social	Change	Personal	Social
Young	0.03 (0.12)	0.06 (0.24)	0 (0)	0.11 (0.30)	0.03 (0.17)	0.18 (0.58)
Middleage	0.04 (0.14)	0 (0)	0.09 (0.29)	0.04 (0.14)	0 (0)	0.09 (0.29)
Old	0.02 (0.10)	0 (0)	0.04 (0.21)	0.09 (0.19)	0.04 (0.21)	0.13 (0.34)
Total	0.03 (0.12)	0.03 (0.16)	0.04 (0.19)	0.08 (0.23)	0.03 (0.16)	0.14 (0.45)

errors, and distortions that occur in the recall of emotional information, especially positive biases in young, middle-aged, and older adults. Although the participants knew they were in an experiment on aspects related to COVID-19, they believed that their task was to rate their level of agreement and disagreement with the ideas and thoughts proposed about the future, and did not expect the task of recalling these contents. It is well known that incidental learning leads to worse performance than intentional learning, but it has also been observed that the effects of positivity are accentuated when participants are free to remember and are not subject to restrictions on how to organize their recall (Reed et al., 2014; García-Bajos et al., 2017). For example, positivity is not evident when the instructions request participants to encode the stimulus valence (Kensinger et al., 2002) or to accurately recall all the information (Grühn et al., 2005).

This study was carried out in the midst of the alarm state, when the population had already been confined for more than 2 weeks, and the streets were deserted and people could only go outside to acquire essential products. With the media and social networks full of bad news, the question arises as to how proposed hypothetical future events will be recalled. It is reasonable to think that people's thoughts and inferences will lead them to focus on the sources of fear implied by COVID-19 (Mathews et al., 1997; Booth and Sharma, 2020) producing a mood-congruent retrieval, that is, focusing recall on events associated with a negative emotional state. There is actually a good deal of evidence for this "mood-congruency" effect for a variety of cognitive processes, including attention and perception, judgment, and various types of recall and recognition procedures (for reviews, see Blaney, 1986; Siemer, 2005; Koster et al., 2010; Sasa, 2013). However, the theory of mood congruence is not met because, even if there was fear of contagion and the unknown consequences of COVID-19 at that particular moment of confinement, there was also expectation, novelty, new activities that were being incorporated into people's routine and thoughts about the positive aspects that this pandemic could imply (e.g., becoming more humane, more supportive or empathetic, uniting more as a society in the face of adversity, and/or improving awareness of the environment). Our data show a phenomenon of positivity. On the one hand, in the subjective ratings of the contents where the participants, regardless of age, agreed more with the positive than the negative content, and, on the other hand, in the recall task, where, despite the health crisis, they retrieved positive and negative thoughts to the same extent. Both effects may reflect a

mood-regulation strategy. Enhancing the idea of positivizing the situation, the participants even biased their recall by transforming content initially presented as negative into more neutral or positive content.

Also in this study, we observed that the positivity phenomenon is enhanced in older people, giving rise to the positivity effect observed in many previous studies (e.g., Carstensen and DeLiema, 2018; Gallo et al., 2011; García-Bajos et al., 2017). Although globally, there are no differences between the recall of positive and negative facts, the interaction between the main variables Group and emotional Valence shows that older people have a greater recall of positive imagined future events and greater resistance to recalling negative facts (Charles et al., 2003), whereas young people present a better performance than the older and middle-aged people in the recall of negative content. At the same time, older and middle-aged adults tend to transform negative statements into neutral ones, eliminating terms that give a negative connotation to the idea or changing the statements to more pleasant possibilities. In other words, older people distance themselves from the negative possibilities and consequences of the future, thinking about it in a more generic, less specific, or more semantic way (Devitt and Schacter, 2020), thus achieving the goal of regulating mood by decreasing negative emotions and increasing positive ones (Rusting and DeHart, 2000). Similar results have been observed for past choices and autobiographical information, where older people show more emotionally gratifying memory distortion than young adults (Mather and Carstensen, 2005). Our data further indicate that this positive bias is also characteristic of middle-aged people, an age group little analyzed in the literature on cognition and memory in general and on the positivity effect in particular.

To what is this positivity effect due? Various theories have been proposed to explain the effect of positivity. Some theories posit that age-related advantages reflect the avoidance of stressors (Charles, 2010) but this is difficult to accept in the current pandemic situation and its effects. Our data are more consistent with the Socio-emotional Selectivity Theory (SST; Carstensen et al., 1999; see also Carstensen and Mikels, 2005; Carstensen et al., 2006; Reed and Carstensen, 2012). The SST is a life-span theory of motivation, which proposes that, because of the realization that the time left to live is growing shorter, older adults are more likely to prioritize their balance and emotional well-being. Priorities change with age, and a preference for the positive emerges. The SST posits that older adults deploy cognitive control mechanisms to suppress negative stimuli and to seek out positive, emotionally rewarding information. Although

at the cognitive level, older people generally show deficits in resource availability, they use their resources to enhance emotion regulation, perhaps using their limited resources to block or inhibit negative thoughts and activate positive ones (Giebl et al., 2016; García-Bajos and Migueles, 2017; García-Bajos et al., 2017; Marsh et al., 2019). Although not as accentuated as in older people, the fact that middle-aged participants also show the positivity effect suggests that the effect is not due to a malfunction of the amygdala that reduces neural and affective responses to negative stimuli (Reed and Carstensen, 2012) or to the fact that the processing of negative content is more complex and cognitively more demanding (Labouvie-Vief et al., 2010). The results rather suggest that cognitive abilities and motivation both contribute to the outcomes obtained from improved emotion regulation as people get older (Mather and Carstensen, 2005).

An interesting aspect is the performance of the group of young people, university students concerned about their future training and work. Although the real impact of COVID-19 on students' education and mental health is still unknown (Araújo et al., 2020; Odriozola-González et al., 2020a; Sahu, 2020), psychological symptoms are common in the university population (Bayram and Bilgel, 2008; Auerbach et al., 2016). A study conducted by Odriozola-González et al. (2020a) analyzed the impact of COVID-19 during the first weeks of confinement in Spanish university students. They applied an online survey to 2530 students and observed that moderate to extremely severe scores of anxiety (21.34%), depression (34.19%), and stress (28.14%), and a total of 50.43% of the respondents presented a moderate to severe impact of the outbreak. In our study, although the young students agreed more with the positive than the negative statements, they showed no positivity effect. In their performance, there were no differences between the recall of positive and negative content, but they recalled significantly more negative content than older and middle-aged adults. In other words, they showed less resilience, which reflected their concern about the potentially negative impact on their academic progress.

Finally, concerning the effects of the social or personal nature of the thoughts about the future, our results have shown that, although young people provided similar ratings for both types of content, older and middle-aged participants rated social thoughts as closer to their way of thinking than personal thoughts, an effect we could call sociability. As for memory, contrary to our expectations of a trend of a greater recall of social than of personal content because it is easier to imagine and recreate the future in third-person than in the first-person (D'Argembeau and Van der Linden, 2004, 2006), there were no differences in recall between the two types of content, which was equivalent in the participants of all age groups. Although young people recruit effective encoding strategies to remember a large amount of information regardless of its nature, older people perceived and categorized the emotional content of a social nature as more relevant. It has been observed that emotional content rated as important reduces the differences between the recall of young and old people (Denburg et al., 2003; Mather, 2004; Spaniol et al., 2008). In general, social statements were more positivized than personal ones, that is, the positivity bias was greater for statements about society as a whole than for statements about

particular individuals, and it would be relevant to analyze these aspects concerning recall and false memories in the future.

Although this study has the limitations of having been carried out online, which leads to a reduction of situational control, it has the advantage of immediacy and of being able to rigorously examine the recall of content and thoughts about the future after a pandemic that is changing the world. Future research should examine whether executive functions also influence the accuracy of recalling content about the future. It would also be interesting to analyze individual differences (especially in the older group) in positivity bias because there may be great variability, as with other types of false memories (Gerrie and Garry, 2007; St Jacques et al., 2015; Greene and Murphy, 2020). The investigation of the particularities of cognitive functioning and memory of middle-aged adults is relevant from a cognitive research perspective.

In short, this study increases our understanding not only of the impact of aging on the memory of imagined positive and negative future events and their transformations and modifications but also middle-aged adults' recall of episodic future events and false memories. The errors, distortions, and transformations observed in this study do not have the numerical scope of the errors observed with the DRM paradigm (Roediger and McDermott, 1995), where a set of associated words (e.g., pin, puncture, pain, syringe) induce the recall of an unrepresented word (needle), or the applied impact in the judicial sphere such as the post-event information procedure, where suggested information is introduced in the recall (Loftus, 1991, 2003); however, they show the mind's ability to spontaneously transform content to make it kinder, more positive, and to help people to adapt to adversity, reduce anxiety and depression, foster resilience, and contribute to feelings of well-being. The positivity bias shows the adaptive value of the memory (e.g., Schacter, 1999, 2021; Schacter et al., 2011), which does not disappear and may even be enhanced in crises, such as the one we are currently experiencing.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

This study was carried out in accordance with the American Psychological Association standards for ethical treatment of participants, the Declaration of Helsinki, and was approved by the Ethics Committee of the University of the Basque Country UPV/EHU (Ref. M10_2016_052). Participants were informed that the experiment dealt with the positive and negative nature of thoughts about the future linked to the COVID-19 pandemic. Participation was voluntary, and consent was presumed by the completion of the survey. Written informed consent was not provided because this experiment was conducted online through Google forms, because the alarm status had been decreed in Spain, and the entire population was confined without the possibility of leaving their homes.

AUTHOR CONTRIBUTIONS

MM and AAI conceived, designed, prepared the materials for the experiment, performed the experiment, collected the data, scored the tasks, and wrote the manuscript. AAI analyzed the data. AAR helped re-analyzing the data and reviewed the final version of the manuscript. All authors critically reviewed the manuscript for important intellectual content, approved the manuscript for publication, and agreed to be accountable for all aspects of the

work in ensuring that issues related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the project of this research.

ACKNOWLEDGMENTS

We thank all the people who generously dedicated their time to participate in this study.

REFERENCES

- Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45, 1363–1377. doi: 10.1016/j.neuropsychologia.2006.10.016
- Araújo, F., Lima, L., Cidade, P., Nobre, C., and Neto, M. (2020). Impact of Sars-Cov-2 and its reverberation in global higher education and mental health. *Psychiatry Res.* 288:112977. doi: 10.1016/j.psychres.2020.112977
- Armitage, R., and Nellums, L. B. (2020). COVID-19 and the consequences of isolating the elderly. *Lancet Public Health* 5:e256. doi: 10.1016/S2468-2667(20)30061-X
- Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., et al. (2016). Mental disorders among college students in the World Health Organization world mental health surveys. *Psychol. Med.* 46, 2955–2970. doi: 10.1017/S0033291716001665
- Balluerka, N., Gómez Benito, J., Hidalgo, M. D., Gorostiaga, A., Espada, J. P., Padilla, J. L., et al. (2020). *Las consecuencias psicológicas de la Covid-19 y el confinamiento*. Ikertzaileak: Nekane Balluerka (UPV/EHU).
- Bayram, N., and Bilgel, N. (2008). The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Soc. Psychiatry.Epidemiol.* 43, 667–672. doi: 10.1007/s00127-008-0345-x
- Benke, C., Autenrieth, L. K., Asselmann, E., and Pané-Farré, C. A. (2020). Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany. *Psychiatry Res.* 293:113462. doi: 10.1016/j.psychres.2020.113462
- Berntsen, D., and Bohn, A. (2010). Remembering and forecasting: The relation. *Memory Cogn.* 38, 265–278. doi: 10.3758/MC.38.3.265
- Berntsen, D., and Jacobsen, A. S. (2008). Involuntary (spontaneous) mental time travel into the past and future. *Conscious. Cogn.* 17, 1093–1104. doi: 10.1016/j.concog.2008.03.001
- Berntsen, D., and Rubin, D. C. (2002). Emotionally charged autobiographical memories across the life span: the recall of happy, sad, traumatic, and involuntary memories. *Psychol. Aging* 17, 636–652. doi: 10.1037/0882-7974.17.4.636
- Blaney, P. H. (1986). Affect and memory: A review. *Psycholog. Bull.* 99, 229–246. doi: 10.1037/0033-2909.99.2.229
- Booth, R. W., and Sharma, D. (2020). Attentional control and estimation of the probability of positive and negative events. *Cogn. Emot.* 34, 553–567. doi: 10.1080/02699931.2019.1657382
- Botzung, A., Denkova, E., and Manning, L. (2008). Experiencing past and future personal events: functional neuroimaging evidence on the neural bases of mental time travel. *Brain Cogn.* 66, 202–212. doi: 10.1016/j.bandc.2007.07.011
- Burr, D. A., Castrellon, J. J., Zald, D. H., and Samanez-Larkin, G. R. (2020). Emotion dynamics across adulthood in everyday life: Older adults are more emotionally stable and better at regulating desires. *Emotion* 20:734. doi: 10.1037/emo0000734
- Carstensen, L. L. (2006). The influence of a sense of time on human development. *Science* 312, 1913–1915. doi: 10.1126/science.1127488
- Carstensen, L. L., Isaacowitz, D., and Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *Am. Psychol.* 54, 165–181. doi: 10.1037/0003-066x.54.3.165
- Carstensen, L. L., and Mikels, J. A. (2005). At the intersection of emotion and cognition aging and the positivity effect. *Curr. Direct. Psychol. Sci.* 14, 117–121. doi: 10.1111/j.0963-7
- Carstensen, L. L., Mikels, J. A., and Mather, M. (2006). “Fifteen - Aging and the Intersection of Cognition, Motivation, and Emotion,” in *Handbook of the Psychology of Aging* (Sixth Edition, eds J. E. Birren, K. W. Schaie, R. P. Abeles, M. Gatz, and T. A. Salthouse (Cambridge, MA: Academic Press.), 343–362. doi: 10.1016/B978-012101264-9/50018-5
- Carstensen, L. L., and DeLiema, M. (2018). The positivity effect: a negativity bias in youth fades with age. *Curr. Opin. Behav. Sci.* 19, 7–12. doi: 10.1016/j.cobeha.2017.07.009
- Carstensen, L. L., Pasupathi, M., Mayr, U., and Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *J. Personal. Soc. Psychol.* 79, 644–655. doi: 10.1037/0022-3514.79.4.644
- Carstensen, L. L., Shavit, Y. Z., and Barnes, J. T. (2020). Age advantages in emotional experience persist even under threat from the COVID-19 pandemic. *Psychol. Sci.* 31, 1374–1385. doi: 10.1177/0956797620967261
- Charles, S. T. (2010). Strength and vulnerability integration: A model of emotional well-being across adulthood. *Psychol. Bull.* 136, 1068–1091. doi: 10.1037/a001232
- Charles, S. T., Mather, M., and Carstensen, L. L. (2003). Aging and emotional memory: the forgettable nature of negative images for older adults. *J. Exp. Psychol. Gen.* 132, 310–324. doi: 10.1037/0096-3445.132.2.310
- Cole, S. N., Staugaard, S. R., and Berntsen, D. (2016). Inducing involuntary and voluntary mental time travel using a laboratory paradigm. *Memory Cogn.* 44, 376–389. doi: 10.3758/s13421-015-0564-9
- D'Argembeau, A., and Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Conscious. Cogn.* 13, 844–858. doi: 10.1016/j.concog.2004.07.007
- D'Argembeau, A., and Van der Linden, M. (2006). Individual differences in the phenomenology of mental time travel: The effect of vivid imagery and emotion regulation. *Conscious. Cogn.* 15, 342–350. doi: 10.1016/j.concog.2005.09.001
- Denburg, N. L., Buchanan, T. W., Tranel, D., and Adolphs, R. (2003). Evidence for preserved emotional memory in normal older persons. *Emotion* 3, 239–253. doi: 10.1037/1528-3542.3.3.239
- Devitt, A. L., and Schacter, D. L. (2020). Looking on the bright side: Aging and the impact of emotional future simulation on subsequent memory. *J. Gerontol. B* 75, 1831–1840. doi: 10.1093/geronb/gbz041
- Faul, F., Erdfelder, E., and Buchner, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4
- Finnbogadottir, H., and Berntsen, D. (2013). Involuntary future projections are as frequent as involuntary memories, but more positive. *Conscious. Cogn.* 22, 272–280. doi: 10.1016/j.concog.2012.06.014
- Fontes, A. P., and Neri, A. L. (2015). Resilience in aging: literature review. *Ciência Saúde Coletiva* 20, 1475–1495. doi: 10.1590/1413-81232015205.00502014
- Gallo, D. A., Korthauer, L. E., McDonough, I. M., Teshale, S., and Johnson, E. L. (2011). Age-related positivity effects and autobiographical memory detail: evidence from a past/future source memory task. *Memory* 19, 641–652. doi: 10.1080/09658211.2011.595723
- García-Bajos, E., and Migueles, M. (2017). Retrieval of past and future positive and negative autobiographical experiences. *Cogn.Emot.* 31, 1260–1267. doi: 10.1080/02699931.2016.1204988

- García-Bajos, E., Migueles, M., and Aizpurua, A. (2017). Age-based positivity effects in imagining and recalling future positive and negative autobiographical events. *Front. Psychol.* 8:1700. doi: 10.3389/fpsyg.2017.01700
- García-Portilla, P., De la Fuente, L., Bobes-Bascarán, T., Jiménez, T. L., Zurrón, P., Suárez, Á.M., et al. (2020). Are older adults also at higher psychological risk from COVID-19? *Aging Mental Health* 2020:1805723. doi: 10.1080/13607863.2020.1805723
- Gerrie, M. P., and Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory* 15, 561–571. doi: 10.1080/09658210701391634
- Giebl, S., Storm, B. C., Buchli, D. R., Bjork, E. L., and Bjork, R. A. (2016). Retrieval-induced forgetting is associated with increased positivity when imagining the future. *Q. J. Exp. Psychol.* 69, 351–360. doi: 10.1080/17470218.2015.1085586
- Gotlib, I. H., and Joormann, J. (2010). Cognition and depression: current status and future directions. *Ann. Rev. Clin. Psychol.* 6, 285–312. doi: 10.1146/annurev.clinpsy.121208.131305
- Greene, C. M., and Murphy, G. (2020). Individual differences in susceptibility to false memories for COVID-19 fake news. *Cogn. Res.* 5:262. doi: 10.1186/s41235-020-00262-1
- Griffiths, M. D., and Mamun, M. A. (2020). COVID-19 suicidal behavior among couples and suicide pacts: Case study evidence from press reports. *Psychiatry Res.* 289:113105. doi: 10.1016/j.psychres.2020.113105
- Grühn, D., Smith, J., and Baltes, P. B. (2005). No aging bias favoring memory for positive material: evidence from a heterogeneity-homogeneity list paradigm using emotionally toned words. *Psychol. Aging* 20, 579–588. doi: 10.1037/0882-7974.20.4.579
- Hamilton, L. J., and Allard, E. S. (2020). Words matter: age-related positivity in episodic memory for abstract but not concrete words. *Aging Neuropsychol. Cogn.* 27, 595–616. doi: 10.1080/13825585.2019.1657556
- Hargis, M. B., and Castel, A. D. (2017). Younger and older adults' associative memory for social information: The role of information importance. *Psychol. Aging* 32, 325–330. doi: 10.1037/pag0000171
- Kennedy, Q., Mather, M., and Carstensen, L. L. (2004). The role of motivation in the age-related positivity effect in autobiographical memory. *Psychol. Sci.* 15, 208–214. doi: 10.1111/j.0956-7976.2004.01503011.x
- Kensinger, E. A. (2008). Age differences in memory for arousing and nonarousing emotional words. *J. Gerontol. B: Psychol. Sci. Soc. Sci.* 63B, 13–18. doi: 10.1093/geronb/63.1.p13
- Kensinger, E. A., Brierley, B., Medford, N., Growdon, J. H., and Corkin, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion* 2, 118–134. doi: 10.1037/1528-3542.2.2.118
- Killgore, W. D. S., Cloonan, S. A., Taylor, E. C., and Dailey, N. S. (2020). Loneliness: A signature mental health concern in the era of COVID-19. *Psychiatry Res.* 290, 113–117. doi: 10.1016/j.psychres.2020.113117
- Koster, E. H. W., De Raedt, R., Leyman, L., and De Lissnyder, E. (2010). Mood-congruent attention and memory bias in dysphoria: Exploring the coherence among information-processing biases. *Behav. Res. Ther.* 48, 219–225. doi: 10.1016/j.brat.2009.11.004
- Kvavilashvili, L., and Rummel, J. (2020). On the nature of everyday prospection: A review and theoretical integration of research on mind-wandering, future thinking, and prospective memory. *Rev. Gen. Psychol.* 24, 210–237. doi: 10.1177/1089268020918843
- Labouvie-Vief, G., Grühn, D., and Studer, J. (2010). "Dynamic Integration of Emotion and Cognition: Equilibrium Regulation in Development and Aging," in *The Handbook of Life-Span Development*, eds R. M. Lerner, M. E. Lamb, and A. M. Freund (Hoboken, NJ: Wiley), doi: 10.1002/9780470880166.hlsd002004
- Loftus, E. F. (1991). "Made in memory: Distortions in recollection after misleading information," in *Psychology of Learning and Motivation*, Vol. 27, ed. G. H. Bower (Cambridge, MA: Academic Press), 187–215. doi: 10.1016/S0079-7421(08)60124-3
- Loftus, E. F. (2003). Our changeable memories: legal and practical implications. *Nat. Rev. Neurosci.* 4, 231–234. doi: 10.1038/nrn1054
- Mammarella, N., Di Domenico, A., Palumbo, R., and Fairfield, B. (2016). When green is positive and red is negative: aging and the influence of color on emotional memories. *Psychol. Aging* 31, 914–926. doi: 10.1037/pag000012
- Marsh, L., Edgington, T., Conway, M. A., and Loveday, C. (2019). Positivity bias in past and future episodic thinking: Relationship with anxiety, depression, and retrieval induced forgetting. *Q. J. Exp. Psychol.* 72, 508–522. doi: 10.1177/1747021818758620
- Mather, M. (2004). "Aging and emotional memory," in *Series in affective science. Memory and emotion*, eds D. Reisberg and P. Hertel (Oxford: Oxford University Press), 272–307. doi: 10.1093/acprof:oso/9780195158564.003.0009
- Mather, M., and Carstensen, L. L. (2005). Aging and motivated cognition: the positivity effect in attention and memory. *Trends Cogn. Sci.* 9, 496–502. doi: 10.1016/j.tics.2005.08.005
- Mathews, A., Mackintosh, B., and Fulcher, E. P. (1997). Cognitive biases in anxiety and attention to threat. *Trends Cogn. Sci.* 1, 340–345. doi: 10.1016/S1364-6613(97)01092-9
- Newby-Clark, I. R., and Ross, M. (2003). Conceiving the Past and Future. *Personal. Soc. Psychol. Bull.* 29, 807–818. doi: 10.1177/0146167203029007001
- Odrizola-González, P., Planchuelo-Gomez, A., Iruñia, M. J., and De Luis-García, R. (2020a). Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry Res.* 290:113108. doi: 10.1016/j.psychres.2020.113108
- Odrizola-González, P., Planchuelo-Gomez, A., Iruñia-Muniz, M. J., and De Luis-García, R. (2020b). Psychological symptoms of the outbreak of the COVID-19 crisis and confinement in the population of Spain. *PsyArXiv* 2020:4. doi: 10.31234/osf.io/mq4fg
- Reed, A., and Carstensen, L. (2012). The theory behind the age-related positivity effect. *Front. Psychol.* 3:339. doi: 10.3389/fpsyg.2012.00339
- Reed, A. E., Chan, L., and Mikels, J. A. (2014). Meta-analysis of the age-related positivity effect: age differences in preferences for positive over negative information. *Psychol. Aging* 29, 1–15. doi: 10.1037/a0035194
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *J. Exp. Psychol.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Rodríguez-Rey, R., Garrido-Hernansaiz, H., and Collado, S. (2020). Psychological Impact of COVID-19 in Spain: Early Data Report. *Psychol. Trauma* 12, 550–552. doi: 10.1037/tra0000943
- Roenker, D., Thompson, C., and Brown, S. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychol. Bull.* 76, 45–48. doi: 10.1037/h0031355
- Rusting, C. L., and DeHart, T. (2000). Retrieving positive memories to regulate negative mood: Consequences for mood-congruent memory. *J. Personal. Soc. Psychol.* 78, 737–752. doi: 10.1037/0022-3514.78.4.737
- Sahu, P. (2020). Closure of universities due to coronavirus disease 2019 (COVID-19): impact on education and mental health of students and academic staff. *Cureus* 12:e7541. doi: 10.7759/cureus.7541
- Senkova, O., and Otani, H. (2012). Category clustering calculator for free recall. *Adv. Cogn. Psychol.* 8, 292–295. doi: 10.2478/v10053-008-0124-y
- Santini, Z. I., Jose, P. E., Cornwell, E. Y., Koyanagi, A., Nielsen, A., Hinrichsen, C., et al. (2020). Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. *Lancet Public Health* 5, e62–e70. doi: 10.1016/S2468-2667(19)30230-0
- Sasa, D. (2013). Evidence for the role of affect in mood congruent recall of autobiographic memories. *Motiv. Emot.* 37, 623–628. doi: 10.1007/s11031-012-9322-5
- Schacter, D. L. (2021). The seven sins of memory: an update. *Memory* 2021:1873391. doi: 10.1080/09658211.2021.1873391
- Schacter, D. L. (1999). The seven sins of memory: insights from psychology and cognitive neuroscience. *Am. Psychol.* 54, 182–203. doi: 10.1037/0003-066X.54.3.182
- Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* 8, 657–661. doi: 10.1038/nrn2213
- Schacter, D. L., Addis, D. R., and Buckner, R. L. (2008). "Episodic simulation of future events: Concepts, data, and applications," in *Annals of the New York Academy of Sciences: Vol. 1124. The year in cognitive neuroscience*, eds A. Kingstone and M. B. Miller (Hoboken, NJ: Blackwell Publishing), 39–60. doi: 10.1196/annals.1440.001
- Schacter, D. L., Guerin, S. A., and Jacques, P. (2011). Memory distortion: an adaptive perspective. *Trends Cogn. Sci.* 15, 467–474. doi: 10.1016/j.tics.2011.08.004

- Schacter, D. L., Koutstaal, W., and Norman, K. A. (1997). False memories and aging. *Trends Cogn. Sci.* 1, 229–236. doi: 10.1016/S1364-6613(97)01068-1
- Schryer, E., and Ross, M. (2014). Does the age-related positivity effect in autobiographical recall reflect differences in appraisal or memory? *J. Gerontol. B: Psychol. Sci. Soc. Sci.* 69, 548–556. doi: 10.1093/geronb/gbt047
- Siemer, M. (2005). Mood-congruent cognitions constitute mood experience. *Emotion* 5, 296–308. doi: 10.1037/1528-3542.5.3.296
- Silva Junior, E. G., Eulalio, M. D. C., Queiroga Souto, R., De Lima Santos, K., Melo, R. L. P. D., and Lacerda, A. R. (2019). The capacity for resilience and social support in the urban elderly. *Ciencia Saude Coletiva* 24, 7–16. doi: 10.1590/1413-81232018241.32722016
- Spaniol, J., Voss, A., and Grady, C. L. (2008). Aging and emotional memory: Cognitive mechanisms underlying the positivity effect. *Psychol. Aging* 23, 859–872. doi: 10.1037/a0014218
- St Jacques, P. L., Montgomery, D., and Schacter, D. L. (2015). Modifying memory for a museum tour in older adults: Reactivation-related updating that enhances and distorts memory is reduced in ageing. *Memory* 23, 876–887. doi: 10.1080/09658211.2014.933241
- Stone, A. A., Schwartz, J. E., Broderick, J. E., and Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proc. Natl. Acad. Sci.* 107, 9985–9990. doi: 10.1073/pnas.1003744107
- Szpunar, K. K. (2010). Episodic future thought: an emerging concept. *Perspect. Psychol. Sci.* 5, 142–162. doi: 10.1177/1745691610362350
- Thakur, V., and Jain, A. (2020). COVID 2019-Suicides: A global psychological pandemic. *Brain Behavior Immun.* 88, 952–953. doi: 10.1016/j.bbi.2020.04.062
- Tulving, E. (1985). Memory and consciousness. *Can. J. Psychol.* 26, 1–12. doi: 10.1037/h0080017
- Tulving, E. (2005). “Episodic Memory and Autonoesis: Uniquely Human?,” in *The missing link in cognition: Origins of self-reflective consciousness*, eds H. S. Terrace and J. Metcalfe (Oxford: Oxford University Press), 3–56. doi: 10.1093/acprof:oso/9780195161564.003.0001
- Vindegard, N., and Benros, M. E. (2020). COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. *Brain Behav. Immun.* 89, 531–542. doi: 10.1016/j.bbi.2020.05.048
- Xiong, J., Lipsitz, O., Nasri, F., Lui, M. M. W., Gill, H., Phan, L., et al. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *J. Affect. Dis.* 277, 55–64. doi: 10.1016/j.jad.2020.08.001
- Zaragoza Scherman, A., Salgado, S., Shao, Z., and Berntsen, D. (2020). Younger adults report more distress and less well-being: A cross-cultural study of event centrality, depression, post-traumatic stress disorder and life satisfaction. *Appl. Cogn. Psychol.* 34, 1180–1196. doi: 10.1002/acp.3707
- Zhai, J., and Du, X. (2020). Addressing collegiate mental health amid COVID-19 pandemic. *Psychiatr. Res.* 288:113003. doi: 10.1016/j.psychres.2020.113003

Conflict of Interest: 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 © 2021 Aizpurua, Migueles and Aranberri. 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) and the copyright owner(s) 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.



Aging Does Not Enhance Social Contagion Effect

Susana Carnero-Sierra*† and Julio Menor†

Área Psicología Básica, Psychology Department, University of Oviedo, Oviedo, Spain

OPEN ACCESS

Edited by:

Sara Cadavid,
Rosario University, Colombia

Reviewed by:

Michelle Meade,
Montana State University,
United States
Ines Bramao,
Lund University, Sweden

*Correspondence:

Susana Carnero-Sierra
carnerosusana@uniovi.es

† These authors have contributed
equally to this work

Specialty section:

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

Received: 05 April 2021

Accepted: 19 July 2021

Published: 04 August 2021

Citation:

Carnero-Sierra S and Menor J
(2021) Aging Does Not Enhance
Social Contagion Effect.
Front. Psychol. 12:691229.
doi: 10.3389/fpsyg.2021.691229

Studies on the social contagion of memory show that it is possible to create false memories from the wrong responses from other people without requiring their physical presence. The current study examined age differences between false memories via the modified social contagion paradigm. Twenty older and twenty younger adults were shown six household scenes and were exposed to the erroneous memory reports of an implied confederate who was not physically present. The presentation time of the scenes and the typicality of the contagion items were manipulated. The participants watched each scene individually and then took turns giving their recall responses with the responses belonging to a fictional participant provided by written cards. The results in a final individual recall test indicated a significant contagion effect in both groups of participants. Additionally, an effect of the typicality of the contagion items was observed, such that the more typical items produced more contagion than the less typical items. In relation to true recall, the older adults remembered significantly fewer items from the scenes than the younger ones and obtained a lower score in the word list subtest of the Wechsler Memory Scale. Although the older group had an episodic memory deficit, they were not more susceptible to being affected by the wrong responses of other people than younger group.

Keywords: false memory, social contagion, memory illusions, aging, false recall

INTRODUCTION

The process of remembering occurs, in many cases, in a social context since one of the purposes of the act of remembering is to transmit information to other people. As different studies have shown (Roediger et al., 2001; Harris et al., 2017), during this process the person who shares the memory can transmit wrong information. The phenomenon involved in this situation is called social contagion of memory. Most of the studies on the social contagion of memory have focused on young adults, but there are less studies that examine how age influences the susceptibility to wrong information socially introduced (Meade et al., 2020).

Regarding the formation of false individual memories and aging, it has been consistently observed that older participants are more susceptible to accepting misinformation (see Schacter et al., 1997, who reviewed the source of memory failures in induced false memory procedures often found in memory tasks in older participants). Additionally, Devitt and Schacter (2016) concluded that several neurological changes related to false memories are associated with aging, increasing susceptibility to false memories. Specifically, Roediger and Geraci (2007) found that older adults were more susceptible to being affected by misinformation interference in the eyewitness paradigm from Loftus and Palmer (1974). Roediger and McDaniel (2007) reviewed the results of four

additional experimental paradigms: the DRM paradigm (Roediger and McDermott, 1995), the categorized word list (Meade and Roediger, 2006), the misinformation paradigm (Loftus and Palmer, 1974), and the imagination inflation paradigm (Goff and Roediger, 1998), ultimately concluding that older adults are more susceptible to false memories than young adults.

However, differences in social false memory paradigms between younger and older participants are a matter of controversy. Meade and Roediger (2009) used a turn-taking collaborative procedure from the recall list of categorized words, describing that older participants have more possibilities to suffer from false recall than younger participants in the final recall test. However, Henkel and Rajaram (2011) observed no age differences between younger and older adults using a free-flowing procedure for collaborating. Gabbert et al. (2003) employed a video and a suggestive questionnaire to induce false memories in younger and older participants. They found no age differences in false memories, although it was observed that false recall in a posterior individual recall is increased when the possibility of prior discussion with other participants is given. The authors understood it as a result of the memory conformity effect. In a subsequent article, Gabbert et al. (2004) went on in depth about a procedure to determine that, in both younger and older, a social chatting with a same-age confederate is more effective for inducing false memories than a biased written narrative. Interestingly, results shown that older were even less susceptible than younger to commit false recall. The third study that explicitly tested age differences in social contagion of memory was carried out by Davis and Meade (2013), who employed the Roediger et al. (2001) paradigm to test false memories in both older and younger participants. No age differences were found, although an effect of confederate age was observed. In summary, the studies carried out so far indicate that the effect of social memory contagion is similar in younger and older adults.

Meade et al. (2020) suggests that social contagion protocols with physical confederates could minimize source monitoring errors which more likely occur in older than in younger. Expressions, emotions, and distinctiveness of recall of having a physical partner could enhance the distinctiveness of erroneous items recalled by the confederate from items actually presented. If this explanation is correct, then providing the wrong answers in written protocols would lead to reducing the distinctiveness of the source and increasing monitoring errors. Since source monitoring errors occur more frequently in older than in younger, the older adults would commit more false memory than the younger adults.

The effect of social contagion on memory induced by a written protocol, that simulates the responses of other people, has already been investigated in some studies. Roediger et al. (2001) established a procedure to induce false memories through a protocol of social contagion. Later, Meade and Roediger (2002) made a substantial variation in the procedure. In original experiments (Roediger et al., 2001) a second experimenter was used, pretending to be a participant who performed the required tasks alongside the real participant. Instead, Meade and Roediger (2002) replaced the false participant with a written protocol

provided by the researcher. The cards presented to the real participant transmitted the same information as the confederate of the previous procedure. In this way, it was explained that responses written on the cards consisted of identical responses given by a subject previously involved in the same task. Thereby, the social component of the experiment was not eliminated. Results of Meade and Roediger (2002) indicated that this modified version of the procedure was just as strong at inducing false memories as the original procedure. Menor and Carnero (2013) confirmed that there were no differences between this non-face-to-face mode of contagion and the face-to-face one, observing similar contagion rates in both conditions from a young adult sample.

The present study aimed to analyze age differences in social contagion of memory by using a virtual confederate. It extends upon previous research of age differences in social contagion of memory, by using, as source of contagion, the virtual paradigm without physical presence of the confederate. The absence of a face, a voice, and other social factors involved in recalling with another person, would decrease source monitoring, which is especially problematic for older participants (Mitchell et al., 2003; Devitt and Schacter, 2016). Two stimulus presentation times (15 and 60 s) were used, and high expectative (objects expected to be in a scene) and low expectative (non-frequent objects about the thematic of the scenes) contagion items were introduced by virtual confederate. In relation to the presentation time of the scenes, it has been found that when the contagion items are introduced by a physical confederate, a shorter presentation time of the scenes increases the false memory in young adults. Thus, Roediger et al. (2001) and Menor and Carnero (2013) found that 15-s rates increased false memory compared to 60-s rates. However, using a written protocol as a source of contagion, Menor and Carnero (2013) did not observe differences between 15 and 60 s. It is possible that the older group was more affected by the shorter presentation time due to their episodic memory deficit, so that older adults would rely more on the responses of the written protocol when they have to remember the scenes.

The results obtained in older adults were compared with those obtained by the group of young adults in Menor and Carnero (2013). Due to a decreased source distinctiveness through the absence of a physical confederate, it was expected to find more social contagion in older than young adults. Additionally, it is expected to find an interaction between the presentation rate and the age group in such a way that the 15-s rate would cause greater social contagion than the 60-s rate in the group of older adults, but not in the young adults group. Lastly, the high-expectation contagion items would cause greater social contagion than the low-expectation items in both groups of participants as observed in other studies (Roediger et al., 2001; Meade and Roediger, 2002).

MATERIALS AND METHODS

Participants

Participants consisted of twenty older adults from municipal social centers for seniors. Seventeen women and three men

between the ages of 60 and 88 ($\bar{x} = 72.37$, $SD = 8.57$) participated, with no specific diagnosed pathology. Older participants scored within the clinically normal range on the MMSE (Folstein et al., 1975; Spanish version: Blesa et al., 2001), $\bar{x} = 27.25$, $SD = 1.68$. The Geriatric Depression Scale (Yesavage et al., 1982) revealed no signs of depression ($\bar{x} = 4.93$, $SD = 3.59$). Furthermore, the mean score on the vocabulary subtest of WAIS (Wechsler, 1999) was 40.93, $SD = 6.36$. Performance in episodic memory was tested using the word list subtest of the Wechsler Memory Scale-III (Wechsler, 2004). The results are shown later in the results section. Older participants' results were then compared with a sample of younger adults (20 undergraduate students between 21 and 34 years old, thirteen women and seven men) that experienced the same procedure of social contagion through a written protocol (Menor and Carnero, 2013). The young adults had more years of education ($\bar{x} = 15.1$, $SD = 0.24$) than the older adults ($\bar{x} = 13.5$, $SD = 3.6$), $t(38) = 2.45$, $p = 0.01$, and $d = 0.63$.

Materials

Six photographs were used to portray six typical scenes of a house: a toolbox, a bathroom, a kitchen, a bedroom, a pantry, and a desk, each containing an average of 21.16 objects. The photographs were composed and made expressly for the investigation and each of the objects that appeared in them were selected with a previous investigation following the same procedure of Roediger et al. (2001). Photographs of these scenes were taken in a real context and with real objects. These same materials were previously used with young adult participants, proving to be able to generate false memories (Menor and Carnero, 2013). Twenty-one people between the ages of 18 and 67 who participated in this pilot study, cited ten objects that could be in those scenes. Objects cited by a minimum of ten people were considered high expectation, while objects cited only by one person were considered low expectation. In each scene, four objects were selected, two with high expectation and another two with low expectation, which would not appear in the photographs and which served as contagion items. To build each of the scenes, the rest of the high and low expectation objects were used.

To carry out the contagion phase, a protocol was developed that replaced the physical subject who functioned as a source of social contagion. This protocol was developed similar to the protocol of Meade and Roediger (2002). Each item of the contagion was written on a white paper card in capital letters, which was presented to the participants at the corresponding time in the recall phase, together with the protocol. The list of items used as contagion for each scene is reflected in **Table 1**. Two contagion items appeared for each contagion scene. A high expectancy contagion item always appeared in the fourth position of the protocol, and a low expectancy item was given in the sixth position, as Meade and Roediger (2002) did. The experimenter also had a main item and a reserve item available, to be able to present if the subject would mention, previously and spontaneously, the main contagion items (see **Table 1**).

Design and Procedure

The experiment followed a mixed $2 \times 2 \times 2 \times 2$ design composed of four variables with two levels each: age (older-younger), scenes

exposed to contagion (contagion – no contagion), expectation of the contagion items (high expectation – low expectation), and presentation time of each scene (15–60 s). Exposure to social contagion and expectation of the contagion items were within-subject variables. The time was classified as inter-subject variable, half subjects were randomly assigned to the 15-s condition and others to the 60-s condition. It may be noted that only three of the six scenes watched were exposed to contagion items by protocol. The other three scenes were accompanied by all veridical items through the written protocol, serving as the control condition. The dependent variable was recorded throughout the number of contagion items recalled in the individual free recall test, expressed in proportion over total contagion items exposed. The correct rate of recall in each scene was also measured, over the total items presented in each photograph.

The experiment was carried out in an isolated room. At first, a presentation containing a sequence of the six photographs was viewed on a computer. To half of the participants, the condition of presenting 15 s of each photograph was used. To the other half, the item exposition condition was 60 s. Immediately afterward, the participants were asked to perform a distracting task for 4 min consisting of simple addition and subtraction. Once this was done, the first recall test and key phase for contagion began. It was explained to the participants that these images had already been presented to other individuals, and their recall responses were collected on the cards that the researcher had. Thus, this first test consisted of remembering six items from each scene, but the real participant had to establish a series of turns between their real responses and the card false responses provided by the experimenter. Half of the subjects participating in the 15-s condition were assigned to Group 1 and the other half to Group 2. The same was true for the participants who watched the 60-s presentation. The difference between Groups 1 and 2 was how the contagion-induced scenes were counterbalanced. For Group 1, the scenes that contained the contagion items in the protocol were the toolbox, the kitchen, and the pantry, while for Group 2, the contagion scenes were the bathroom, the bedroom, and the desk. Collaborative recall of the rest of the scenes followed the same procedure, but items written in cards were all items presented in the photographs, without contagion items. The proportion of contagion items that appear by chance in the three no-contagion scenes, served as the control condition. The next phase consisted of an individual recall test, in which

TABLE 1 | Contagion items: The main contagion items for each scene highlighted in bold.

	High		Low	
Toolbox	Pliers	Adjustable wrench	Torch	Silicone
Bathroom	Bar of soap	Sponge	Razor	Nail clippers
Kitchen	Pan	Sink	Coffee maker	Napkins
Bedroom	Lamp	Carpet	Pillow	Slippers
Pantry	Milk	Rice	Dustpan	Potatoes
Desk	Pen tin	Stapler	Ruler	Magazines

The reserve items appear in the adjacent column. Original cards were written in Spanish, the native language of participants.

the participant had to name all the objects that he was able to remember for each scene and with a time limit of 2 min for each scene. When presenting the scenes, the order for the viewing phase remained the same throughout all the tests.

RESULTS

A mixed ANOVA with mentioned variables was done with status of contagion items and expectation items as within-subject factors, and time of presentation and age group as between-subject factors. To begin with, the counterbalance of contagion and non-contagion scenes was analyzed to rule out any differences depending on the features of the scene. For this, the total number of contagion items remembered by the participants in half of the scenes was compared with the participants who suffered the contagion in the others (Group 1 and Group 2). It is confirmed that the counterbalance did not influence the total contagion based on the scenes, $F(1, 38) = 1.09$, $p = 0.301$, and $\eta_p^2 = 0.03$.

Differences of contagion items remembered in the final individual test, between contagion and non-contagion scenes, indicated that participants remembered objects that did not actually appear in the scenes displayed but were suggested in the written protocol during the joint recall phase. A main effect of the total contagion was found, the mean proportion of false memory was higher in the contagion scenes than in the control scenes (see **Figure 1**), $F(1, 36) = 12.49$, $p = 0.001$, and $\eta_p^2 = 0.25$. However, the age group factor was not significant, $F(1, 36) = 2.20$, $p = 0.14$, and $\eta_p^2 = 0.05$, nor was its interaction with the contagion item, $F(1, 36) = 0.22$, $p = 0.63$, and $\eta_p^2 = 0.06$. In fact, no other interactions were significant, $F_s(1, 36) < 1.02$, $p_s > 0.317$, and $\eta_p^2 < 0.02$. Expectation of the item was also found to be significant, $F(1, 36) = 15.28$, $p < 0.001$, and $\eta_p^2 = 0.29$. Time for presentation was not significant, $F(1, 36) = 0.42$, $p = 0.51$, and $\eta_p^2 = 0.01$.

In relation to veridical free-recall performance, the average of correctly recalled items belonging to the control scenes, in which no contagion items were introduced in the collaborative recall phase, was calculated. This was done in order to prevent the false memory from contaminating the veridical memory. An ANOVA was performed with exposition time (15 and 60 s) and age group (younger and older). This analysis revealed a significant main effect of age group, $F(1, 36) = 11.98$, $p = 0.001$, and $\eta_p^2 = 0.25$, whereas older group participants recalled less items ($\bar{x} = 0.31$) from scenes than younger group participants ($\bar{x} = 0.45$). The exposition time of the scene was marginally significant, $F(1, 36) = 3.56$, $p = 0.06$, and $\eta_p^2 = 0.09$, participants recalled more items when scenes were shown for 60 s each ($\bar{x} = 0.42$) than when scenes were shown for 15 s each ($\bar{x} = 0.34$). No interaction effect was observed, $F(1, 36) = 0.009$, and $p = 0.92$.

Regarding episodic memory performance, significant differences were found between younger and older participants in three measures of the word list subtest of Wechsler Memory Scale-III: first trial [younger $\bar{x} = 6.30$, older $\bar{x} = 4.55$, $t(38) = 2.89$, $p = 0.006$, and $d = 0.92$], total recall score after four trials [younger $\bar{x} = 36.35$, older $\bar{x} = 26.55$, $t(38) = 5.02$, $p < 0.001$, and $d = 1.59$], and learning slope [younger $\bar{x} = 5$, older $\bar{x} = 3.75$, $t(38) = 2.10$, $p = 0.04$, and $d = 0.67$]. No significant correlations were found between the total recall scores and social contagion scores in both younger ($r = -0.23$) and older participants ($r = 0.30$, $p_s > 0.10$).

DISCUSSION

The aim of this study was to verify whether the social memory contagion is greater in a group of older people than in a group of younger people, using a contagion protocol that does not require the physical presence of the confederate. The results showed that there were no significant differences between younger and older adults in the total contagion score. The presentation time for the

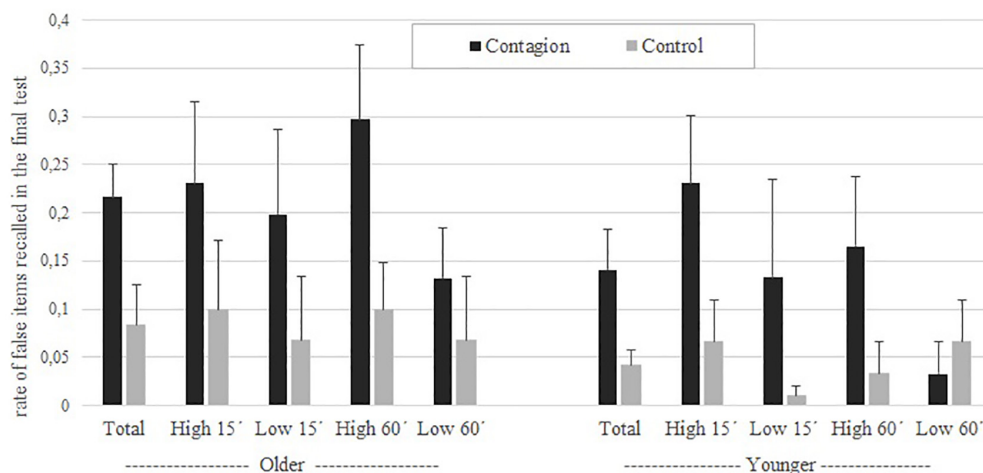


FIGURE 1 | Rate of false items recalled between contagion and control conditions in both older and younger participants. Dark bars show contagion rate whereas light ones depict control performance. Corresponding standard errors are drawn over each bar.

scenes did not affect the false memory, however, expectation of the items showed a relevant effect, due to high expectation objects generating more contagion than low ones. The older group of participants did not show more expectation influence than the younger sample. Consequently, an effect of social contagion was obtained in both groups of participants, but social contagion was not superior in the older adults group. We had hypothesized more social contagion effect in older than younger participants due to the loss of distinctive clues in the written protocol. However, this did not occur. Therefore, written protocol seems to be as powerful as physical confederate, at least for generating the same level of source distinctiveness. Furthermore, these results extend those obtained by Davis and Meade (2013) who used a similar procedure with a physical confederate. In addition, although the group of older adults had a lower score in episodic memory than the group of younger adults, the magnitude of the effect of social contagion did not significantly correlate with episodic memory in either group.

Thus, it can be stated that the procedure for finding social contagion of memory through a written protocol seems adequate in older population. Therefore, it was enough for older participants to consider that the answers provided came from other participants in the same situation. In this regard, it is interesting to note that when the contagion items were presented in the protocol, the participants frequently claimed to remember those objects that did not appear in the photographs at any time.

The presentation time of the scenes did not significantly affect the social contagion of memory, nor did it interact with the age group. Unlike what was hypothesized, the shorter presentation time of the scenes did not increase the effect of social contagion in either of the groups. This result replicates that obtained by Menor and Carnero (2013), who found that the presentation time increased social contagion only when the contagion items were presented through a physical confederate, but not when it was virtual. It is important to note that the effects of presentation time are not consistent in the literature on false recall. Using other false memory paradigms (i.e., McDermott and Watson, 2001) an inverted U-shape function has been found, and Meade and Roediger (2002, exp. 2) did not find presentation time effects (5 vs 15 s per slide) using a physical confederate. Therefore, more research is needed to clarify the conditions under which the presentation time of scenes affects the paradigm of social contagion.

It is not infrequent to find diversity of results regarding the differences in tasks on false memories and the interference of

false information in young and old adults. Roediger and Geraci (2007) explain these differences as the result of the different methodology used and the diversity of samples of older adults. However, as have been observed in Gabbert et al. (2003) following other procedure and Davis and Meade (2013) with the original protocol from Roediger et al. (2001), the results found in the present study supports the absence of age differences in a social contagion paradigm employing for the first time a written confederate protocol. Apparently, the social aspects included in contagion paradigms, even when using written protocols, contribute to neutralizing the enhanced false memories usually found in older people performing individual tasks.

The present study has some limitations that should be noted. The absence of differences between younger and older participants in social contagion could be due to the lack of power of the test to obtain these differences. A power sample analysis showed the need to improve the sample size used in this study to replicate this result, despite its similarity with Meade and Roediger (2002) experiments. Further research should also analyze individual differences among older adults since their cognitive performance is more variable than that of younger adults (Lindenberger and Oertzen, 2006). It is possible that the social memory contagion differs among older people due, for example, to variations in the ability to monitor memories and executive functioning (Colombel et al., 2016).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Oviedo. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Blesa, R., Pujol, M., Aguilar, M., Santacruz, P., Bertran-Serra, I., Hernández, G., et al. (2001). Clinical validity of the "mini-mental state" for Spanish speaking communities. *Neuropsychologia* 39, 1150–1157. doi: 10.1016/s0028-3932(01)00055-0
- Colombel, F., Tessoulin, M., Gilet, A.-L., and Corson, Y. (2016). False memories and normal aging: links between inhibitory capacities and monitoring processes. *Psychol. Aging* 31, 239–248. doi: 10.1037/pag0000086
- Davis, S. D., and Meade, M. L. (2013). Both young and older adults discount suggestions from older adults on a social memory test. *Psychon. Bull. Rev.* 20, 760–765. doi: 10.3758/s13423-013-0392-5
- Devitt, A. L., and Schacter, D. L. (2016). False memories with age: neural and cognitive underpinnings. *Neuropsychologia* 91, 346–359. doi: 10.1016/j.neuropsychologia.2016.08.030
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental State". a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

- Gabbert, F., Memon, A., and Allan, K. (2003). Memory conformity: can eyewitnesses influence each other's memories for an event? *Appl. Cogn. Psychol.* 17, 533–543. doi: 10.1002/acp.885
- Gabbert, F., Memon, A., Allan, K., and Wright, D. B. (2004). Say it to my face: examining the effects of socially encountered misinformation. *Legal Criminol. Psychol.* 9, 215–227. doi: 10.1348/1355325041719428
- Goff, L. M., and Roediger, H. L. (1998). Imagination inflation for action events: repeated imaginings lead to illusory recollections. *Mem. Cognit.* 26, 20–30. doi: 10.3758/BF03211367
- Harris, C. B., Barnier, A. J., Sutton, J., and Khan, T. (2017). Social contagion of autobiographical memories. *J. Appl. Res. Mem. Cog.* 6, 319–327. doi: 10.1016/j.jarmac.2017.07.006
- Henkel, L. A., and Rajaram, S. (2011). Collaborative remembering in older adults: age-invariant outcomes in the context of episodic recall deficits. *Psychol. Aging* 26, 532–545. doi: 10.1037/a0023106
- Lindenberger, U., and Oertzen, T. (2006). “Variability in cognitive aging: from taxonomy to theory,” in *Lifespan Cognition: Mechanisms of Change*, eds E. Byalystok and F. I. Craik (Oxford: Oxford University Press), 297–314. doi: 10.1093/acprof:oso/9780195169539.003.0021
- Loftus, E. F., and Palmer, J. C. (1974). Reconstruction of automobile destruction. *J. Verbal Learn. Verbal Behav.* 13, 585–589. doi: 10.1016/S0022-5371(74)80011-3
- McDermott, K. B., and Watson, J. M. (2001). The rise and fall of false recall: the impact of presentation duration. *J. Mem. Lang.* 45, 160–176. doi: 10.1006/jmla.2000.2771
- Meade, M. L., Perga, V. A., and Hart, K. M. (2020). “Memory and aging in social contexts,” in *Handbook of Cognitive Aging: A Life Course Perspective*, eds A. K. Thomas and A. Gutches (Cambridge, MA: Cambridge University Press), 281–298. doi: 10.1017/9781108552684.018
- Meade, M. L., and Roediger, H. L. (2002). Explorations in the social contagion of memory. *Mem. Cognit.* 30, 995–1009. doi: 10.3758/BF03194318
- Meade, M. L., and Roediger, H. L. (2006). The effect of forced recall on illusory recollection in younger and older adults. *Am. J. Psychol.* 119, 433–462. doi: 10.2307/20445352
- Meade, M. L., and Roediger, H. L. (2009). Age differences in collaborative memory: the role of retrieval manipulations. *Mem. Cognit.* 37, 962–975. doi: 10.3758/MC.37.7.962
- Menor, J., and Carnero, S. (2013). La importancia de la fuente de contagio en la formación de falsos recuerdos. *Rev. Iberoam. Psicol. Salud.* 4, 77–91.
- Mitchell, K. J., Johnson, M. K., and Mather, M. (2003). Source monitoring and suggestibility to misinformation: adult age-related differences. *Appl. Cognit. Psychol.* 17, 107–119. doi: 10.1002/acp.857
- Roediger, H. L., and Geraci, L. (2007). Aging and the misinformation effect: a neuropsychological analysis. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 321–334.
- Roediger, H. L., and McDaniel, M. A. (2007). “Illusory recollections in older adults: testing mark twain’s conjecture,” in *Do Justice and Let the Sky Fall: Elizabeth F. Loftus and her Contributions to Science, Law, and Academic Freedom*, eds M. Garry and H. Hayne (Hillsdale, NJ: Lawrence Erlbaum Associates), 105–136.
- Roediger, H. L., and McDermott, K. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 804–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, H. L., Meade, M. L., and Bergman, E. T. (2001). Social contagion of memory. *Psychon. Bull. Rev.* 8, 365–371. doi: 10.3758/BF03196174
- Schacter, D. L., Koutstaal, W., and Norman, K. A. (1997). False memories and aging. *Trends Cogn. Sci.* 1, 229–236. doi: 10.1016/S1364-6613(97)01068-1
- Wechsler, D. (1999). *WAIS-III. Escala de Inteligencia de Wechsler Para Adultos-III*. Madrid: TEA.
- Wechsler, D. (2004). *WMS-III. Escala de Memoria de Wechsler-III*. Madrid: TEA.
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* 17, 37–49. doi: 10.1016/0022-3956(82)90033-4

Conflict of Interest: 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.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Carnero-Sierra and Menor. 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) and the copyright owner(s) 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.



Working Memory Beats Age: Evidence of the Influence of Working Memory on the Production of Children's Emotional False Memories

Chiara Mirandola^{1,2*} and Francesca Pazzaglia²

¹ Department of Philosophy, Sociology, Education and Applied Psychology, University of Padua, Padua, Italy, ² Department of General Psychology, University of Padua, Padua, Italy

OPEN ACCESS

Edited by:

Beatriz Martín-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

Kate Cockcroft,
University of the Witwatersrand,
South Africa
Carlos M. Gómez,
University of Seville, Spain

*Correspondence:

Chiara Mirandola
chiara.mirandola@unipd.it

Specialty section:

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

Received: 25 May 2021

Accepted: 07 July 2021

Published: 18 August 2021

Citation:

Mirandola C and Pazzaglia F
(2021) Working Memory Beats Age:
Evidence of the Influence of Working
Memory on the Production
of Children's Emotional False
Memories.
Front. Psychol. 12:714498.
doi: 10.3389/fpsyg.2021.714498

Emotional valence and working memory ability (WM) affect false memories' production in adults. Whereas a number of studies have investigated the role of emotional valence in children's tendency to produce spontaneous false memories, individual differences in WM have not been previously included. In the current article, we were interested in investigating whether emotion and WM would interact in influencing the propensity to incur inferential false memories for scripted events. Ninety-eight typically developing children (first-, third-, and eighth- graders) were administered the Emotional false memory paradigm – allowing to study false memories for negative, positive, and neutral events – and a WM task. Results showed that regardless of age, valence influenced false memories' production, such that positive events protected against incurring distortions. Furthermore, WM interacted with valence, such that children with higher WM abilities produced fewer false memories for negative events. Concerning confidence judgments, only the youngest group of children claimed to be overconfident when committing false memories for negative and neutral events. Results are discussed in terms of the role of individual differences in higher cognitive abilities interacting with the emotional content of to-be-remembered events.

Keywords: emotion, false memories, working memory, children, confidence

INTRODUCTION

Emotionally laden events typically lead to better memories than mundane events. A deluge of studies proved this memory-enhancing effect in adults (e.g., Kensinger, 2009). Nonetheless, adults are not exempted from memory distortions when retrieving emotional events (e.g., Kensinger and Schacter, 2005; Brainerd et al., 2008b; Gallo et al., 2009; Mirandola et al., 2014b, 2017). Whether children benefit from a similar effect of emotion on memory and what factors may influence the development of false memories for emotional events are still debated. The current study was aimed at investigating the relation between the emotional content of to-be-remembered events and individual differences in higher cognitive processes, such as working memory capabilities, in children's tendency to falsely remember everyday events.

For neutral material, when false memories stem from elaborating semantically related wordlists [as in the widely used Deese-Roediger-McDermott (DRM) paradigm;

Roediger and McDermott, 1995], a developmental reversal emerges in typically developing children: i.e., age-related increases in memory errors. The DRM paradigm consists of the presentation of several wordlists; within each list, words are semantically related to each other (e.g., *sick*, *nurse*, and *medicine*) and to a critical lure not presented at encoding (e.g., *doctor*). Claiming to remember such critical lures at retrieval represent committing a false memory. Younger children recall or recognize fewer critical lures than older children and adults. This developmental reversal applies also to emotional stimuli (Howe, 2007; Brainerd et al., 2010). Both the associative-monitoring framework (AAT; Howe, 2005, 2008) and the fuzzy-trace theory (FTT; Brainerd et al., 2008a) may explain the developmental reversal. The former predicts that semantically related but non-presented critical lures are activated at encoding more so than do semantically unrelated words, and that this effect is even boosted when the critical lures are emotionally charged; thus, it follows that younger children – whose ability to process semantic relations among words is not fully developed and undergoes important changes due to maturation and education during development – produce false memories to a lesser extent than older peers and adults. The FTT instead predicts a similar developmental increase in false memories since emotionally charged events support gist (i.e., thematic) connections more than neutral events and gist connections are responsible for false memory formation. Thus, it follows that the developing ability to detect gist connections among experienced events leads to increasing production of false memory with increasing age.

A more ecologically valid paradigm, the Emotional False Memory Paradigm (Mirandola et al., 2017) has been recently widely used to test spontaneous false memories for emotionally valenced events in the entire life-span: older adults (Toffalini et al., 2019), adults (Mirandola et al., 2014b, 2017, 2020; Mirandola and Toffalini, 2016), adolescents (Toffalini et al., 2014, 2015) and both typically developing children (Melinder et al., 2017) and children with developmental disorders (Mirandola et al., 2014a; Solomon et al., 2019). This paradigm consists of the presentation of photographs depicting different scripts or episodes at encoding. Each episode may end either in a positive, negative or neutral way. For example, in the dating episode, a boy and a girl are shown while they are getting ready to go out on a date (e.g., getting dressed, combing their hair, text messaging on the cellphone, etc.). The episode may have a positive (the boy and the girl are fiancée, and they kiss each other), negative (the boy is aggressive toward the girl for being late) or neutral (the boy and girl are friends and meet for exchanging a textbook) ending. The causal antecedent – that is the scene depicting what happens in the story right before the episode ending – is not shown during encoding, but it is presented at recognition. Claiming to remember the causal antecedent represents committing an inferential causal error (see **Supplementary Figure 1** for the pictorial example of this episode). Another type of error elicited within this paradigm is the gap-filling error, which corresponds to erroneously remembering non-presented but script-consistent pictures (e.g., remembering the girl combing her hair while she was brushing her teeth). The available evidence in typically developing children shows that positive but not negative events

protect children – regardless of age – from incurring false memories, especially causal errors (Melinder et al., 2017). The authors suggest that the underdeveloped working memory in children (as young as 6 years in their study) may have played a role. Evidence in adults with the same paradigm (Mirandola et al., 2017) shows that both individual differences in working memory (Exp. 1) and a double task at encoding (Exp. 2) influence false memories for negative events: indeed, individuals with lower WM abilities produce more false memories for negative events, thus discarding the protecting effect of emotion on false memory production. The hypothesized mechanism is that individuals with higher WM are better able to manipulate both useful and irrelevant information, excluding the latter ones when unnecessary for the ongoing task. Given that negative information is more difficult to be inhibited than positive information (Osaka et al., 2013), it follows that people with reduced WM would struggle to manipulate negative events and thus would be more prone to include negative events in their memory, even if not presented (Mirandola et al., 2017). As far as children are concerned, given that younger children have lower WM abilities than older children and adults, one would expect more false memories for negative events in younger children. Melinder et al. (2017) found that regardless of age only positive events protected against false memories and that negative false memories were produced to a similar extent than neutral ones. However, they could only speculate on the possible role of WM given that it was not tested in their sample of children. The current study was aimed at specifically testing the influence of WM abilities on the production of emotional false memories in children. To this end, we administered the Emotional False Memory Paradigm and the Letter-number sequencing WM subtest of the Wechsler Intelligence Scale for Children to three groups of children in order to detect possible developmental changes: 1st graders, 3rd graders, and 8th graders. We could hypothesize that younger children – due to lower WM ability – would produce more false memories for negative events compared to older children. Alternatively, we could hypothesize that regardless of age, WM would interact with emotional valence protecting children from incurring false memories for negative events.

METHODS

Participants

Ninety-eight typically developing children participated to this study. Specifically, 23 first-grade children ($M_{age} = 82.6$, $SD = 3.4$; females = 10), 40 third-grade children ($M_{age} = 105$, $SD = 4.4$; females = 20), and 35 eight-grade children ($M_{age} = 156.6$, $SD = 6.6$; females = 12). We determined the number of participants on the basis of a power analysis using the G*Power program (Erdfeiler et al., 1996), which indicated that a total sample of 87 participants would be needed to detect differences using a repeated measure ANOVA design, with power ($1 - \beta$) set at 0.90, alpha at 0.05, and with an estimated $\eta_p^2 = 0.04$ (based on previous work with the same false memory paradigm and the same repeated measure design; Mirandola et al., 2017).

Inclusion criteria were: Italian as either first language or second language but with advanced knowledge (documented by the teachers) and absence of a diagnosis of learning disorders or other neurodevelopmental disorders (documented by the school records). Due to school restrictions we could not assess participants' IQ; however, we shall note that the WM task that we employed in this study is included into the Wechsler intelligence scale for children to calculate the total IQ; furthermore, WM ability and IQ are highly correlated (see Giofrè and Cornoldi, 2015). We obtained written informed consent from all children's parents. The study was approved by the Local Ethical Board of the University of Padua (no. 2232).

Materials and Procedure

Emotional False Memory Paradigm

Encoding

A sequence of color photographs depicting nine episodes was employed. The episodes were the following: going grocery shopping, waking up, going to a bike trip, rock climbing, track competition, homecoming after a long trip, dating, birthday party, playing at the slot machine (see Mirandola et al., 2014b, 2017 for pictorial examples of the episodes). For each episode, 14 photographs depicted actions that typically occur during the event (11 were used as target photographs in the encoding phase and three were used as gap-filling distractors in the recognition phase), and two photographs depicted cause-effect scenes (the effect scene was studied whereas the cause scene was presented only during the recognition test). The emotionality of the effect photographs was balanced across episodes, such that the same cause could have three different outcomes: positive, negative, and neutral. Finally, the stimuli also included 10 photographs that were inconsistent with any of the nine episodes, such as children playing on the beach, shown at the beginning and at the end of the presentation in order to avoid primacy and recency effects on the relevant material. Participants saw the nine episodes in sequence without any interruption between them.

Children were tested individually in a quiet room at their school. They were told that they would see several photographs depicting other children and young people doing different daily activities and that they would have to pay close attention and try to understand what the stories represented. The encoding phase consisted of a series of 126 photographs; each photo was presented for 2 s and was followed by a black screen lasting 2 s. The nine episodes were presented – using Microsoft PowerPoint program – in a fixed order, with target-distractor photographs and valence of the episode-ending varying across participants. The encoding phase was followed by a 15-min retention interval. During this interval, children were administered filler tasks.

Recognition

Children received a surprise memory test. Stimuli for the recognition phase consisted of a series of 45 target and 45 distractor photographs in a randomized order. For each episode, four targets and four distractors were included (one of the four distractor photographs were the causal antecedent whose outcome had been presented during study). Further, 18 photographs inconsistent with any of the episodes were included

(nine targets and nine distractors). The memory test consisted of a self-paced recognition task. For each photograph, children had to utter “yes” or “no” whether they could, respectively, remember having seen the photograph during the encoding phase or not. Furthermore, children had to provide confidence judgments for each response given, using the Confidence Rating Board (CRB), proved effective even with children as young as 5 years.

Confidence Rating Board

The Confidence Rating Board (CRB; Ghetti et al., 2002) consists of two photographs that depict either a child with a confident expression or the same child with a doubtful expression. These photographs are positioned on the opposite sides of a white board. Three dots are drawn between these photographs which represent the three degrees of confidence (very sure, somewhat sure, and not sure at all). Children were instructed to utter how sure they were that they saw/did not see the picture before, using the board as a help; they could only point to the dot near the picture of the child with a confident facial expression when they were very sure (that they saw or that they did not see the photograph), the middle dot, when they were somewhat sure, and the dot near the doubtful facial expression when they were not at all sure. See the photographs used in the CRB in the **Supplementary Figure 2**.

At the end of the experimental session, participants were also asked to orally describe what the actors of each script were doing and feeling. All children were able to explain in their own words the content of the stories and the feelings of the actors represented.

Working Memory Task

After the Emotional False Memory paradigm, children were administered the WM task (Letter-Number Sequencing, WISC-IV; Wechsler, 2004). The WM task consists of the presentation of a sequence of digits and letters in “scrambled” order; the child is required to immediately repeat first the digits from the smallest to the biggest, and second the letters in alphabetical order. Sequences of alphanumerical elements increase from two to ten strings. The task is self-paced, such that after the erroneous repetition of two strings within the same sequence, it is interrupted and the corresponding WM span is calculated.

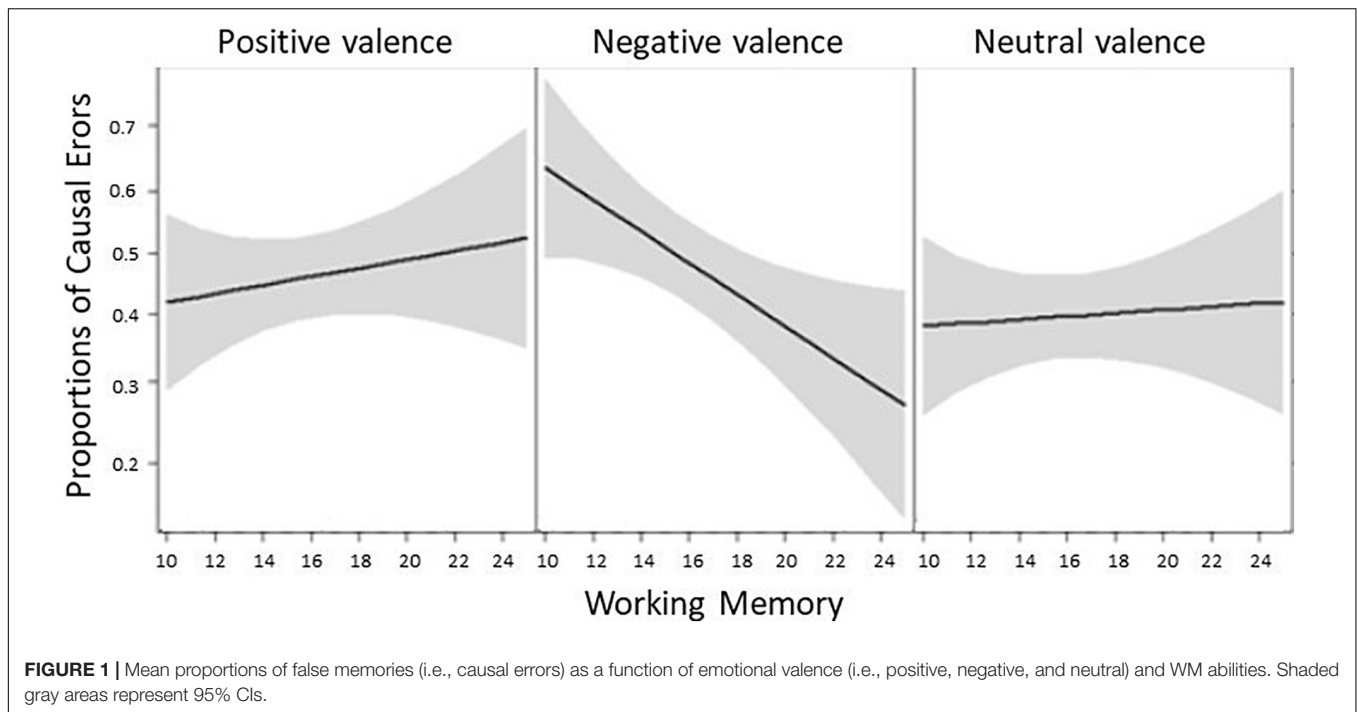
RESULTS

Working Memory

A univariate ANOVA revealed a main effect of Grade (1st vs. 3rd vs. 8th) on the WM span, $F(2,95) = 25.53$, $p < 0.001$, $\eta_p^2 = 0.35$, showing a clear developmental improvement: 8th graders had a significantly higher WM span ($M = 18.83$, $SD = 3.28$) than 3rd graders ($M = 16.15$, $SD = 2.25$), who in turn had a significantly higher WM span than 1st graders ($M = 13.52$, $SD = 2.82$).

False Memories

Within the Emotional False Memory paradigm two types of false memories were calculated, namely causal errors and gap-filling errors. Mean proportions of “yes” responses to causal distractor



images represent causal errors, whereas mean proportions of “yes” responses to script-consistent distractor images represent gap-filling errors. Preliminary analysis with gender as the between-participant factor and proportions of either causal errors ($p = 0.44$) and gap-filling errors ($p = 0.25$) as the dependent measures, revealed no main effect. Thus, gender was not included as a covariate in the following analyses. A linear model (ANCOVA) was computed with Grade (1st vs. 3rd vs. 8th) as the between-participants factor, Valence (positive vs. negative vs. neutral) as the within-participants factor and WM score (continuous variable treated as a covariate) over the proportion of causal errors as the dependent measure. A main effect of Valence was found, $F(2,188) = 3.95$, $p = 0.02$, $\eta_p^2 = 0.04$, such that regardless of age, causal errors for positive events ($M = 0.39$, $SD = 0.28$) were produced to a lesser extent than both negative ($M = 0.46$, $SD = 0.32$) and neutral ($M = 0.47$, $SD = 0.28$) events. Furthermore, a significant interaction between Valence and WM emerged, $F(2,188) = 3.70$, $p = 0.03$, $\eta_p^2 = 0.04$. As a *post hoc* analysis for this interaction we calculated single correlations between WM and each of the three levels of valenced causal errors. Only the correlation between WM and negative causal errors was significant ($r = -0.25$, $p = 0.01$; positive causal errors: $r = -0.07$, $p = 0.48$; neutral causal errors: $r = -0.09$, $p = 0.33$) (see **Figure 1**). The main effect of Grade was not found, $F(2,94) = 0.77$, $p = 0.46$, $\eta_p^2 = 0.01$.

We then computed a similar analysis with grade, valence, and WM as the predictors and gap-filling errors as the dependent measure. No main effect of Valence [$F(2,188) = 0.05$, $p = 0.95$, $\eta_p^2 = 0.001$] nor interactive effect with Grade [$F(4,188) = 0.68$, $p = 0.60$, $\eta_p^2 = 0.01$] and WM [$F(2,188) = 0.08$, $p = 0.92$, $\eta_p^2 = 0.001$] emerged. We found a main effect of Grade, $F(2,94) = 3.14$, $p = 0.048$, $\eta_p^2 = 0.06$, with the tendency of

older children to produce fewer gap-filling errors than 8-year-old children (8th graders: $M = 0.20$, $SD = 0.13$; 3rd graders: $M = 0.28$, $SD = 0.16$; 1st graders: $M = 0.23$, $SD = 0.12$); however, Bonferroni-corrected *post hoc* comparisons did not reach statistical significance ($p = 0.073$).

Confidence Judgments Associated With False Memories

When we computed the analysis on the mean confidence judgments associated with causal errors, the main effect of Valence [$F(2,186) = 0.65$, $p = 0.52$, $\eta_p^2 = 0.007$], the main effect of Grade [$F(2,93) = 0.59$, $p = 0.55$, $\eta_p^2 = 0.01$] nor the interactive effect between Valence and WM [$F(2,186) = 1.43$, $p = 0.24$, $\eta_p^2 = 0.01$] did not emerge. Interestingly, the interaction between Valence and Grade was significant, $F(4,186) = 2.39$, $p = 0.05$, $\eta_p^2 = 0.05$; Bonferroni-corrected *post hoc* comparisons showed that only within first Grade did children claim to be overconfident when committing causal errors for negative ($p = 0.001$) and neutral events ($p = 0.01$), compared to positive ones (see **Figure 2**). A similar analysis conducted on the mean confidence judgments associated with gap-filling errors did not show any main nor interactive effect (all $ps > 0.34$).

Accuracy

An analysis of variance with Grade as the between-participant factor, WM as the covariate and the “yes” responses to target photographs (i.e., hit rate) as the dependent measure did not reveal the main effect of Grade [$F(2,94) = 0.63$, $p = 0.53$, $\eta_p^2 = 0.01$] nor WM [$F(1,94) = 1.63$, $p = 0.21$, $\eta_p^2 = 0.01$]. A similar analysis conducted over the mean

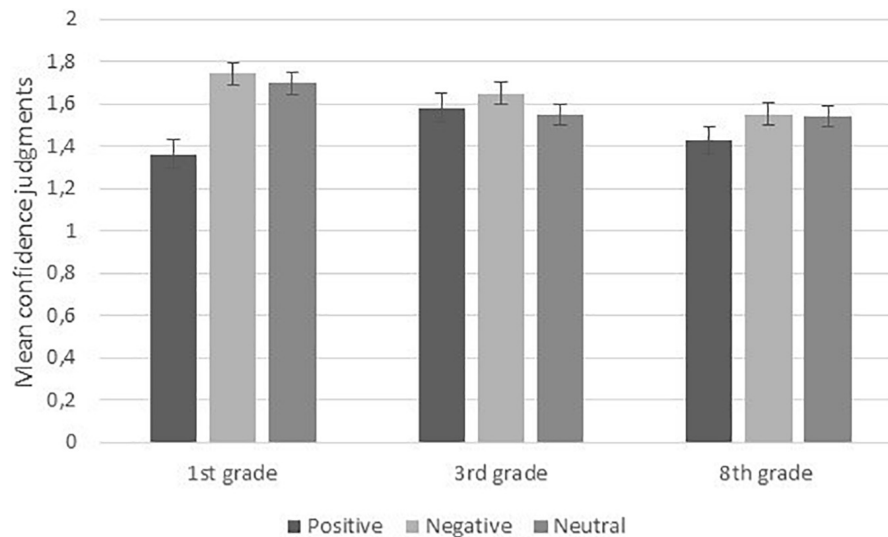


FIGURE 2 | Mean confidence judgments associated with causal errors as a function of emotional valence and grade. Bars represent standard errors.

confidence judgments for the hit rate did not show any significant effect (all $F_s < 1$).

DISCUSSION

With the current article, we wanted to deepen the understanding of the relation between emotion and WM abilities on the creation of false memories during development. Two hypotheses were put forward: (1) younger children could produce more false memories for negative events, given their reduced WM, compared to older children; and (2) regardless of age, WM could play a stronger influence, with decreasing production of false memories for negative events with increasing WM abilities in all children. We found that individual differences in WM indeed interact with emotion and influence false memory formation to a different extent. Specifically, regardless of age, all children with higher WM abilities produced fewer causal errors for negative events, sustaining the second hypothesis. This finding replicates what previously found in the adult population (Mirandola et al., 2017). Thus, even during development, individual differences in WM are relevant in determining a different tendency to distort memories. Working memory allows to attend to a selection of stimuli while controlling for interfering/irrelevant information (e.g., De Beni et al., 1998; Engle, 2002). We argue that within the current paradigm the negative non-presented events are inferred at encoding, thus becoming interfering information into the newly formed memory trace. People with a lower ability at manipulating the inferred information and later excluding it from memory, incorporate negative false events into their memory and are no longer able to distinguish them from true events. It follows, that people with a lower WM produce a higher amount of false memories for negative events. The current article suggests that this reasoning may be applied to a developmental sample as well. We may hypothesize that the central executive component of

Baddeley's WM model (for a more recent review see Baddeley, 2012) plays an important role; indeed, the central executive has the function to control the processes at hand in ongoing complex tasks (such as the Letter-number sequencing used in the current study which requires children to simultaneously process letters in their alphabetical order and numbers from the smallest to the biggest). Individuals with higher ability in terms of multi-task processing (possible through the central executive) should be better able to distinguish target/experienced events from inferred but not experienced ones while performing an episodic memory task (see also Gómez et al., 2018 for a thorough review of the neurodevelopment of working memory). Furthermore, these findings suggest the need for working memory training in typically developing children with lower WM abilities or in children with disabilities who have hindered WM capacity, in order to investigate the potential benefit on episodic memory, in terms of both accuracy and false recognitions (see Capodieci et al., 2019 for WM training in typically developing children and children with ADHD; Cornoldi et al., 2015 for WM and metacognitive training in typically developing children; and Fraser and Cockcroft, 2020 for WM intervention in special populations of adolescents).

The current work also replicated previous findings, that positive emotional valence *per se* – but not negative valence – protects children from incurring false memories (Melinder et al., 2017). In particular, the effect of emotional valence pertained to causal errors and not gap-filling errors – as found in previous research with the same paradigm (Mirandola et al., 2014b, 2017; Melinder et al., 2017). This is easily explained by the fact that causal photographs are directly linked to the emotional consequence of the episodes, whereas gap-filling errors are not. We did not find developmental differences in causal errors, again replicating available evidence with the same paradigm in typically developing children (Melinder et al., 2017), nor in the accurate recognition of target photographs (i.e., hit rate).

While it was possible to hypothesize increasing accuracy relative to correct recognitions of target photographs we must specify that the current Emotional false memory paradigm is specifically designed to elicit false memories and it does not usually lead to developmental differences in accuracy measures nor other effects of the variables of interest on overall memory accuracy (e.g., Mirandola et al., 2014a, 2017).

Another interesting finding concerns the qualitative nature of memory, that is confidence judgments relative to false memories. Younger children did claim to be very confident after falsely recognizing both negative and neutral events – but not positive ones – suggesting again that emotion does not influence only the quantitative aspects of memory, but also the qualitative ones. Previous evidence with the same Emotional false memory paradigm in atypical development (Mirandola et al., 2014a) – but that investigated subjective remembering through the Remember-know paradigm – showed that children with non-verbal learning disabilities claimed to subjectively remember causal errors to a higher extent than typically developing children, regardless of emotional content. These findings suggest that children with disabilities not only produce more false memories, but when that happens, they also associate a higher subjective feeling of vividness to non-experienced events. In the current study, only younger children associated a higher subjective confidence to their negative and neutral false memories compared to the positive ones, suggesting that non-experienced but inferred negative events may be subjectively more compelling than positive ones in younger children. This is conceptually similar to the higher subjective judgments of recollection-based false memories in children with developmental disabilities (Mirandola et al., 2014a).

Taken together, these results show that when studying emotional memory for everyday events, children are protected against distortions only when facing positive events, but not when facing negative events, which are produced to a similar extent than the neutral non-emotional ones. From an applied forensic perspective, it is worth noting the relevance of this evidence. We may expect that children incorporate more false negative events (such as maltreatment episodes) into their memory traces compared to positive ones. However, working memory abilities play an important protecting role in this mechanism, such that even during development, children who have higher WM abilities

may benefit from this, reducing their tendency to incur memory distortions for negative events.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Data are available on Figshare at doi: 10.6084/m9.figshare.14672622.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee for the Psychological Research of the University of Padua. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CM and FP contributed to the conception and design of the study and read and approved the submitted version. CM organized the dataset, performed the statistical analysis, and wrote the first draft of the manuscript. FP contributed to manuscript revision. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Elena Oteri for her involvement with her students in the current project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.714498/full#supplementary-material>

REFERENCES

- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Brainerd, C. J., Holliday, R. E., Reyna, V. F., Yang, Y., and Toglia, M. P. (2010). Developmental reversals in false memory: effects of emotional valence and arousal. *J. Exp. Child Psychol.* 107, 137–154. doi: 10.1016/j.jecp.2010.04.013
- Brainerd, C. J., Reyna, V. F., and Ceci, S. J. (2008a). Developmental reversals in false memory: a review of data and theory. *Psychol. Bull.* 134, 343–382. doi: 10.1037/0033-2909.134.3.343
- Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., and Reyna, V. F. (2008b). How does negative emotion cause false memories? *Psychol. Sci.* 19, 919–925. doi: 10.1111/j.1467-9280.2008.02177.x
- Capodieci, A., Re, A. M., Fracca, A., Borella, E., and Carretti, B. (2019). The efficacy of a training that combines activities on working memory and metacognition: transfer and maintenance effects in children with ADHD and typical development. *J. Clin. Exp. Neuropsychol.* 41, 1074–1087. doi: 10.1080/13803395.2019.1651827
- Cornoldi, C., Carretti, B., Drusi, S., and Tencati, C. (2015). Improving problem solving in primary school students: the effect of a training programme focusing on metacognition and working memory. *Br. J. Educ. Psychol.* 85, 424–439. doi: 10.1111/bjep.12083
- De Beni, R., Palladino, P., Pazzaglia, F., and Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *Q. J. Exp. Psychol.* 51A, 305–320. doi: 10.1080/027249898391648
- Engle, R. W. (2002). Working memory capacity as executive attention. *Curr. Direct. Psychol. Sci.* 11, 19–23. doi: 10.1111/1467-8721.00160
- Erdfelder, E., Faul, F., and Buchner, A. (1996). GPOWER: a general power analysis program. *Behav. Res. Methods Instrum. Comp.* 28, 1–11. doi: 10.3758/bf03203630

- Fraser, S., and Cockcroft, K. (2020). Working with memory: a working memory intervention for adolescents living with HIV. *Child Neuropsychol.* 26, 612–634. doi: 10.1080/09297049.2019.1676407
- Gallo, D. A., Foster, K. T., and Johnson, E. L. (2009). Elevated false recollection of emotional pictures in young and older adults. *Psychology and Aging* 24, 981–988. doi: 10.1037/a0017545
- Ghetti, S., Qin, J., and Goodman, G. S. (2002). False memories in children and adults: age, distinctiveness, and subjective experience. *Dev. Psychol.* 38, 705–718. doi: 10.1037/0012-1649.38.5.705
- Giofrè, D., and Cornoldi, C. (2015). The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. *Intelligence* 52, 36–43. doi: 10.1016/j.intell.2015.07.002
- Gómez, C. M., Barriga-Paulino, C. I., Rodríguez-Martínez, E. I., Rojas-Benjumea, M. Á., Arjona, A., and Gómez-González, J. (2018). The neurophysiology of working memory development: from childhood to adolescence and young adulthood. *Rev. Neurosci.* 29, 261–282. doi: 10.1515/revneuro-2017-0073
- Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychol. Sci.* 16, 927–931. doi: 10.1111/j.1467-9280.2005.01638.x
- Howe, M. L. (2007). Children's emotional false memories. *Psychol. Sci.* 18, 856–860.
- Howe, M. L. (2008). Visual distinctiveness and the development of children's false memories. *Child Dev.* 79, 65–79. doi: 10.1111/j.1467-8624.2007.01111.x
- Kensinger, E. A. (2009). *Emotional Memory Across The Adult Lifespan*. New York, NY: Psychology Press.
- Kensinger, E. A., and Schacter, D. L. (2005). Emotional content and reality-monitoring ability: fMRI evidence for the influence of encoding processes. *Neuropsychologia* 43, 1429–1443. doi: 10.1016/j.neuropsychologia.2005.01.004
- Melinder, A., Toffalini, E., Geccherle, E., and Cornoldi, C. (2017). Positive events protect children from causal false memories for scripted events. *Memory* 25, 1366–1374. doi: 10.1080/09658211.2017.1306080
- Mirandola, C., Losito, N., Ghetti, S., and Cornoldi, C. (2014a). Emotional false memories in children with learning disabilities. *Res. Dev. Disabil.* 35, 261–268. doi: 10.1016/j.ridd.2013.11.004
- Mirandola, C., and Toffalini, E. (2016). Arousal—but not valence— reduces false memories at retrieval. *PLoS One* 11:e0148716. doi: 10.1371/journal.pone.0148716
- Mirandola, C., Toffalini, E., Ciriello, A., and Cornoldi, C. (2017). Working memory affects false memory production for emotional events. *Cogn. Emot.* 31, 33–46. doi: 10.1080/02699931.2015.1075379
- Mirandola, C., Toffalini, E., Ferruzza, E., and Pazzaglia, F. (2020). The power of extraversion? Reduced false memories for positive events. *Pers. Individ. Differ.* 159, 1–5.
- Mirandola, C., Toffalini, E., Grassano, M., Cornoldi, C., and Melinder, A. (2014b). Inferential false memories of events: negative consequences protect from distortions when the events are free from further elaboration. *Memory* 22, 451–461. doi: 10.1080/09658211.2013.795976
- Osaka, M., Yaoi, K., Minamoto, T., and Osaka, N. (2013). When do negative and positive emotions modulate working memory performance? *Sci. Rep.* 3:1375. doi: 10.1038/srep01375
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Solomon, M., Iosif, A. M., Krug, M. K., Nordahl, C. W., Adler, E., Mirandola, C., et al. (2019). Emotional false memory in autism spectrum disorder: more than spared. *J. Abnorm. Psychol.* 128, 352–363. doi: 10.1037/abn0000418
- Toffalini, E., Mirandola, C., Altoè, G., and Borella, E. (2019). Inferential false memories for emotional events in older adults. *Br. J. Psychol.* 110, 686–706. doi: 10.1111/bjop.12371
- Toffalini, E., Mirandola, C., Coli, T., and Cornoldi, C. (2015). High trait anxiety increases inferential false memories for negative (but not positive) emotional events. *Pers. Individ. Differ.* 75, 201–204. doi: 10.1016/j.paid.2014.11.029
- Toffalini, E., Mirandola, C., Drabik, M. J., Melinder, A., and Cornoldi, C. (2014). Emotional negative events do not protect against false memories in young adults with depressive–anxious personality traits. *Pers. Individ. Differ.* 66, 14–18. doi: 10.1016/j.paid.2014.02.042
- Wechsler, D. (2004). *The Wechsler Intelligence Scale for Children*, 4th Edn. London: Pearson Assessment.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mirandola and Pazzaglia. 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) and the copyright owner(s) 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.



Imagination Reduces False Memories for Everyday Action Sentences: Evidence From Pragmatic Inferences

María J. Maraver*, Ana Lapa, Leonel Garcia-Marques, Paula Carneiro and Ana Raposo

CIGPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal

OPEN ACCESS

Edited by:

Rui Paulo,
Bath Spa University, United Kingdom

Reviewed by:

Marie Geurten,
University of Liège, Belgium
Marek Nieznanski,
Cardinal Stefan Wyszyński University,
Poland
Naveen Kashyap,
Indian Institute of Technology
Guwahati, India

*Correspondence:

María J. Maraver
mjmaraver@psicologia.ulisboa.pt

Specialty section:

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

Received: 17 February 2021

Accepted: 26 July 2021

Published: 20 August 2021

Citation:

Maraver MJ, Lapa A,
Garcia-Marques L, Carneiro P and
Raposo A (2021) Imagination
Reduces False Memories for
Everyday Action Sentences: Evidence
From Pragmatic Inferences.
Front. Psychol. 12:668899.
doi: 10.3389/fpsyg.2021.668899

Human memory can be unreliable, and when reading a sentence with a pragmatic implication, such as “the karate champion hit the cinder block,” people often falsely remember that the karate champion “broke” the cinder block. Yet, research has shown that encoding instructions affect the false memories we form. On the one hand, instructing participants to imagine themselves manipulating the to-be-recalled items increase false memories (imagination inflation effect). But on the other hand, instructions to imagine have reduced false memories in the DRM paradigm (imagination facilitation effect). Here, we explored the effect of imaginal encoding with pragmatic inferences, a way to study false memories for information about everyday actions. Across two experiments, we manipulated imaginal encoding through the instructions given to participants and the after-item filler task (none vs. math operations). In Experiment 1, participants were either assigned to the encoding condition of imagine + no filler; pay attention + math; or memorize + math. In Experiment 2, the encoding instructions (imagine vs. memorize) and the filler task (none vs. math) were compared across four separate conditions. Results from the two experiments showed that imagination instructions lead to better memory, by showing a higher proportion of correct responses and better performance in a memory benefit index. Similarly, a significant reduction of false memories was observed across both experiments, even though a complementary Bayesian analysis only supported this conclusion for Experiment 1. The findings show that imaginal encoding improves memory, suggesting the engagement of a distinctiveness heuristic and source-monitoring process.

Keywords: false memories, pragmatic inferences, imagination, retrieval, memory

INTRODUCTION

Human memory can be untrustworthy. When reading the sentence “the karate champion hit the cinder block” we might very often infer that the cinder block was broken, although this outcome was not explicitly stated in the sentence. The generation of inferences depends on constructive non-intentional processes that often lead to memory errors and distortions (Carpenter and Schacter, 2017). The use of sentences embedded with pragmatic implications, such as the previous example is thus a useful way to induce false memories for everyday

actions and to study the reconstructive nature of memory. This article explores the consequences of different encoding instructions in memory retrieval for information on day-to-day actions using the pragmatic inference paradigm. Given the damaging consequences of inaccurate memories, it is of great interest to understand the mechanisms behind memory for pragmatic inferences.

In pragmatic inference sentences, an implication normally occurs when, from the information presented, the reader expects something that was not explicitly stated or not necessarily implied, changing the original meaning of the sentence. In the example above, the sentence pragmatically implies that the cinder block was broken, although this consequence was never made explicit. A strategy to test if a sentence implies a pragmatic inference is whether it can be joined by a “but not” conjunction and result in a consistent sentence. That is, “*the karate champion hit the cinder block, but did not break it.*” The effectiveness of pragmatic inferences in eliciting false memories is well documented, since it has been repeatedly demonstrated that participants tend to falsely recall pragmatic implications of sentences (Brewer, 1977; Chan and McDermott, 2006; Carneiro et al., 2020). Moreover, they represent a sensitive and robust measure for the study of false memories since they allow for the dissociation between the semantic and episodic memory levels. In other words, what is inferred and remembered from a character in a sentence depends, in part, on what the reader knows about the characteristics of the character (Barclay et al., 1984). In the example, the term “*champion*” implies qualities of capability and strength, which may lead the reader to infer that he/she was able to break the cinder block when, in fact, he/she might not have. Therefore, although the inference might be semantically consistent with what is presented in the sentence, it was not explicitly stated at the episodic level (Brewer, 1977; Barclay et al., 1984; McKoon and Ratcliff, 1992; Graesser et al., 1994; Raposo and Marques, 2013).

It has been demonstrated that later memory recall for pragmatic inference sentences is sensitive to manipulations in the encoding phase, such as, for example, the repetition of the materials during study (McDermott and Chan, 2006), or whether subjects encode sentences using or not a semantic strategy (Barclay et al., 1984). Thus, despite the robustness of false memories, experimental manipulations in the way items are processed during encoding can modulate retrieval in different false memory paradigms. For example, imagining that an event might have happened can increase confidence that it really happened. This *imagination inflation* effect has been demonstrated for autobiographical memories (Hyman and Pentland, 1996; Mazzoni and Memon, 2003), as well as self-performed actions or actions performed by others (Goff and Roediger, 1998; Thomas and Loftus, 2002; Lindner and Echterhoff, 2015; Calvillo et al., 2019). Winograd et al. (1998) observed a positive correlation between false recall in the DRM paradigm (Deese, 1959; Roediger and McDermott, 1995) and individual differences in mental imagery (measured by the Vividness of Visual Imagery Questionnaire), suggesting that the more vivid the mental images, the more likely to be falsely remembered. The authors explain this effect by difficulties in source monitoring between

the words presented externally and the internally generated images (Winograd et al., 1998). In contrast to the imagination inflation effect, several reports have demonstrated an effect of *imagination facilitation*. That is, imaginal encoding can increase correct recall and reduce memory errors. Imaginal encoding typically improves memory (Paivio, 1991), and specifically for false memories, deliberately generating images reduces false memories compared to a control condition in the DRM paradigm (Foley et al., 2006; Robin, 2011; Oliver et al., 2016). Of note, in the DRM, false memories are elicited by relational processing (i.e., a list level or inter-item effect; Hege and Dodson, 2004; Huff et al., 2020). Generating images during the encoding of DRM items (associative word lists or objects) seem to improve item-specific processing, thus decreasing the chances of activating the critical item (non-presented lure) during encoding, and enhancing retrieval of studied items, both in recall and recognition tests (Robin, 2011; Robin and Mahé, 2015). Moreover, according to Foley et al. (2006), imaginary strategies at encoding allow participants to benefit from a richer context for successful monitoring, which results in improved veridical memories and reduced false memories. Still other studies have shown that imaginal encoding has no effect on false recall (Newstead and Newstead, 1998).

Perhaps, the different outcomes that have been reported for the effect of imaginal encoding on false memories – in particular the imagination inflation and the imagination facilitation effects – might be explained by differences between the paradigms where this encoding strategy was employed. As described above, the imagination inflation effect was found when the imaginal encoding strategy was employed for autobiographical events and actions, while the imagination facilitation effect was found when such strategy was employed for verbal materials, such as the associative wordlists used in the DRM paradigm. Considering these divergent patterns of results across different false memory paradigms, it is of great interest to compile evidence and explore the impact of imaginal encoding using pragmatic inference sentences. This paradigm involves verbal information describing everyday events and actions. Similarly to the DRM, it elicits false memories for non-presented material. Yet, while the DRM effects occur at the inter-item (list) level, pragmatic inference effects do not rely on relational processing, occurring at an intra-item level.

In the present study, we wanted to extend the existing knowledge on the effects of different encoding instructions on false memories using a paradigm that resembles the type of memory errors generated for everyday actions. Here, we investigated how an imagination strategy at encoding affected false memories for information about everyday actions produced by pragmatic inferences. For that purpose, sixty pragmatic inference sentences were used. These sentences were presented in Portuguese since pragmatic inferences have been shown to be culture and language specific (Carneiro et al., 2020). Responses were coded following the standard criteria proposed by Brewer (1977), as detailed in the methods' section. Across two experiments, the effect of imagination was studied by manipulating the instructions given to participants at the encoding phase as well as the after-item filler task (none, to

allow time for participants to imagine vs. solving math operations). By manipulating the instruction task, we contrasted our experimental condition (imagination) against two control conditions: a memory condition, akin to those commonly used in memory research and an instruction to simply pay attention, to serve as a baseline. After encoding, participants either engaged in a no filler task, allowing elaborative rehearsal and the engagement in deeper processing or, alternatively, they performed math operations, which constrained rehearsal and elaboration (Craik and Watkins, 1973). This manipulation of the filler task is important as imaginal encoding is associated with rehearsal and elaboration, and as such it may depend on the time available for such processes to take place. Indeed, similar manipulations of the filler task have been repeatedly used for the study of different levels of processing in memory tasks (Lewandowsky and Oberauer, 2015; Bartsch and Oberauer, 2021), and specifically in false recall paradigms (Rhodes and Anastasi, 2000). Experiment 1 was a laboratory-based experiment, where participants were randomly assigned to one out of three conditions: imagine+no filler; pay attention+math; or memorize+math. In Experiment 2, we manipulated the encoding instructions (imagine vs. memorize) and filler tasks (none vs. math) orthogonally. Participants were randomly assigned to one of four conditions: imagine+no filler; imagine+math; memorize+no filler; or memorize+math. Based on the revised literature, we expected imagination to significantly affect false memory performance compared to our control instructions to memorize or pay attention to the sentences. We had two contrasting possible hypotheses regarding the direction of this effect. On the one side, one could expect imagination encoding to increase the proportion of false memories compared to the control encodings, as previously found for the performance of actions (Goff and Roediger, 1998; Thomas and Loftus, 2002; Lindner and Echterhoff, 2015; Calvillo et al., 2019). This imagination inflation effect would reflect an increased proneness to source-monitoring errors, due to the difficulty in differentiating between externally presented information (i.e., the information presented in the sentence) and internally generated images (i.e., imagining the information that is pragmatically inferred). Alternatively, as shown in the DRM, an instruction to imagine could reduce false memories, by promoting item-specific processing and decreasing the probability of generating a pragmatic inference during encoding. Moreover, the instruction to imagine may lead to a more deliberate consideration of what was presented (i.e., why the sentence included “hit” instead of “broke”), which could facilitate memory and monitoring compared to the control conditions. One possible mechanism supporting this could be the use of a distinctiveness heuristic (Dodson and Schacter, 2002). That is, the generation of images at encoding could provide more distinctive diagnostic cues that would help making monitoring decisions at the moment of retrieval (Dodson and Schacter, 2002; Hege and Dodson, 2004). The use of the distinctive heuristic for the reduction of false memories has been repeatedly demonstrated in the DRM paradigm (Dodson and Schacter, 2002; Hege and Dodson, 2004; Foley et al., 2006; Oliver et al., 2016). However, it remains unexplored how different encoding contexts would affect

pragmatic inferences, a paradigm in which false memories are not generated from relational processing and resembles closely the type of memory errors that occur in everyday actions. Therefore, with the present study, we aimed to extend the current understanding on the effects of different encoding conditions in memory errors and to clarify the role of imagination in increasing or reducing errors generated by pragmatic inferences.

EXPERIMENT 1

Materials and Methods

Sample

A total of 120 participants (87 females; $M_{\text{age}} = 25.63 \pm 7.76$) voluntarily agreed to participate in a laboratory experiment and were rewarded with 10€ gift vouchers for their time. For both experiments, the sample size was determined based on the number of participants used in previous studies with similar designs (Carneiro et al., 2017, 2020; Soro et al., 2017) and depending on the availability of resources. Participants were randomly assigned to one out of three conditions based on the encoding instructions: imagine, memorize, or pay attention. All participants were provided with information from the study and gave informed consent according with the Declaration of Helsinki (World Health Organisation, 2013). All experimental procedures were approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon.

Materials and Procedure

Sixty pragmatic implication sentences in Portuguese adapted by Carneiro et al. (2020) were used in the current experiment. Participants performed their task individually in a laboratory computer and the experiment was programmed and ran in the online platform Qualtrics (Qualtrics, Provo, UT). Participants were randomly distributed to the experimental conditions. Likewise, the order of the sentences at the encoding and retrieval phases was randomized anew for each participant. Participants first performed a practice block of the encoding phase with five sentences (without pragmatic implications). After, during the encoding phase, participants were presented with individual sentences in the computer screen for 4.5s and were instructed to read and either memorize, imagine, or pay attention to them (i.e., “*The karate champion hit the cinder block*”). After each sentence, participants in the imagine condition were presented with a blank screen for 5s, while participants in the pay attention or memorize conditions had to solve a simple math operation for 5s (i.e., $25 - 9 = ?$). We included a blank screen after the imagine condition to ensure that participants had enough time to engage in mental imagery. This is important given the well-established individual differences in the ability to generate mental images (Marks, 1973; Cui et al., 2007; Pearson et al., 2011). In the middle of the encoding phase, immediately after presentation of the 30th sentence, participants could take a self-paced break, before continuing with the encoding of the remaining sentences. To allow for this break, sentences were counterbalanced in two

blocks of 30 sentences each. Two sets of materials were created to counterbalance the sentences across these two blocks. Sentences were randomized within each block. After the encoding of all 60 sentences, participants performed a distractor task for 5 min, where they were asked to count the differences between four pairs of images. Finally, participants in all conditions performed a cued-recall test where the 60 sentences were presented, in a random order, without its critical words and participants had to fill in the gaps (i.e., “*The karate champion ____ the cinder block*”). The experiment lasted, on average, 40 min to be completed. After completion of this experiment, participants took part in two additional experiments (independent of the current study) for 20 min, totaling a 60 min experimental session, after which they were thanked and rewarded for their time.

Participants’ responses were recoded following an adaptation of the standard criterion proposed by Brewer (1977). Answers that match the original sentence or synonyms that maintained the meaning of the original sentence were considered correct responses, answers that matched the expected pragmatic inferences or their synonyms were considered pragmatic inferences responses, and other alternative answers were considered intrusions. Responses left blank were considered as omissions. For example, for the sentence “*the karate champion hit the cinder block*,” responses were classified as either: (a) correct responses (i.e., “*hit*”); (b) pragmatic inferences (i.e., “*broke/destroyed/crashed*”); (c) intrusions (i.e., “*kicked/lifted/moved*”); and (d) omissions (no response). **Supplementary Table 1** includes the complete list of the 60 sentences used as well as their coding criteria.

We calculated the proportion of recall for the four different response types (correct responses, pragmatic inferences, intrusions, and omissions) by adding the number of each response type and dividing them by the total number of sentences (60). **Table 1** summarizes the proportion of recall for each encoding condition group and response type.

Results

Response Type

Statistical analyses were performed using IBM SPSS Statistics software version 25 (IBM Corp., Armonk, NY, United States). Because the four dependent variables (i.e., correct responses, pragmatic inferences, intrusions, and omissions) have different statistical distributions, we ran separate one-way ANOVAs for

each response type, with the proportion of recall as the dependent variable and the encoding condition as the independent factor (imagine vs. memorize vs. pay attention). Results were further analyzed within the Bayesian hypothesis testing framework to quantify the evidence for differences between the conditions (H_1) in favor of no differences (H_0). Bayesian ANOVAs and t -tests were run in JASP (JASP Team, 2020) using the default settings [r scale fixed effects=0.5, r scale random effects=1, samples=auto (10,000)]. In short, according to the Lee and Wagenmakers’ classification (Ly et al., 2016), Bayes factors (BF_{10}) above 3 provide evidence in favor of H_1 , below 0.3 support H_0 , while intermediate values are interpreted as inconclusive (van Doorn et al., 2019).

Results for correct responses revealed a main effect of encoding condition, $F(2,117)=28.89$, $p<0.01$, $\eta_p^2=0.33$, $BF_{10}=1.41e+8$. *Post-hoc* Bonferroni comparisons showed that the group instructed to imagine ($M=0.41\pm0.16$) recalled a higher proportion of correct responses than the groups instructed to memorize [$M=0.20\pm0.11$; $p<0.01$; 95% CI (0.14–0.28)] and to pay attention [$M=0.25\pm0.12$; $p<0.01$; 95% CI (0.09–0.23)]. No differences in correct responses were observed between the instruction to memorize and pay attention [$p=0.26$; 95% CI (–0.12–0.02)].

Regarding memory errors, results for pragmatic inferences responses revealed a main effect of encoding condition, $F(2,117)=17.83$, $p<0.01$, $\eta_p^2=0.23$, $BF_{10}=82080.45$. *Post-hoc* Bonferroni comparisons showed that the group instructed to imagine ($M=0.36\pm0.09$) recalled a lower proportion of pragmatic inferences than the groups instructed to memorize [$M=0.49\pm0.09$; $p<0.01$; 95% CI (–0.18––0.07)] and to pay attention [$M=0.46\pm0.11$; $p<0.01$; 95% CI (–0.15––0.04)]. No difference in the recall of pragmatic implications was observed between the instruction to memorize and pay attention [$p=0.55$; 95% CI (–0.02–0.08)].

As for intrusions, results showed that the evidence for an effect of encoding condition was inconclusive, $F(2, 117)=2.78$, $p=0.06$, $\eta_p^2=0.04$, $BF_{10}=0.78$; and results for omissions provided moderate evidence for no differences between the encoding conditions, $F(2, 117)=1.18$, $p=0.31$, $\eta_p^2=0.20$, $BF_{10}=0.21$.

Memory Benefit Index

In addition to the analyses of the different response types, we derived a memory benefit index which captures the size

TABLE 1 | Mean proportions and standard deviations of the cued-recall tests for correct responses, pragmatic implications, omissions, and intrusions for the two experiments as a function of the experimental conditions.

	<i>N</i>	Correct responses	Pragmatic inferences	Intrusions	Omissions
Experiment 1					
Imagine + no filler	40	0.41 ± 0.16	0.36 ± 0.09	0.02 ± 0.03	0.05 ± 0.08
Memorize + math	40	0.20 ± 0.11	0.49 ± 0.09	0.03 ± 0.02	0.06 ± 0.13
Pay attention + math	40	0.25 ± 0.12	0.46 ± 0.11	0.02 ± 0.02	0.09 ± 0.12
Experiment 2					
Imagine + no filler	27	0.46 ± 0.14	0.38 ± 0.12	0.09 ± 0.05	0.07 ± 0.14
Imagine + math	29	0.34 ± 0.15	0.44 ± 0.12	0.12 ± 0.05	0.09 ± 0.10
Memorize + no filler	28	0.32 ± 0.15	0.45 ± 0.13	0.14 ± 0.08	0.08 ± 0.12
Memorize + math	29	0.29 ± 0.13	0.48 ± 0.14	0.14 ± 0.09	0.08 ± 0.11

of the difference between correct recall and recall of pragmatic implications for each sentence. We computed this index by subtracting the proportion of pragmatic inferences from the proportion of correct responses and this difference was divided by the overall recall (i.e., $\text{Correct responses} - \text{Pragmatic inferences} / (\text{Correct responses} + \text{Pragmatic inferences})$). This index captures the effect of error production relative to correct recall, while controlling for overall memory performance. It is thus a more fine-grained measure of memory accuracy as it is independent of the total number of responses generated. A higher score means greater memory accuracy (with higher veridical recall and lower pragmatic inferences committed), and a lower score reflects lower memory accuracy (with lower correct recall and higher pragmatic inferences committed). This variable was introduced in a one-way ANOVA with encoding condition (imagine vs. memorize vs. pay attention) as the between-groups factor. Results revealed a main effect of condition, $F(2,117) = 29.34$, $p < 0.01$, $\eta_p^2 = 0.33$, $BF_{10} = 2.72e + 8$, and *post-hoc* Bonferroni comparisons showed that the group instructed to imagine ($M = 0.04 \pm 0.32$) outperformed the groups instructed to memorize [$M = -0.43 \pm 0.24$; $p < 0.01$; 95% CI (0.32–0.63)] and to pay attention [$M = -0.30 \pm 0.29$; $p < 0.01$; 95% CI (0.19–0.50)]. No differences in memory performance were observed between the instruction to memorize and pay attention [$p = 0.13$; 95% CI (–2.82–0.25)].

Discussion

Results show that the instruction to imagine benefitted memory performance compared to the instructions to memorize and pay attention. Participants showed a higher proportion of correct responses, and a lower proportion of pragmatic inference errors. We should note, however, that the group instructed to imagine had a longer time to rehearse because of the presentation of a blank screen after each sentence, while the memorize and pay attention groups had to perform math operations after each sentence and thus were exposed to a harder environment for rehearsal. To determine if the benefit of imagination comes from the longer time to rehearse or from a reliable effect of the instruction to imagine, we ran a second experiment in which we improved methodological control and compared four different conditions resulting from the crossing of two factors: the encoding instruction (memorize vs. imagine) and the filler task (none vs. math operations). For this second experiment, we decided to exclude the pay attention condition since it held no significant differences from the typical memory instruction.

EXPERIMENT 2

Materials and Methods

Sample

A total of 179 university students from the Faculty of Psychology of the University of Lisbon participated in the study and received course credits for compensation. The first 29 participants were tested in the experimental laboratory of the Faculty of

Psychology. Due to COVID-19 outbreak and subsequent national lockdown, data collection moved to online for the remaining 150 participants. For online data collection and following previous work (Carneiro et al., 2020), several attention checks were introduced across the experiment (see procedure below) to guarantee that participants completed the task successfully and attentively. Sixteen participants failed one or more attention checks during cued recall, and their responses were excluded from further analysis. Moreover, other 50 responses had to be excluded for the following reasons: 40 participants indicated that they were not native speakers of European Portuguese, and 10 participants did not complete the experiment or were interrupted during the session. Therefore, the final sample consisted of 113 valid participants (85 online and 28 from the laboratory; 94 females; $M_{\text{age}} = 20.65 \pm 6.37$). To rule out possible differences in memory performance as a function of the testing setting, we ran independent samples *t*-test in the four dependent variables. Despite the unbalanced distributions of participants across settings (28 in the laboratory vs. 85 online), results provide evidence for no differences between participants tested in the laboratory and online: correct responses: $t(111) = 0.21$, $p = 0.83$, 95% CI (–0.06–0.07), $BF_{10} = 0.23$; pragmatic inferences: $t(111) = -0.35$, $p = 0.72$, 95% CI (–0.07–0.05), $BF_{10} = 0.24$; intrusions: $t(111) = 0.25$, $p = 0.80$, 95% CI (–0.03–0.03), $BF_{10} = 0.23$; and omissions: $t(111) = -0.03$, $p = 0.97$, 95% CI (–0.05–0.05), $BF_{10} = 0.23$.

Participants were randomly assigned to one out of four conditions resulting from the factorial combination of encoding instructions and filler tasks: imagine + no filler; imagine + math; memorize + no filler; or memorize + math. Similarly, all participants were provided with information from the study and gave written informed consent. All procedures were approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon.

Procedure

Materials and procedure were similar to those used in Experiment 1 with the following exceptions. First, online participants were instructed to perform the task in a quiet environment without interruptions. Second, several attention checks were introduced across the study to guarantee successful performance in the task (following Carneiro et al., 2020). At the encoding phase, participants either read instructions to imagine or to memorize the sentences and, after each sentence, were either presented with a blank screen for 5 s or a math problem to solve, according to the condition. Two attention checks were presented at the encoding phase. These consisted in pressing an arrow button under 10 s to resume the presentation of the sentences (a timer was displayed). To allow for these attention checks to appear amid encoding, the sentences were counterbalanced across three blocks of 20 sentences each: the first attention check appeared at the end of the first block, after the 20th sentence; the small break appeared amid the second block, after the 30th sentence; and the second attention check appeared before the final block, after the 40th sentence. Three sets of materials were created to counterbalance the

sentences across the three blocks of the encoding phase, a first block of 20 sentences, a second block of 20 divided by a break after the 10th sentence, and a final block of 20 sentences. Within each block, sentences were presented in a randomized order. The distractor task and test phase were the same as those in the previous experiment, but three additional attention checks were added at test: these consisted of fragmented sentences, similar to those presented for recall, but explicitly asking participants to write a given word (seven, backpack, and red) in the response field. At the end of the experiment, participants were asked to rate their level of attention ($M=5.80\pm0.75$) and the quality of their data ($M=5.15\pm1.16$) in a 7-point Likert scale. The presentation order for the fragmented sentences at test was randomized. Experiment 2 took, on average, 40 min to be completed.

Results

An independent samples *t*-test revealed no difference in the overall math accuracy between the memorize ($M=0.75\pm0.10$) and imagine ($M=0.74\pm0.09$) encoding conditions, $t(56)=0.25$; $p=0.76$; 95% CI $(-0.06-0.04)$, $BF_{10}=0.27$.

Akin to Experiment 1, responses of the final cued-recall test were recoded and classified using an adaptation of the standard criterion proposed by Brewer (1977; see **Supplementary Table 1**).

Response Type

We conducted four separate two-way ANOVAs with 2 (encoding condition: imagine vs. memorize) $\times$ 2 (filler task: no filler vs. math), one for each dependent variable. Performance on the different variables as a function of condition is included in **Table 1**.

Results for correct responses revealed a main effect of encoding condition, $F(1,109)=12.56$, $p<0.01$, $\eta_p^2=0.10$, $BF_{10}=41.78$, suggesting that the groups instructed to imagine ($M=0.40\pm0.16$) recalled a higher proportion of correct responses than the groups instructed to memorize [$M=0.30\pm0.14$; $p<0.01$; 95% CI $(0.04-0.15)$]. There was also a main effect of filler task, $F(1,109)=7.65$, $p<0.01$, $\eta_p^2=0.06$, $BF_{10}=6.59$, with no filler ($M=0.39\pm0.16$) leading to a better performance than doing math operations [$M=0.31\pm0.14$; $p=0.01$; 95% CI $(0.01-0.13)$]. Results showed moderate evidence for an interaction between the factors, $F(1,109)=2.92$, $p=0.09$, $\eta_p^2=0.03$, $BF_{10}=3.31$. This means that while for participants solving math operations there is no conclusive evidence for a difference in the proportion of correct responses between the instruction to imagine ($M=0.34\pm0.15$) and to memorize ($M=0.29\pm0.13$), $t(56)=1.32$, $p=0.19$, 95% CI $(-0.17-0.86)$, $BF_{10}=0.55$; for participants that saw a blank screen after each sentence, those in the imagine condition had a higher proportion of correct responses ($M=0.46\pm0.14$), than instructed to memorize ($M=0.32\pm0.15$), $t(56)=3.65$, $p<0.01$, 95% CI $(0.42-1.54)$, $BF_{10}=47.89$.

For pragmatic inferences responses, results revealed a significant main effect of encoding condition, $F(1,109)=5.20$, $p=0.02$, $\eta_p^2=0.05$, $BF_{10}=1.48$, suggesting that the group

instructed to imagine ($M=0.41\pm0.12$) committed a significantly lower proportion of pragmatic inference errors than the groups instructed to memorize [$M=0.47\pm0.17$; $p=0.03$; 95% CI $(-0.10-0.00)$]. A main effect of filler task was also observed, $F(1, 109)=4.48$, $p=0.04$, $\eta_p^2=0.04$, $BF_{10}=1.13$; with the no filler condition ($M=0.41\pm0.13$) leading to lower levels of pragmatic inference errors than doing math operations [$M=0.46\pm0.13$; $p=0.04$; 95% CI $(0.00-0.10)$]. The interaction between encoding condition and filler task did not reach significance, $F(1,109)=0.50$, $p=0.48$, $\eta_p^2<0.01$, $BF_{10}=0.51$. Yet, in Bayesian analyses, the evidence for these effects is deemed inconclusive, meaning that, with the current sample size, we did not have enough power to provide evidence for reliable effects (Ly et al., 2016; Quintana and Williams, 2018).

Lastly, results for intrusions revealed a main effect of encoding condition, $F(1, 109)=10.31$, $p<0.01$, $\eta_p^2=0.08$, $BF_{10}=12.17$, revealing that the group instructed to imagine ($M=0.10\pm0.05$) committed a lower proportion of intrusion errors than the group instructed to memorize [$M=0.14\pm0.08$; $p<0.01$; 95% CI $(-0.06-0.02)$]. No other significant results were found (all $ps>0.29$, $BF_{10}<0.37$). No significant differences were found for omissions (all $ps>0.47$, $BF_{10}<0.15$).

Memory Benefit Index

Similar to Experiment 1, we computed a memory benefit index that was analyzed in a two-way ANOVA, with 2 (encoding instruction: imagine vs. memorize) $\times$ (filler task: no filler vs. math) as between group factors. Consistent with Experiment 1, results revealed a main effect of encoding instruction, $F(1,109)=9.77$, $p<0.01$, $\eta_p^2=0.08$, $BF_{10}=12.04$, meaning that participants instructed to imagine ($M=-0.03\pm0.33$) outperformed those instructed to memorize [$M=-0.21\pm0.33$; $p<0.01$; 95% CI $(0.07-0.31)$]. Results also revealed a main effect of filler task, $F(1,109)=7.08$, $p=0.01$, $\eta_p^2=0.06$, $BF_{10}=4.45$, such that performing math operations after encoding each of the sentences led to worse memory performance ($M=-0.20\pm0.34$) than to stare at a blank screen without doing any explicit task for 5 s [$M=-0.04\pm0.32$; $p<0.01$; 95% CI $(-0.28-0.04)$]. Finally, the interaction between the factors did not reach the level of significance, with the Bayesian analysis revealing inconclusive evidence for the absence or presence of an interaction, $F(1,109)=2.18$, $p=0.14$, $\eta_p^2=0.02$, $BF_{10}=1.90$.

Discussion

Experiment 2 aimed at disentangling whether the benefits that imagination had on memory in Experiment 1 were solely due to this encoding strategy or if these benefits were promoted by the task participants carried out after the presentation of each sentence (no filler task, allowing for rehearsal, vs. math operations). To do so, in Experiment 2, both filler tasks were manipulated across different encoding strategies. Results were partly consistent with Experiment 1: imaginal encoding seems to be beneficial for memory, compared to a memory encoding strategy. The instruction to imagine promoted an overall better performance, with higher levels of correct responses and lower levels of intrusion errors, compared to an instruction to memorize.

GENERAL DISCUSSION

The purpose of this research was to explore the effect of imagination as an encoding strategy on false memories induced by pragmatic inference sentences. In Experiment 1, we observed that instructions to imagine improved overall memory performance, by enhancing the level of correct responses recalled and by reducing the proportion of pragmatic inference errors committed, compared to other encoding strategies. However, given the methodological confound derived from the longer time to rehearse in the imagination group against the other conditions, caution is warranted in the interpretation of the results. Experiment 2 addressed this methodological limitation and revealed an effect of imagination in increasing the proportion of correct responses and improving overall performance (memory benefit index), compared to the memorize condition. In addition, a significant reduction in pragmatic inference errors was also observed in the imagination condition, indicating a decrease in false memories. Yet, while results from the Bayesian framework provided support for the reduction of pragmatic inference errors in Experiment 1, results from Experiment 2 suggested that with the current sample size, we did not have enough power to claim for such an effect in the imagine, compared to the memorize condition.

Regarding the differences as a function of the filler task, we observed that participants in the imagine condition had a higher proportion of correct recall compared to the memory condition, when they saw a blank screen after each sentence. We must note that the blank screen displayed after each sentence had a fixed presentation time of 5 s. Therefore, although participants had the same time to rehearse the sentence, it was in the imagine condition where rehearsal benefitted subsequent memory recall. Furthermore, the overall performance in solving math equations did not differ between the instruction to imagine or to memorize, suggesting that the amount of time and elaboration was the same for both encoding conditions, and it was the combination of the instruction to imagine and the time to rehearse that led to more accurate recall (Bower, 1970).

Imagination has long been recognized for its enhancing effects on veridical memory (Paivio, 1991), and imaginal encoding strategies are often used as a mnemonic aid that can lead to increases in the amount of information that can be stored and retrieved (Paivio, 1991). The effectiveness of instructions that invite participants to use imaginal encoding strategies in verbal learning has generally provided consistent, reliable, and substantial improvements in tests related to retention, recall performance, and recognition (Richardson, 1998). According to the DRM literature, imagination is also beneficial by reducing false memories: When participants are instructed to imagine objects that correspond to the presented word-list items or to images presented on the screen, veridical memory is improved, and false memories are reduced compared to control instructions (Hege and Dodson, 2004; Foley et al., 2006, 2009; Oliver et al., 2016). Our results replicated this imagination facilitation effect and extended it to pragmatical implications on everyday actions: Imaginal encoding improved veridical memory of the stated information.

Several explanations can be proposed for this result. On the one hand, according to the impoverished relational-encoding view, imaginal encoding improves item-specific processing and reduces relational processing of information and thus false memories (Hege and Dodson, 2004; Foley et al., 2006; Oliver et al., 2016). According to Hege and Dodson (2004), encoding images interfere with the encoding of relational information – the main cause of false memories in the DRM – so that critical lures are less likely to be falsely recalled at the final memory test, resulting in an improved memory performance. In our experimental design, it was not possible to directly measure what people were actually imagining, in order to assess if the generated images matched what was explicitly stated in the sentence. Yet, the observed imagination facilitation effect suggests that indeed, instructing participants to imagine the sentences promote item-specific processing and attenuate the probability of generating pragmatic inferences. It would be interesting for future studies to ask participants what they imagined and evaluate whether relational information had been encoded or not.

On the other hand, imagination instructions seem to promote encoding of the presented material, by providing more specific characteristics or diagnostic cues (Foley et al., 2009; Robin, 2011). These cues can later be used to make better source-monitoring decisions, that is, to accept (veridical) information about which one has more mental images and to disregard (wrong) information that lacks such cues. This process may rely on the distinctiveness heuristic (Hege and Dodson, 2004; McCabe and Smith, 2006; Oliver et al., 2016), by which source-monitoring decisions would be guided by distinctiveness. Because the non-imagined material lacks distinctive details, it would, therefore, be rejected (Foley et al., 2009). This interpretation is supported by the results of Rajaram's study showing that the recognition of pictures compared to words is based on the retrieval of distinctive features from memory rather than the familiarity of the events (Rajaram, 1993). Yet, results from Robin (2011) showed that, even though imaginal encoding reduces false memories in free recall tests, it does not reduce false memories in recognition tests, suggesting that distinctiveness was not at play. According to Robin (2011), the benefits of imaginal instructions for memory could either stem from an enhancement of the specific characteristic of the information encoded that will act as cues, facilitating veridical retrieval and only by that avoiding errors or from the use of the distinctiveness heuristic only when participants can benefit from appropriate contextual support – as it is the case of cued-recall tests but not for recognition tests where participants are exposed to false information.

Besides this, the distinctiveness heuristic is an inferential strategy that comes into play during retrieval, when an individual fails to remember sufficient information about a past event (Dodson and Schacter, 2002; Hege and Dodson, 2004). The distinctiveness heuristic depends on individual source-monitoring processes and on metamemorial beliefs of what should be remembered (Johnson et al., 1993; Dodson and Schacter, 2002). In this regard, failing to remember inferable information about an event (expected, but not necessarily veridical) can

sign that the event never occurred. Dodson and Schacter (2002) demonstrated that encoding images biased toward the use of a distinctiveness heuristic and, in fact, previous research as shown that the use of this heuristic reduces false memories both in recognition and in recall tests (McCabe and Smith, 2006; Robin, 2011). In our case, we believe that the experience of imagination provides at encoding rich information details, regarding both the stimuli and the residual traces of the mental operations performed at encoding that can be used, promoting item-specific processing with distinctive and diagnostic details. This leads to richer and more detailed representations in memory for the imagined action and, at retrieval, the use of the distinctiveness heuristic – as a metamemorial decision-based strategy – allows the exclusion of items that are recalled in the absence of such details. Therefore, imaginal encoding could lead to a more careful dissociation between explicitly stated vs. inferable information.

The Fuzzy Trace Theory (FTT) provides an alternative but complementary explanation to our findings since it distinguishes between the parallel encoding of two different types of memory traces. On the one hand, gist traces encode general meaning-based representations, and on the other, verbatim traces encode superficial representations based on perception. According to the FTT, memory performance is based on the retrieval of both verbatim and gist traces. While both types of traces can support the accurate memory reconstruction, gist-irrelevant information can co-occur with the activation of gist traces, but they are suppressed by the retrieval of verbatim traces (Reyna and Kiernan, 1994; Brainerd and Reyna, 2002). Thus, the memory errors generated by pragmatic inference sentences would be consistent with the retrieval of gist traces, since they represent a deductive interpretation of concepts (meanings, relations, and patterns) that integrate world knowledge with textual information (Kintsch, 1974; Reyna and Kiernan, 1994). In this sense, in our study, the imaginary activity generated at encoding promoted the processing of verbatim traces (i.e., episodically instantiated representations of the presented items), which resulted in the imagination facilitation effect, leading to an increase in correct responses and a reduction in pragmatic inference errors.

The imagination facilitation effect is very robust for DRM-induced false memories and, supported by our results, for pragmatic inference sentences. Nonetheless, there are reports of imagination inflation effects with other memory paradigms (Hyman and Pentland, 1996; Goff and Roediger, 1998; Mazzoni and Memon, 2003; Lindner and Echterhoff, 2015; Calvillo et al., 2019). Such discrepant results may be explained by differences in the paradigms employed. As mentioned at the outset, while the imagination facilitation effect was found using the DRM paradigm, that relies on verbal material and their semantic association, the imagination inflation effect was found for autobiographical events and actions. That is one of the reasons that makes the use of pragmatic inference sentences particularly interesting: These materials describe everyday actions and events but rely on semantic extraction and association; and yet the imagination facilitation effect persisted.

Another possible approach to the discrepant result patterns, less parsimonious but perhaps more conciliatory, is that these

findings may reflect two sides of the same coin. As it was already stated, imagination leads to a better retrieval of the studied information (Paivio, 1991; Richardson, 1998; Foley et al., 2006). Perhaps imagination increases overall the retrieval of the imagined information – whether true (explicitly studied) or false (not presented). In the DRM paradigm and in the pragmatic inference paradigm employed here, participants are asked to imagine the presented words and sentences, respectively. So, presumably, they imagined veridical information, which in turn promoted veridical retrieval and reduced false memories for non-imagined lures or pragmatic implications. However, in the cases where an imagination inflation effect was found, participants were asked to imagine events and actions that they did not perform and were thus considered as false (not presented) information. Perhaps both inflation and facilitation effects reflect an increase of retrieval for all imagined information, true or false, as long as imagined. Our results provide evidence for a beneficial effect of imagination in memory using pragmatic inferences. However, caution should be warranted when generalizing our results to other materials and paradigms. Although we consider that in our data no additional analysis could allow us to differentiate between imagination inflation and facilitation, it might be interesting for future studies to compare memory performance between different types of pragmatic inference sentences, for example, manipulating the agent (i.e., third vs. first person), as this has been argued to be a critical factor in false memory paradigms on action performance (Lindner and Echterhoff, 2015). This would allow testing the extent to which our results (a facilitation effect) generalize to conditions similar to those used in other paradigms.

Some limitations of the current study are worth acknowledging. First, a subset of participants in Experiment 2 was tested in a different setting, and results from Bayesian analyses suggest that the sample size was not large enough to show conclusive results. Second, the level of imaginability of each sentence was not considered. Future studies should explore the impact of each stimulus' imaginary value and differentiate between high and low imaginability sentences. Besides, it is important for future studies to consider individual mental imagery abilities, since we could expect that differences in imagery abilities might explain whether the instruction to imagine results in a positive or negative effect on memory performance. Third, participants' confidence level was not measured, a factor that has been shown to modulate the proportion of false memories (Brewer et al., 2005). It might be interesting for future research to assess the confidence levels of both imagination (how accurate they think their generated image was) and retrieval (how confident they are of their response at the final memory test). Finally, the substantial reduction in false recognition errors that we obtained suggests that it may also be worth exploring imaginal encoding as a strategy for the correction of false memories.

Taken together, our findings suggest how an imagery strategy at encoding improves memory. Perhaps the limits of imagining are not the same as the limits of remembering: Our memories are constrained but maybe our imagination is free (McCarroll, 2020). Imaginary encoding surely seems

to be a key strategy to create better learning environments, less prone to false memories, and empowering of our veridical memories.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the Open Science Framework online repository: https://osf.io/v8apj/?view_only=9969d17536f54053a72be19c050c4767. The names of the repository/repositories and accession number(s) can also be found in the article/Supplementary Material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PC conceived the research question and the research design, supervised the data collection, and conducted the preliminary data analysis of Experiment 1. LG-M, MM, AL, and AR developed the research design of Experiment 2. AL programmed and supervised the data collection of Experiment 2. MM conducted

the data analysis of Experiments 1 and 2 and wrote the first draft of the manuscript. All authors were involved in the discussion of the theoretical implications of this research. MM, AL, LG-M, and AR revised and approved the final version of the manuscript. PC deceased prior to the development of Experiment 2 and finalization of the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Fundação para a Ciência e Tecnologia (Foundation for Science and Technology of Portugal) through a project grant (PTDC/PSI-ESP/28414/2017) to PC and AR and a Ph.D. Studentship (PD/BD/114066/2015) to AL.

ACKNOWLEDGMENTS

We would like to thank André Vaz for his help in implementing Experiment 1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.668899/full#supplementary-material>

REFERENCES

- Barclay, C. R., Toglia, M. P., and Chevalier, D. S. (1984). Pragmatic inferences and type of processing. *Am. J. Psychol.* 97, 285–296. doi: 10.2307/1422601
- Bartsch, L. M., and Oberauer, K. (2021). The effects of elaboration on working memory and long-term memory across age. *J. Mem. Lang.* 118:104215. doi: 10.1016/j.jml.2020.104215
- Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *J. Verbal Learn. Verbal Behav.* 9, 529–533. doi: 10.1016/S0022-5371(70)80096-2
- Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Curr. Dir. Psychol. Sci.* 11, 164–169. doi: 10.1111/1467-8721.00192
- Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Mem. Cogn.* 5, 673–678. doi: 10.3758/BF03197414
- Brewer, W. F., Sampaio, C., and Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *J. Mem. Lang.* 52, 618–627. doi: 10.1016/j.jml.2005.01.017
- Calvillo, D. P., Vasquez, A. N., and Pesavento, A. (2019). Imagination inflation effects are unrelated across two imagination inflation tasks. *Psychol. Conscious. Theory Res. Pract.* 6, 90–98. doi: 10.1037/cns0000178
- Carneiro, P., Garcia-Marques, L., Lapa, A., and Fernandez, A. (2017). Explaining the persistence of false memories: a proposal based on associative activation and thematic extraction. *Memory* 25, 986–998. doi: 10.1080/09658211.2016.1239742
- Carneiro, P., Lapa, A., Reis, J., and Ramos, T. (2020). Testing pragmatic inferences: the impact of language and culture. *Psicológica J.* 41, 1–20. doi: 10.2478/psicolj-2020-0001
- Carpenter, A. C., and Schacter, D. L. (2017). Flexible retrieval: when true inferences produce false memories. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 335–349. doi: 10.1037/xlm0000340
- Chan, J. C. K., and McDermott, K. B. (2006). Remembering pragmatic inferences. *Appl. Cogn. Psychol.* 20, 633–639. doi: 10.1002/acp.1215
- Craik, F. I. M., and Watkins, M. J. (1973). The role of rehearsal in short-term memory. *J. Verbal Learn. Verbal Behav.* 12, 599–607. doi: 10.1016/S0022-5371(73)80039-8
- Cui, X., Jeter, C. B., Yang, D., Montague, P. R., and Eagleman, D. M. (2007). Vividness of mental imagery: individual variability can be measured objectively. *Vis. Res.* 47, 474–478. doi: 10.1016/j.visres.2006.11.013
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Dodson, C. S., and Schacter, D. L. (2002). When false recognition meets metacognition: the distinctiveness heuristic. *J. Mem. Lang.* 46, 782–803. doi: 10.1006/jmla.2001.2822
- Foley, M. A. N. N., Hughes, K., Librot, H., and Paysnick, A. M. Y. (2009). Imagery encoding effects on memory in the DRM paradigm: a test of competing predictions. *Appl. Cogn. Psychol.* 23, 828–848. doi: 10.1002/acp.1516
- Foley, M. A., Wozniak, K. H., and Gillum, A. (2006). Imagination and false memory inductions: investigating the role of process, content and source of imaginations. *Appl. Cogn. Psychol.* 20, 1119–1141. doi: 10.1002/acp.1265
- Goff, L. M., and Roediger, H. L. (1998). Imagination inflation for action events: repeated imaginings lead to illusory recollections. *Mem. Cogn.* 26, 20–33. doi: 10.3758/BF03211367
- Graesser, A. C., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101, 371–395. doi: 10.1037/0033-295X.101.3.371
- Hege, A. C. G., and Dodson, C. S. (2004). Why distinctive information reduces false memories: evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 787–795. doi: 10.1037/0278-7393.30.4.787
- Huff, M. J., Bodner, G. E., and Gretz, M. R. (2020). Reducing false recognition in the Deese-Roediger/McDermott paradigm: related lures reveal how distinctive encoding improves encoding and monitoring processes. *Front. Psychol.* 11:602347. doi: 10.3389/fpsyg.2020.602347

- Hyman, I. E. J., and Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *J. Mem. Lang.* 35, 101–117. doi: 10.1006/jmla.1996.0006
- JASP Team (2020). JASP (Version 0.14.1) [Computer software]. Available at: <https://jasp-stats.org/faq/how-do-i-cite-jasp/>
- Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3
- Kintsch, W. (1974). *The Representation of Meaning in Memory*. Oxford, England: Lawrence Erlbaum.
- Lewandowsky, S., and Oberauer, K. (2015). Rehearsal in serial recall: an unworkable solution to the nonexistent problem of decay. *Psychol. Rev.* 122, 674–699. doi: 10.1037/a0039684
- Lindner, I., and Echterhoff, G. (2015). Imagination inflation in the mirror: can imagining others' actions induce false memories of self-performance? *Acta Psychol.* 158, 51–60. doi: 10.1016/j.actpsy.2015.03.008
- Ly, A., Verhagen, J., and Wagenmakers, E. J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: explanation, extension, and application in psychology. *J. Math. Psychol.* 72, 19–32. doi: 10.1016/j.jmp.2015.06.004
- Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *Br. J. Psychol.* 64, 17–24. doi: 10.1111/j.2044-8295.1973.tb01322.x
- Mazzoni, G., and Memon, A. (2003). Imagination can create false autobiographical memories. *Psychol. Sci.* 14, 186–188. doi: 10.1046/j.1432-1327.2000.01821.x
- McCabe, D. P., and Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory* 14, 570–583. doi: 10.1080/09658210600624564
- McCarroll, C. J. (2020). Remembering the personal past: beyond the boundaries of imagination. *Front. Psychol.* 11:585352. doi: 10.3389/fpsyg.2020.585352
- McDermott, K. B., and Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Mem. Cogn.* 34, 1273–1284. doi: 10.3758/BF03193271
- McKoon, G., and Ratcliff, R. (1992). Inference during reading. *Psychol. Rev.* 99, 440–466. doi: 10.1037/0033-295X.99.3.440
- Newstead, B. A., and Newstead, S. E. (1998). False recall and false memory: the effects of instructions on memory errors. *Appl. Cogn. Psychol.* 12, 67–79. doi: 10.1002/(SICI)1099-0720(199802)12:1<67::AID-ACP492>3.0.CO;2-1
- Oliver, M. C., Bays, R. B., and Zabrocky, K. M. (2016). False memories and the DRM paradigm: effects of imagery, list, and test type. *J. Gen. Psychol.* 143, 33–48. doi: 10.1080/00221309.2015.1110558
- Paivio, A. (1991). Dual coding theory: retrospect and current status. *Can. J. Psychol. Can. Psychol.* 45, 255–287. doi: 10.1037/h0084295
- Pearson, J., Rademaker, R. L., and Tong, F. (2011). Evaluating the mind's eye: the metacognition of visual imagery. *Psychol. Sci.* 22, 1535–1542. doi: 10.1177/0956797611417134
- Quintana, D. S., and Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry* 18:178. doi: 10.1186/s12888-018-1761-4
- Rajaram, S. (1993). Remembering and knowing: two means of access to the personal past. *Mem. Cogn.* 21, 89–102. doi: 10.3758/BF03211168
- Raposo, A., and Marques, J. F. (2013). The contribution of fronto-parietal regions to sentence comprehension: insights from the Moses illusion. *NeuroImage* 83, 431–437. doi: 10.1016/j.neuroimage.2013.06.052
- Reyna, V. F., and Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Dev. Psychol.* 30, 178–191. doi: 10.1037/0012-1649.30.2.178
- Rhodes, M. G., and Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychon. Bull. Rev.* 7, 158–162. doi: 10.3758/BF03210735
- Richardson, J. T. E. (1998). *Imagery*. Hove, East Sussex: John Wiley & Sons, Ltd.
- Robin, F. (2011). Imagination and false memories. *Imagin. Cogn. Pers.* 30, 407–424. doi: 10.2190/IC.30.4.e
- Robin, F., and Mahé, A. (2015). Effects of image and verbal generation on false memory. *Imagin. Cogn. Pers.* 35, 26–46. doi: 10.1177/0276236615574488
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Soro, J. C., Ferreira, M. B., Semin, G. R., Mata, A., and Carneiro, P. (2017). Ad hoc categories and false memories: Memory illusions for categories created on-the-spot. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1779–1792. doi: 10.1037/xlm0000401
- Thomas, A. K., and Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Mem. Cogn.* 30, 423–431. doi: 10.3758/BF03194942
- van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., et al. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis. *PsyArxiv* [Preprint]. doi: 10.31234/osf.io/yqxf
- Winograd, E., Peluso, J. P., and Glover, T. A. (1998). Individual differences in susceptibility to memory illusions. *Appl. Cogn. Psychol.* 12, S5–S27. doi: 10.1002/(SICI)1099-0720(199812)12:7<S5::AID-ACP553>3.0.CO;2-D
- World Health Organisation (2013). Declaration of Helsinki World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *J. Am. Med. Assoc.* 310, 2191–2194. doi: 10.1001/jama.2013.281053

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Maraver, Lapa, Garcia-Marques, Carneiro and Raposo. 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) and the copyright owner(s) 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.



Generalization and False Memory in an Acquired Equivalence Paradigm: The Influence of Physical Resemblance Across Related Episodes

Caitlin R. Bowman^{1,2}, Maria-Alejandra de Araujo Sanchez¹, William Hou¹, Sarina Rubin¹ and Dagmar Zeithamova^{1*}

¹ Department of Psychology, University of Oregon, Eugene, OR, United States, ² Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

OPEN ACCESS

Edited by:

Sara Cadavid,
Rosario University, Colombia

Reviewed by:

G. Elliott Wimmer,
University College London,
United Kingdom
Edgar Erdfelder,
Universität Mannheim, Germany

*Correspondence:

Dagmar Zeithamova
dasa@uoregon.edu

Specialty section:

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

Received: 18 February 2021

Accepted: 08 June 2021

Published: 20 August 2021

Citation:

Bowman CR, de Araujo Sanchez M-A, Hou W, Rubin S and Zeithamova D (2021) Generalization and False Memory in an Acquired Equivalence Paradigm: The Influence of Physical Resemblance Across Related Episodes. *Front. Psychol.* 12:669481. doi: 10.3389/fpsyg.2021.669481

The ability to make inferences about related experiences is an important function of memory that allows individuals to build generalizable knowledge. In some cases, however, making inferences may lead to false memories when individuals misremember inferred information as having been observed. One factor that is known to increase the prevalence of false memories is the physical resemblance between new and old information. The extent to which physical resemblance has parallel effects on generalization and memory for the source of inferred associations is not known. To investigate the parallels between memory generalization and false memories, we conducted three experiments using an acquired equivalence paradigm and manipulated physical resemblance between items that made up related experiences. The three experiments showed increased generalization for higher levels of resemblance. Recognition and source memory judgments revealed that high rates of generalization were not always accompanied by high rates of false memories. Thus, physical resemblance across episodes may promote generalization with or without a trade-off in terms of impeding memory specificity.

Keywords: acquired equivalence associative learning task, generalization, inference, associative memory, source memory, false memory

INTRODUCTION

Memory integration—the ability to link information across related experiences—is an important function of memory. It allows individuals to build knowledge to support inferences and generalize prior experience to novel situations. For example, after hearing that your friend from New York, Kyle, is spending his summer vacation at the Jersey Shore, you may generalize this preference to another friend from New York, Brad, assuming that he might make similar vacation plans. Memory generalization, however, may come at the expense of memory specificity, as integration may lead to false memories (Carpenter and Schacter, 2017, 2018). That is, you might falsely remember that both Kyle and Brad told you that they were going to the Jersey Shore when, in fact, you had merely inferred Brad's plans. In this study, we were interested in the relationship between false

memories and generalization, and whether contexts that promote false memories also tend to promote generalization in decision-making.

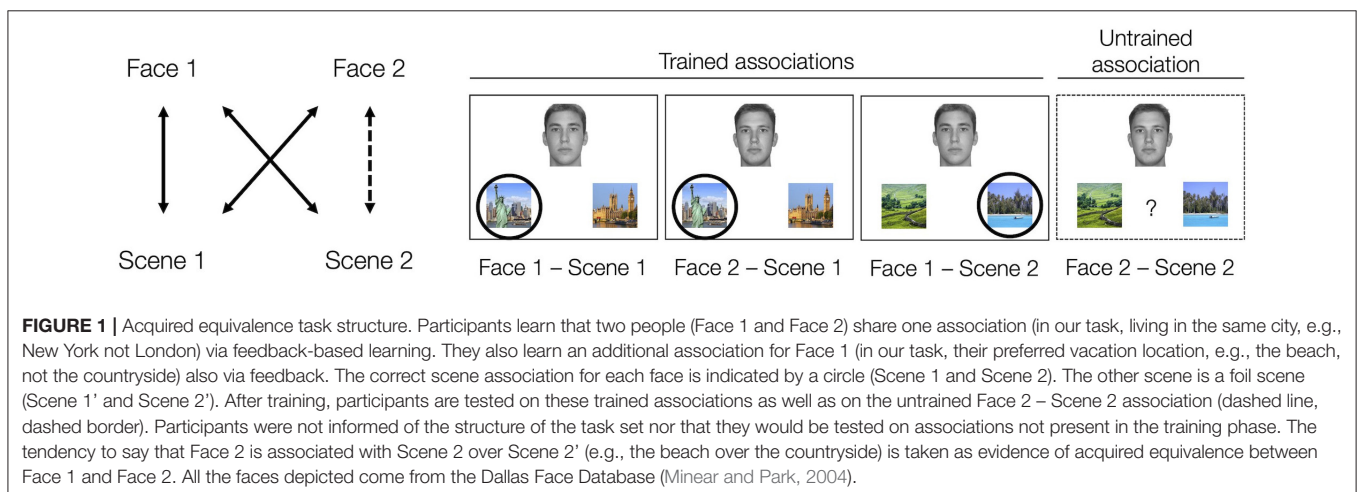
Acquired equivalence is one form of generalization, which involves assuming that when a pair of stimuli share one commonality (e.g., two people who are both from New York), they are likely to share other characteristics (e.g., preferred vacation location) (Edwards et al., 1982; Honey and Hall, 1989; Shohamy and Wagner, 2008). Importantly, acquired equivalence occurs when characteristics are extended from one individual to the other through inference rather than through direct learning. In this paradigm (see **Figure 1**), participants generally undergo training to learn a set of associations, like that Face 1 and Face 2 both prefer Scene 1 over Scene 1'. They also learn that Face 1 prefers Scene 2 over Scene 2'. The sets of associations that the participants learn directly through training are known as the trained associations. The participants are subsequently tested on the trained associations as well as on the critical untrained association between Face 2 and Scene 2. When the participants indicate that Face 2 is associated with Scene 2 over Scene 2' at rates reliably above 50%, it is taken as evidence that the participants have generalized across Face 1 and Face 2, showing "acquired equivalence" between them.

Integrative encoding is one mechanism by which acquired equivalence can occur. It involves combining memory representations of events at the time of learning based on their commonalities (Shohamy and Wagner, 2008; Zeithamova et al., 2012a; Richter et al., 2016; Schapiro et al., 2017). For example, meeting Brad and hearing he is from New York may trigger a memory for Kyle because of their shared association with New York. The memory representations of Brad and Kyle may then become integrated, because they were active at the same time, making it more likely that new information learned about either person will be applied to both of them. Prior study has supported the notion that acquired equivalence can be a function of integrative encoding, showing that acquired equivalence judgments can be made as quickly as judgments about associations learned directly (Shohamy and Wagner, 2008;

but see de Araujo Sanchez and Zeithamova, 2020). Inferences do not always require additional processing time at retrieval suggests that the inferred relationship can be established prior to retrieval. Further, hippocampal processing at the time of encoding predicted later generalization, suggesting that links across episodes were formed during learning rather than on-the-fly during generalization itself (Shohamy and Wagner, 2008).

A key component of integrative encoding is pattern completion. Pattern completion occurs when a partial cue triggers retrieval of a complete engram, leading to recognition (Marr, 1971; McClelland et al., 1995; Hunsaker and Kesner, 2013; Horner et al., 2015). In the above example, pattern completion occurs when Brad mentions being from New York, which then triggers a memory for Kyle who is also associated with New York. This process can lead to integrative encoding when reactivated prior events are then re-encoded along with current experience (Zeithamova et al., 2012a; Schapiro et al., 2017). Further, a prior study suggests that the likelihood of pattern completion occurring increases with the degree of resemblance between past experience and the current retrieval cue (Guzowski et al., 2004; Lee et al., 2004; Vazdarjanova and Guzowski, 2004; Liu et al., 2016). Recent evidence has also shown that higher similarity between related episodes increases the likelihood that individuals will make new inferences that combine information across those episodes (Molitor et al., 2021). The first goal of this study was to extend this finding to an acquired equivalence paradigm, testing whether increasing physical resemblance between stimuli comprising overlapping associations would lead to increases in rates of acquired equivalence.

While integrative encoding may support inference and facilitate later generalization, it may also come with a trade-off: reduced memory specificity. Combining representations of related events may lead unique aspects of those experiences to be lost or conflated. Increases in false memories, in particular, have sometimes been posited as a consequence of generalization (Carpenter and Schacter, 2017, 2018; Varga et al., 2019). For example, Carpenter and Schacter (2017, 2018) showed that



when participants made inferences across related experiences, they tended to misattribute unique contextual details from one episode to the other, conflating the two experiences. Likewise, Shohamy and Wagner (2008) reported anecdotal evidence that individuals who generalized well in an acquired equivalence paradigm were those who conflated directly learned and inferred relationships during learning, although this finding has been difficult to replicate (de Araujo Sanchez and Zeithamova, 2020). Outside inference paradigms, prior research shows that rates of false memories tend to increase with increasing perceptual and/or conceptual overlap between new and old information (Roediger et al., 1995; Arndt and Reder, 2003; Gutchess and Schacter, 2012; Bowman and Dennis, 2015). This effect is thought to occur, in part, because pattern completion can be triggered by the partial overlap between new and old information (Toner et al., 2009; Yassa et al., 2011; Vieweg et al., 2015). Thus, while manipulating the physical resemblance across stimuli within an acquired equivalence paradigm may lead to more generalization by promoting integrative encoding, it may also lead to higher rates of false memories.

Alternatively, generalization may not lead to false memories if inferences are retrieval-based rather than formed through integrative encoding (Zeithamova and Bowman, 2020). Retrieval-based inference occurs when related episodes are encoded separately and then retrieved in parallel to make generalization decisions on the fly (Zeithamova and Preston, 2010; Banino et al., 2016). Importantly, when generalization occurs *via* this flexible retrieval mechanism, it is possible to maintain separate representations of related episodes while still generalizing. In this case, memory specificity and generalization may go hand-in-hand, because the quality of the separate representations determines the accuracy of later decision-making (Kumaran, 2012; Kumaran and McClelland, 2012). Recent study has formally tested both generalization and memory specificity in inference paradigms, sometimes showing no trade-off (de Araujo Sanchez and Zeithamova, 2020), or even a positive relationship between them (Banino et al., 2016). Thus, false memories are not always a consequence of generalizing, and the degree of overlap across experiences may have different effects on false memory and generalization when related experiences are coded separately (Zeithamova and Bowman, 2020).

To investigate the parallels between memory generalization and false memories, we conducted three experiments using an acquired equivalence paradigm while manipulating the physical resemblance between items that comprised related experiences using computer-blended face stimuli. In Experiment 1, some pairs of faces shared a common parent image, while other pairs were blended without a shared parent. In Experiment 2, all pairs of faces shared a common parent, but the weight given to the shared parent varied parametrically. We then tested both generalization and rates of false memory for the source of the inferred information, measuring whether physical resemblance had similar or distinct effects across these two memory judgments. In Experiment 3, we added a recognition test to determine whether pair-mate faces could be discriminated from one another. We hypothesized that physical resemblance

would increase the likelihood that prior related experiences would be reactivated during new learning, promoting integrative encoding and supporting generalization. However, we also hypothesized that reliance on integrative encoding would lead to conflating of separate experiences, with higher rates of false memories for the source of information when items resembled one another.

EXPERIMENT 1

Materials and Methods

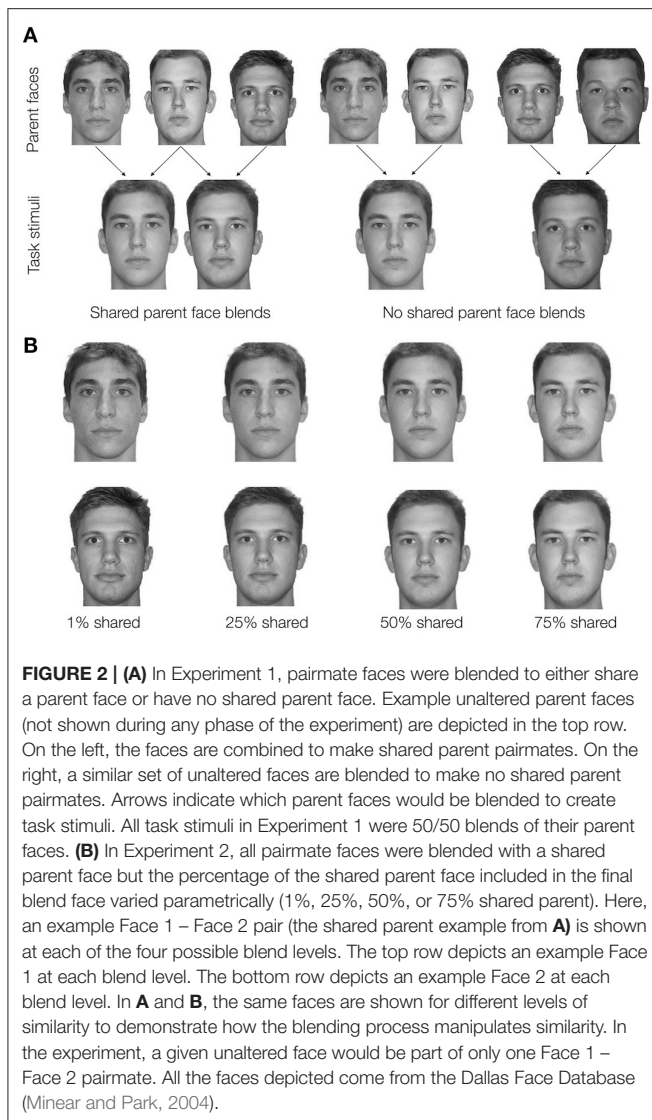
Participants

Forty participants from the University of Oregon completed the experiment for course credit (27 females, mean age = 19.5 years, SD age = 1.3 years, age range = 18–23 years). This sample size was within the range of prior acquired equivalence studies collected in the laboratory (de Araujo Sanchez and Zeithamova, 2020). While there was no prior study on which to base the expected effect size of physical similarity, we chose a moderate effects size ($d = 0.5$) *a priori* as the effect size for which we aimed to have sufficient power. All the participants completed written informed consent, and the Institutional Review Board of the University of Oregon approved all procedures.

Materials

The stimuli were eight colored images of scenes and eight gray scale images of Caucasian male faces. Faces and scenes were paired to form four quadruplets. Each quadruplet included two faces (Face 1 and Face 2, called pair-mates) each paired with two scenes (Scene 1 and Scene 2). There were four possible face-scene pairs within each quadruplet: Face 1-Scene 1, Face 1-Scene 2, Face 2-Scene 1, and Face 2-Scene 2. Three of these face-scene pairs were used for training (Face 1-Scene 1, Face 1-Scene 2, and Face 2-Scene 1). The last one (Face 2-Scene 2) was untrained and only used during the subsequent test to measure generalization *via* acquired equivalence (Figure 1). Of the eight colored scenes, four were well-known cities and four were nature scenes. For each quadruplet, Scene 1 was a city scene and Scene 2 was a nature scene. The specific pairing of scenes to quadruplets was randomly assigned for each participant.

Faces within each quadruplet were constructed by blending two unaltered face images together using FantaMorph Version 5 by Abrosoft. For each participant, the pair-mate faces for two of the four quadruplets were blended with a shared parent that made up 50% of the blend (Figure 2A). For example, unaltered Face A would be 50/50 blended with unaltered Face B to create Face 1, and unaltered Face C would also be 50/50 blended with unaltered Face B to create Face 2. Thus, these pair-mate faces were manipulated to share a physical similarity. The faces in the other two quadruplets were also created as 50/50 blends but did not share a parent face. For example, unaltered Face D would be 50/50 blended with unaltered Face E to create Face 1, and unaltered Face F would be 50/50 blended with unaltered Face G to create the Face 2. These pair-mates will be referred to as no shared parent pair-mates. For each participant, 14 unaltered faces were randomly selected from a set of 18 possible faces and randomly assigned to create shared parent pair-mates (six parent faces,



three for each of two shared parent pair-mates) or no shared parent pair-mates (eight parent faces, four for each of two no shared parent pair-mates). The resulting blended faces were then randomly paired with their nature/city scenes.

Procedure

Initial Exposure

To familiarize the participants with the faces prior to asking them to form associations between faces and scenes, the participants first passively viewed each face in isolation three times across a single block. Each face was presented for 2 s followed by a 1-s fixation cross. The participants made no responses and were simply instructed to remember the faces without any indication of the upcoming task structure.

Training

During training, the participants learned three face-scene associations for each quadruplet: Face 1-Scene 1, Face 2-Scene

1, and Face 1-Scene 2. They were instructed to try to learn where each person vacationed (nature scene) and where they lived (city scene). On each trial, a face cue was presented at the top of the screen with two scene options below. For nature scene trials, the cue “Where does he vacation?” appeared at the top of the screen above the face. “Where does he live?” appeared on the screen for city scene trials. The face and scene options were presented for 3 s during which time the participants selected which of the two presented scenes they thought was associated with the cue face by pressing a button on the keyboard. The participants were then given feedback on a separate screen: “Correct” if they selected the correct scene, “Incorrect” if they selected the incorrect scene, and “Too late” if they did not respond within 3 s. Feedback appeared on the screen for 1 s, followed by a 1-s fixation cross.

The scene options were generated so that (1) options were either both city scenes or both nature scenes and (2) scenes were paired together so that a given scene was always presented as an option along with the same other scene, once serving as the target (correct) scene and once as the foil (incorrect) scene. For example, for quadruplet A, the nature scene options might be a beach and a field with the beach being the correct answer. The city scene options might be New York and London with New York being the correct answer. Quadruplet B would then be yoked to Quadruplet A such that it has the same scene options for both nature and city scenes, but the opposite correct answers (i.e., the field and London). This ensures that the participants must learn the association between the face and the scene and not merely which scene is correct more often or correct when compared with a particular other scene, which could occur if the foil scene was randomized. The left/right presentation of the scene options was counterbalanced within the participants across trials.

The participants underwent 16 blocks of training containing 12 trials each, totaling 192 training trials. They took a self-paced break between each block. The order of trials was randomized for each participant with no cue face-scene option combination shown more than three times consecutively.

Acquired Equivalence Test

Immediately following training, the participants were tested on the trained pairs as well as the untrained Face 2-Scene 2 pairing to test for acquired equivalence between Face 1 and Face 2. As in the training phase, a face cue was presented with two scene options on each trial with a corresponding question (“Where does he live?” or “Where does he go on vacation?”) presented above the cue face. The cue face and scene options were presented for 3 s during which time the participants were to indicate which scene was associated with the cue face. Unlike the training phase, the participants did not receive feedback during the test. Each trial ended with a fixation cross that was presented for 1 s. The structure of the scene options remained the same as in the training phase.

To improve the reliability of estimates of acquired equivalence given the limited trials (one for each of four quadruplets), each association was tested six times throughout the test (Shohamy and Wagner, 2008). The order of presentation was randomized with the constraint that the same association was not tested more than twice in a row. The acquired equivalence test was completed

in a single block with a total of 96 trials (four associations for each of four quadruplets each presented six times).

Source Memory Test

Following the acquired equivalence test, the participants underwent a source memory test to measure how well they were able to remember when they encountered each association. On each trial, the participants were presented with one face and two scenes arranged vertically above each other to remove the left-right organization of scenes present in training and test. The task was to indicate whether and when the three pictures were presented together previously, regardless of their spatial arrangement.

There were four possible response options, presented from left to right: “study,” “test,” “both,” and “never.” “Study” responses indicated that the images were seen together only during the training phase—an answer that was never correct but was included so that the response options would not give away the task structure. “Test” responses indicated that the images were seen together only during the acquired equivalence test the participants had just completed—an answer that was correct only for Face 2-Scene 2 acquired equivalence items. “Both” responses indicated that the images had been seen together during both the training phase and the acquired equivalence test—an answer that was correct for the three types of trained pairs (i.e., Face 1-Scene 1, Face 2-Scene 1, and Face 1-Scene 2). Lastly, “never” responses indicated that the three images had not been shown together in previous phases of the experiment. We constructed three types of recombined trials for which “never” was the correct answer. “Recombined all” trials consisted of a face and two scenes in which none of the components had been presented with one another. “Recombined face” trials consisted of two scenes that had been previously presented together (e.g., Scene 1 and Scene 1') and a face that was never presented with those scenes. “Recombined scene” trials consisted of a cue face presented with its corresponding Scene 1 and Scene 2. Thus, each scene was presented with the face previously, but the scenes were never presented together. While this trial type has been used in a prior study with random scene types (de Araujo Sanchez and Zeithamova, 2020), it was particularly obvious in this experiment that these images had not been presented together because of the nature vs. city scene manipulation, as scene options were always either two cities or two nature scenes. One face from each of the four quadruplets was used to create all seven types of source memory trials: three trained (“both”), 1 untrained (“test”), and three recombined (“never”) for a total of 28 source memory test trials. The source memory test was self-paced, with a 1-s fixation cross between each trial.

Of primary interest in the source memory test was how often the participants reported having seen the untrained association during both training and test. We refer to this as a source false memory, but it could also reflect a participant generating the F2-S2 association during encoding through reactivation, then later misattributing that internal experience to the training task. Other trial types were included to help differentiate integration-related source confusion (i.e., false memory for untrained associations

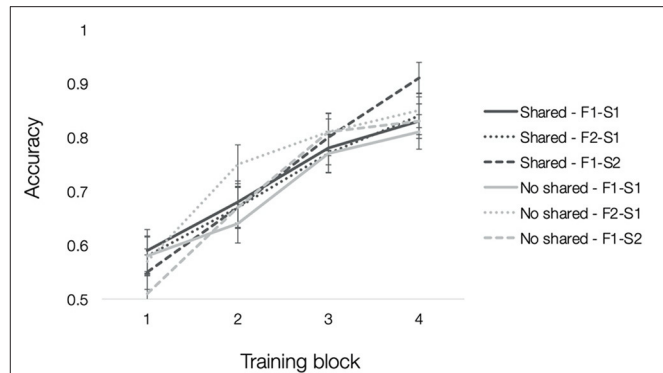


FIGURE 3 | Experiment 1 training accuracy. Mean accuracy from each block of training separated by trained association type (F1-S1 solid lines, F2-S1 dotted lines, F1-S2 dashed lines) and pair-mate type (shared parent in dark gray, no shared parent in light gray). Error bars depict the standard error of the mean across subjects.

being presented during the study) from overall poor source memory or response biases.

Design and Statistical Analyses

The primary independent variable of interest was a physical resemblance between pair-mate faces. There were two levels of physical resemblance (sharing a parent, not sharing a parent), and this was manipulated within subjects. We had two primary dependent variables of interest: acquired equivalence for untrained pairs measured during the acquired equivalence test and false memories for the source of untrained associations. For all analyses, a Greenhouse–Geisser correction was applied (denoted with “GG”) when the sphericity assumption was violated. A Bonferroni correction for multiple comparisons was applied when multiple independent statistical tests were computed, such as following up on a significant omnibus ANOVA with multiple pairwise comparisons.

Results

Training

Mean accuracies for each type of trained association are depicted in **Figure 3**. To test how resemblance between faces within a quadruplet affected acquisition of trained associations, we computed a 2 (pair-mate type: shared parent, no shared parent) $\times$ 3 (trained association type: F1-S1, F2-S1, and F1-S2) $\times$ 4 (training block: 1–4) repeated-measures ANOVA. There was a significant main effect of training block [$F_{(3, 117)} = 76.89, p < 0.001, \eta_p^2 = 0.66$] with an accompanying linear effect of training block [$F_{(1, 39)} = 164.22, p < 0.001, \eta_p^2 = 0.81$], which showed increasing accuracy with training. No other effects reached significance (all F 's $< 2, p$'s > 0.07).

Acquired Equivalence Test

Figure 4A depicts results from the acquired equivalence test in terms of accuracy for trained associations (F1-S1, F2-S1, and F1-S2) and the proportion of trials showing acquired equivalence for untrained associations (F2-S2). For the ease of report of the ANOVA results, we will refer to the proportion of untrained

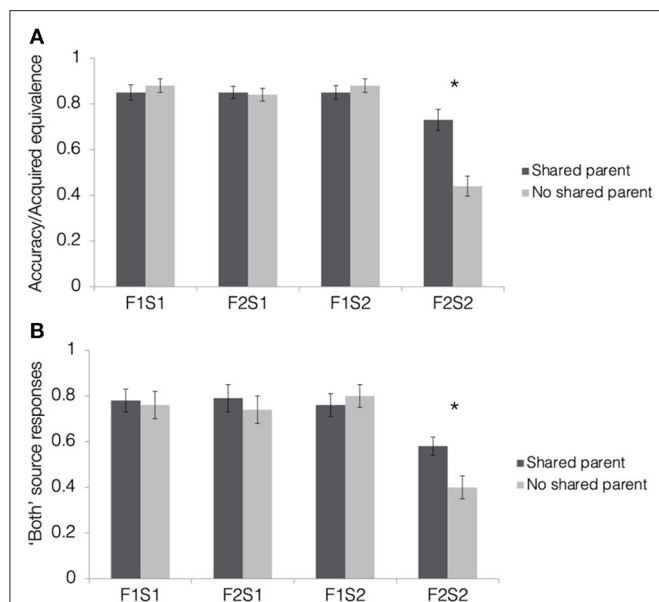


FIGURE 4 | Results from the tests of acquired equivalence and source memory in Experiment 1. **(A)** Acquired equivalence test accuracy for trained associations (F1-S1, F2-S1, and F1-S2) and rates of acquired equivalence for untrained associations (F2-S2). **(B)** For the source memory test, the proportion of trials where the participants responded that they had seen the items together during both study and test. This was the correct response for the trained associations but constituted a false memory for the untrained associations, since they were presented only during the test. All results are depicted separately for quadruplets with faces sharing a parent (dark gray bars) and quadruplets with faces not sharing a parent (light gray bars). Stars indicate a significant paired difference, following a significant test item type $\times$ quadruplet type interaction effect (corrected alpha = 0.0125).

associations on which a participant responded consistently with acquired equivalence as “accuracy,” but generalization of scene associations from one face to another is not inherently correct or incorrect in this paradigm. First, we tested whether rates of acquired equivalence differed significantly from chance (50% for two alternative forced choices) using one-sample t -tests for shared and no shared parent pair-mates separately. Rates of acquired equivalence were significantly higher than chance for pair-mates sharing a parent [$M = 0.73$, $SD = 0.29$; $t_{(39)} = 4.93$, $p < 0.001$, $d = 0.76$; Bonferroni-corrected alpha-level: $p < 0.025$] but not for pair-mates without a shared parent [$M = 0.44$, $SD = 0.28$; $t_{(39)} = -1.46$, $p = 0.15$, $d = -0.23$].

To test how resemblance between pair-mate faces affected the tendency to show acquired equivalence, we computed a 2 (pair-mate type: shared parent, no shared parent) $\times$ 4 (test item type: F1-S1, F2-S1, F1-S2, and F2-S2) repeated-measures ANOVA. There was a significant main effect of test item type [$F_{(1.94, 75.67)} = 55.52$, $p < 0.001$, $\eta_p^2 = 0.59$] with all trained associations having higher accuracy (F1-S1: $M = 0.86$, $SD = 0.17$; F2-S1: $M = 0.84$, $SD = 0.13$; F1-S2: $M = 0.86$, $SD = 0.16$) than untrained associations (F2-S2: $M = 0.58$, $SD = 0.19$; t 's > 8 , p 's < 0.001 ; Bonferroni-corrected alpha level: $p < 0.0083$). Accuracy for the trained associations did not differ from one another (t 's < 1 ,

TABLE 1 | Experiment 1 source memory responses separated by pair-mate and trial type.

	Response type			
	Study only	Test only	Both study and test	Never
Face 1—Scene 1				
Shared parent	0.06 (0.20)	0.11 (0.27)	0.78 (0.34)	0.05 (0.15)
No shared parent	0.10 (0.28)	0.11 (0.27)	0.76 (0.39)	0.03 (0.11)
Face 2—Scene 1				
Shared parent	0.06 (0.23)	0.09 (0.25)	0.79 (0.36)	0.06 (0.20)
No shared parent	0.08 (0.21)	0.10 (0.23)	0.74 (0.39)	0.09 (0.22)
Face 1—Scene 2				
Shared parent	0.04 (0.13)	0.14 (0.25)	0.76 (0.34)	0.06 (0.17)
No shared parent	0.05 (0.15)	0.11 (0.24)	0.80 (0.32)	0.04 (0.18)
Face 2—Scene 2				
Shared parent	0.06 (0.17)	0.24 (0.34)	0.58 (0.37)	0.13 (0.25)
No shared parent	0.06 (0.17)	0.38 (0.35)	0.40 (0.41)	0.16 (0.31)
Recombined face	0.04 (0.11)	0.14 (0.20)	0.25 (0.28)	0.58 (0.32)
Recombined scenes	0.04 (0.13)	0.06 (0.19)	0.16 (0.23)	0.73 (0.36)
Recombined all	0.01 (0.06)	0.07 (0.13)	0.06 (0.14)	0.86 (0.23)

Mean (SD). Bolded rate = correct response for a given trial type.

p 's > 0.3). There was also a significant main effect of pair-mate type [$F_{(1, 39)} = 5.32$, $p = 0.03$, $\eta_p^2 = 0.12$] with higher overall accuracy for shared parent pair-mates ($M = 0.82$, $SD = 0.16$) compared with those with no shared parent ($M = 0.76$, $SD = 0.15$). These main effects were qualified by a significant pair-mate type $\times$ test item type interaction [$F_{(1.73, 67.57)} = 14.13$, $p < 0.001$, $\eta_p^2 = 0.27$]. There was no difference between shared parent pair-mates and no shared parent pair-mates for any of the trained associations (t 's < 1.1 , $p > 0.3$), but participants showed higher rates of acquired equivalence when pair-mates shared a parent compared with when they did not [$t_{(39)} = 4.38$, $p < 0.001$, $d = 0.69$; Bonferroni-corrected alpha level: $p < 0.0125$]. Thus, the participants learned all trained associations equally well, but were more likely to generalize across pairs of faces when they were more physically similar to one another.

Source Memory Test

Mean response rates for all source memory response options are presented in Table 1. Mean proportion of “both study and test” responses for each pair-mate and association type is presented in Figure 4B. To test whether the participants tended to falsely remember untrained pairs as having been presented during both the study and test phase, we computed a 2 (pair-mate type: shared parent, no shared parent) $\times$ 4 (test item type: F1-S1, F2-S1, F1-S2, F2-S2) repeated-measures ANOVA on the proportion of “both study and test” responses. There was a marginal main effect of pair-mate type [$F_{(1, 39)} = 3.93$, $p = 0.055$, $\eta_p^2 = 0.09$] with numerically higher rates of “both study and test” responses for pair-mates sharing a parent ($M = 0.73$, $SD = 0.27$) compared with those not sharing a parent ($M = 0.68$, $SD = 0.28$). There was a significant main effect of test item type [$F_{(1.93, 75.27)} = 15.25$, $p < 0.001$, $\eta_p^2 = 0.27$, GG] with lower rates of “both study and

test” responses for untrained associations (F2-S2: $M = 0.49$, $SD = 0.33$) than for all of the trained association types (F1-S1: $M = 0.77$, $SD = 0.34$; F2-S1: $M = 0.76$, $SD = 0.35$; F1-S2: $M = 0.78$, $SD = 0.3$; all t 's > 4 , p 's < 0.001). There was no difference between shared parent pair-mates and no shared parent pair-mates for any of the trained associations (t 's < 1.1 , $p > 0.3$), but the participants showed higher rates of source memory errors when pair-mates shared a parent compared with when they did not [$t_{(39)} = 4.38$, $p < 0.001$, $d = 0.69$; Bonferroni-corrected alpha level: $p < 0.0125$]. There were no overall differences in source memory across the three types of trained associations (all t 's < 0.5 , p 's > 0.6). Importantly, there was a significant pair-mate type $\times$ test item type interaction effect [$F_{(3, 117)} = 3.31$, $p = 0.02$, $\eta_p^2 = 0.08$]. Follow-up paired t -tests revealed no differences between shared and no shared parent pair-mates for the trained associations (all t 's < 1.2 , p 's > 0.25) but increased false memory for having seen the untrained test items (F2-S2) during both study and test when pair-mates shared a parent compared with when they did not [$t_{(39)} = 2.66$, $p = 0.01$, $d = 0.42$; Bonferroni-corrected alpha level: $p < 0.0125$]. Thus, the physical resemblance between pair-mate faces led to increases in false memories of the source (observed instead of inferred) in addition to the increases in generalization.

Lastly, we tested whether the participants falsely remembered all new pairings on the source memory test as having been presented during both the study and test phase or if this effect was particular to the untrained inference pairs. To do so, we compared the rate of “both” responses for untrained associations to each type of recombined trial, separately for the shared parent pair-mates and no shared parent pair-mates. In all cases, the participants responded “both” to the untrained pairs at numerically higher rates than any type of recombined trial. All comparisons of untrained vs. recombined trials reached Bonferroni corrected threshold of $p < 0.0083$ (t 's > 4.7 , p 's < 0.001), with the exception of untrained no shared parent condition compared with the recombined face condition [$t_{(39)} = 2.24$, $p = 0.03$, $d = 0.35$] and untrained no shared parent condition compared with the recombined scene condition [$t_{(39)} = 2.75$, $p = 0.009$, $d = 0.43$] that only reached an uncorrected threshold. Thus, the evidence generally points to false source memories being specific to the inference pairings.

Comparing Physical Resemblance Effects in Generalization and False Source Memories

The results showed that greater physical resemblance led to more acquired equivalence as well as more false memories for the source of learning for these associations. We then tested whether the sizes of the effect in generalization and source memory were similar or if one was larger than the other. If acquired equivalence was based mostly or entirely on integration at encoding, then we would expect the two effects to be of a similar size. However, if some acquired equivalence judgments were based on flexible retrieval, memory for the source of these associations might be maintained even in the face of successful generalization, making the effect of physical resemblance smaller in source memory judgments. To test this idea, we computed a 2 (test: acquired equivalence, source memory) $\times$ 2 (pair-mate

type: shared parent, no shared parent) ANOVA on rates of generalization and false memory for the inference pairs (F2-S2) and were specifically interested in the interaction effect. The test $\times$ pair-mate type interaction effect was not significant [$F_{(1, 39)} = 1.28$, $p = 0.27$, $\eta_p^2 = 0.03$]. Thus, we did not see strong evidence for a differential effect of physical resemblance on generalization and false memory.

Discussion

In Experiment 1, we tested the hypothesis that greater physical resemblance between pair-mates in an acquired equivalence paradigm would lead to increased generalization across related faces. We also hypothesized that increases in generalization would be accompanied by increases in false memory for the source of generalized associations. The results confirmed both hypotheses. Rates of acquired equivalence were higher when pair-mate faces were blended with a shared parent compared with when pair-mate faces were blended without a shared parent. This difference in generalization for the shared vs. no-shared parent conditions arose despite comparably high performance for the associations presented during training. False memory rates were also higher for the shared parent pair-mates than the no shared parent pair-mates: the participants tended to erroneously believe that the untrained associations had been presented during both the training and test phases rather than in the test phase alone. This difference in source memory for the untrained associations emerged, while source memory for the trained associations did not differ for the shared and no shared parent pair-mates. Together, these findings provide initial evidence that physical resemblance may serve as a cue to reactivate prior experience while encoding related information, leading to integration of the related experiences to support generalization but potentially losing some contextual details.

EXPERIMENT 2

In Experiment 1, we manipulated physical resemblance in a binary manner: pair-mate faces either shared a parent or did not share a parent. Prior study has shown that the degree of similarity, not just its presence or absence, can be an important factor affecting memory fidelity (Turney and Dennis, 2017; Bowman et al., 2019). Graded similarity between items also affects the likelihood of inference across related events, although in a more all-or-none manner (Molitor et al., 2021). However, as these prior studies have not measured memory fidelity and inference across the same events, it is unclear whether graded vs. all-or-none effects of similarity were driven by differences in task parameters across studies or a true divergence across these types of memory judgments. In Experiment 2, we tested how rates of acquired equivalence and false memory for the source of inferred associations varied across levels of pair-mate resemblance that varied parametrically. We constructed pair-mate faces that all shared a parent face but differed in the degree to which the shared parent influenced the final blend (**Figure 2B**). As in Experiment 1, we tested how likely the participants were to treat the pair-mate faces as equivalent and how often the participants falsely remembered encountering inferred associations during training.

Materials and Method

Participants

Thirty-eight participants from the University of Oregon completed the experiment for course credit. One participant was excluded because of incomplete data, leaving 37 participants reported in all analyses (26 females, mean age = 19.24 years, SD age = 2.41 years, age range = 18–28 years). This sample size was within the range of prior acquired equivalence studies collected in the laboratory (de Araujo Sanchez and Zeithamova, 2020) and was selected, because it would be well-powered to detect a moderate effect size ($d = 0.5$). We did not have prior data to estimate an effect size across a physical similarity gradient, but data from Experiment 1 suggested that comparisons between individual blend levels (i.e., 50 and 1%) would likely be within this range. All the participants completed written informed consent, and the Institutional Review Board of the University of Oregon approved all procedures.

Materials

Blended faces were combined with nature scenes and city scenes to create four quadruplets in a manner similar to Experiment 1. However, rather than having pair-mate faces either share a parent face or not, all pair-mate faces shared a parent face. The pair-mates differed in the level at which the shared parent face was blended: pair-mates with the least overlap between Face 1 and Face 2 had a shared parent blended at 1%, followed by pair-mates with a shared parent blended at 25%, followed by pair-mates with a shared parent blended at 50%, and lastly the pair-mates with the most overlap between Face 1 and Face 2 had a shared parent blended at 75% (**Figure 2B**). Each blend level was applied to one quadruplet in the set. Besides the difference in the blend procedure, stimulus generation was the same as in Experiment 1 with the exception that there were 20 possible unaltered faces to choose from based on the creation of additional face blends in the time between experiments.

Procedure

The overall procedure followed that of Experiment 1 with the addition of a recognition test following the source memory test. Deviations from the Experiment 1 procedure are noted below.

Initial Exposure

Rather than being shown in a fixed order, stimulus order was randomized with the constraint that no face was shown twice in a row.

Training

The participants underwent six blocks of training, each containing 24 trials for a total of 144 training trials. In each block, the participants saw each of the three trained associations from each of the four quadruplets twice in a random order. The question cues were indicated by “Vacation?” or “Live?” displayed at the top of the screen, and the feedback for trials in which the participants did not respond within the 3 s allotted was changed to “Sorry, too slow!”

Acquired Equivalence Test

The acquired equivalence test was as in Experiment 1, except that the question cues at the top of the screen were “Vacation?” and “Live?” as in the training phase of this experiment.

Source Memory Test

The source memory test for the trained and untrained associations was as in Experiment 1 with the exception that these associations were tested twice to obtain stable source memory estimates. Differing from Experiment 1, we revised the “recombined scene” condition to match the scene options to be both city or both nature scenes. For recombined conditions, a coding error resulted in some trials being the same as the F2-S2 trials. These trials were excluded from all analyses. Altogether, there were 40 trials in the source memory test.

Face Recognition Test

In addition to the acquired equivalence and source memory tests that were of primary interest, we also included a face recognition test. However, it was aimed at testing the discriminability of a face at its different blend levels (e.g., 50 vs. 75%), which was not relevant for the current goals. We, thus, do not discuss this recognition test further.

Design and Statistical Analysis

The design and analytical approach were as in Experiment 1 except that there were four levels of physical resemblance rather than two.

Results

Training

Mean accuracies for each type of trained association are presented in **Table 2**. To test how resemblance between pair-mate faces affected learning, we computed a 4 (shared parent blend level: 1, 25, 50, and 75%) $\times$ 3 (trained association type: F1-S1, F2-S1, and F1-S2) $\times$ 6 (training block: 1–6) repeated-measures ANOVA. There was a significant main effect of training block [$F_{(2.31, 82.97)} = 39.75, p < 0.001, \eta_p^2 = 0.53, GG$] accompanied by a significant linear effect [$F_{(1, 36)} = 70.88, p < 0.001, \eta_p^2 = 0.66$] showing increasing accuracy across training. There was also a significant main effect of trained association type [$F_{(2, 72)} = 4.38, p = 0.02, \eta_p^2 = 0.11$]. Overall learning was significantly better for F2-S1 pairs ($M = 0.75, SD = 0.16$) compared with F1-S2 pairs [$M = 0.69, SD = 0.16; t_{(36)} = 2.68, p = 0.011, d = 0.44$; Bonferroni-corrected alpha level: $p < 0.016$]. No other difference passed the corrected alpha level (t 's $< 2.2, p$'s > 0.04). There was also a significant main effect of pair-mate blend level [$F_{(3, 108)} = 3.27, p = 0.02, \eta_p^2 = 0.08$]. Overall, accuracy was numerically poorest for 25% shared parent pair-mates ($M = 0.7, SD = 0.17$), followed by those with 50% shared parent ($M = 0.71, SD = 0.17$), then those with 1% shared parent ($M = 0.74, SD = 0.17$), and those with 75% shared parent had the highest overall accuracy ($M = 0.77, SD = 0.17$). However, only the difference between 75 and 25% blends reached significance [$t_{(36)} = 3.09, p = 0.004, d = 0.51$; all other t 's $< 2.4, p$'s > 0.01 ; Bonferroni-corrected alpha level: $p < 0.0083$]. No interaction effect reached significance (all F 's $< 1.8, p$'s > 0.1).

TABLE 2 | Experiment 2 training accuracy separated by block, shared parent blend level, and trained association type.

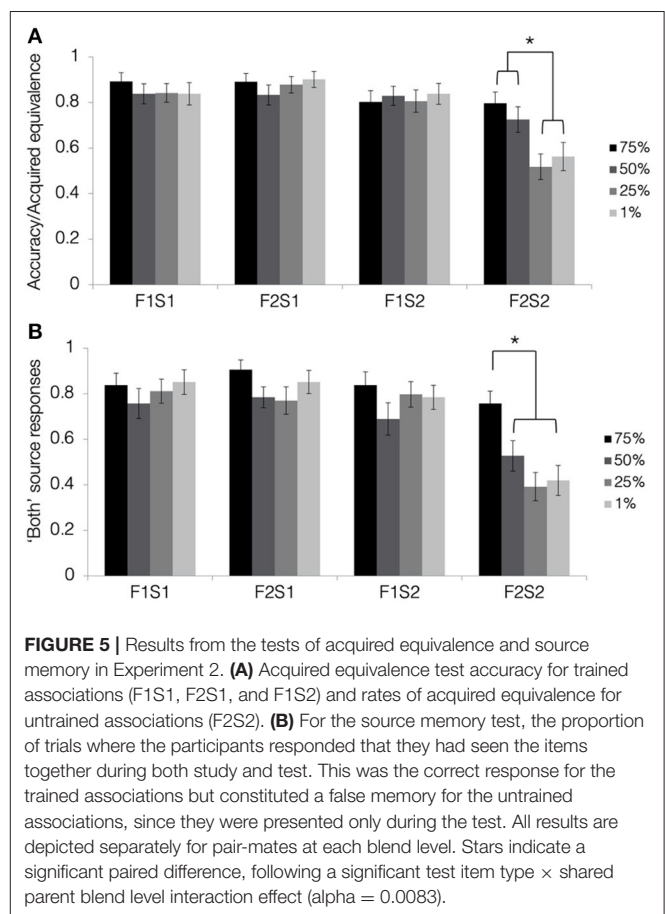
Experiment 2	Training block					
	1	2	3	4	5	6
Face 1—Scene 1						
1%	0.54 (0.34)	0.62 (0.40)	0.78 (0.32)	0.78 (0.34)	0.85 (0.29)	0.88 (0.27)
25%	0.53 (0.37)	0.57 (0.39)	0.66 (0.37)	0.69 (0.43)	0.74 (0.35)	0.84 (0.29)
50%	0.55 (0.37)	0.73 (0.37)	0.73 (0.37)	0.76 (0.35)	0.76 (0.33)	0.88 (0.25)
75%	0.76 (0.35)	0.76 (0.37)	0.82 (0.32)	0.85 (0.26)	0.87 (0.30)	0.82 (0.34)
Face 2—Scene 1						
1%	0.51 (0.38)	0.69 (0.34)	0.72 (0.32)	0.76 (0.35)	0.88 (0.30)	0.87 (0.30)
25%	0.57 (0.36)	0.53 (0.41)	0.72 (0.38)	0.82 (0.29)	0.92 (0.25)	0.85 (0.26)
50%	0.54 (0.40)	0.61 (0.36)	0.76 (0.33)	0.84 (0.26)	0.81 (0.32)	0.84 (0.29)
75%	0.57 (0.36)	0.72 (0.36)	0.82 (0.34)	0.88 (0.27)	0.84 (0.31)	0.91 (0.20)
Face 1—Scene 2						
1%	0.57 (0.34)	0.72 (0.38)	0.80 (0.36)	0.72 (0.38)	0.77 (0.32)	0.81 (0.32)
25%	0.49 (0.34)	0.57 (0.38)	0.70 (0.34)	0.72 (0.36)	0.80 (0.34)	0.87 (0.28)
50%	0.46 (0.40)	0.60 (0.42)	0.68 (0.34)	0.72 (0.34)	0.77 (0.38)	0.70 (0.38)
75%	0.51 (0.38)	0.60 (0.41)	0.74 (0.33)	0.66 (0.37)	0.78 (0.34)	0.87 (0.28)

Mean (SD).

Acquired Equivalence

Figure 5A depicts results from the acquired equivalence test in terms of accuracy for trained associations (F1-S1, F2-S1, and F1-S2) and the proportion of trials showing acquired equivalence for untrained associations (F2-S2). First, we tested whether rates of acquired equivalence differed significantly from chance using a one-sample *t*-test at each blend level. Results revealed that rates of acquired equivalence differed from chance for pair-mates blended at 75 and 50% (both *t*'s > 4, *p*'s < 0.001; Bonferroni-corrected alpha-level: *p* < 0.0125) but not those blended at 25 or 1% (both *t*'s < 1.1, *p*'s > 0.31).

To test how resemblance between pair-mate faces affected the tendency to show acquired equivalence, we computed a 4 (shared parent blend level: 1, 25, 50, and 75%) × 4 (test item type: F1-S1, F2-S1, F1-S2, and F2-S2) repeated-measures ANOVA. There was a significant main effect of test item type [$F_{(2.14, 77.08)} = 33.9$, *p* < 0.001, $\eta_p^2 = 0.49$, GG], with F2-S1 pairs having the highest numeric accuracy (*M* = 0.88, *SD* = 0.19), followed by F1-S1 pairs (*M* = 0.85, *SD* = 0.21), and F1-S2 pairs (*M* = 0.82, *SD* = 0.23), and with F2-S2 pairs showing the lowest accuracy (*M* = 0.65, *SD* = 0.21). Accuracy for all types of trained associations was significantly higher than for untrained associations (all *t*'s > 5, *p*'s < 0.001; Bonferroni-corrected alpha level: *p* < 0.0083). Differences between the trained associations did not pass correction for multiple comparisons (*t*'s < 2.8, *p*'s > 0.009). There was also a significant main effect of shared parent blend level [$F_{(3, 108)} = 3.58$, *p* = 0.02, $\eta_p^2 = 0.09$]. Overall accuracy was highest for pair-mates with a shared parent blended at 75% (*M* = 0.85, *SD* = 0.24), followed by those blended at 50% (*M* = 0.81, *SD* = 0.21), those blended at 1% (*M* = 0.79, *SD* = 0.2), and lastly those blended at 25% (*M* = 0.76, *SD* = 0.21). Those blended at 75% differed significantly from those blended at 25% [$t_{(36)} = 3.75$, *p* = 0.001, *d* = 0.62; Bonferroni-corrected alpha level: *p* < 0.0083]. No other pairwise comparisons passed a correction for multiple comparisons (all *t*'s < 2.2, *p*'s > 0.03).



Lastly, there was a significant test item type × blend level interaction effect [$F_{(5.85, 210.75)} = 3.9$, *p* = 0.001, $\eta_p^2 = 0.1$, GG]. To determine the nature of this interaction, we computed a separate

one-way ANOVA for each test item type, testing for an effect of the blend level. Accuracy for the trained associations did not differ significantly based on the percentage of the shared parent making up the faces (all F 's < 2, p 's > 0.2), but the blend level did affect rates of acquired equivalence for the untrained associations (i.e., F2-S2 pairs) [$F_{(3, 108)} = 6.73$, $p < 0.001$, $\eta_p^2 = 0.16$; Bonferroni-corrected alpha level: $p < 0.0125$]. Rates of acquired equivalence were highest for pair-mates with the shared parent blended at 75% ($M = 0.8$, $SD = 0.3$), followed by those blended at 50% ($M = 0.73$, $SD = 0.34$), those blended at 1% ($M = 0.56$, $SD = 0.38$), and lastly those blended at 25% ($M = 0.52$, $SD = 0.34$). Rates of acquired equivalence for the two highest blend levels (75 and 50%) differed significantly from the two lowest blend levels (1 and 25%; all t 's > 3, p 's < 0.003; Bonferroni-corrected alpha level: $p < 0.0083$) but did not differ significantly from one another [$t_{(36)} = 1.31$, $p = 0.2$, $d = 0.22$]. Rates of acquired equivalence for the two lowest blend levels likewise did not differ significantly from one another [$t_{(36)} = -0.56$, $p = 0.58$, $d = 0.09$]. Thus, the similarity of items across related trials affected generalization but not memory for trained associations. Despite the apparent discontinuity in rates of generalization across blend levels, only the linear effect of blend level reached significance [$F_{(1, 36)} = 13.43$, $p = 0.001$, $\eta_p^2 = 0.27$]. Neither the quadratic effect [$F_{(1, 36)} = 1.48$, $p = 0.23$, $\eta_p^2 = 0.04$] nor the cubic effect reached significance [$F_{(1, 36)} = 3.15$, $p = 0.08$, $\eta_p^2 = 0.08$].

Source Memory Test

Mean response rates for all source memory response options are presented in Table 3. Mean proportion of “both study and test” responses for each pair and quadruplet type are presented in Figure 5B. To test whether the participants tended to falsely remember untrained pairs as having been presented during both the study and test phase, we computed a 4 (shared parent blend level: 1, 25, 50, and 75%) $\times$ 4 (test item type: F1-S1, F2-S1, F1-S2, and F2-S2) repeated-measures ANOVA on the proportion of “both study and test” responses. There was a significant main effect of test item type [$F_{(3, 108)} = 33.8$, $p < 0.001$, $\eta_p^2 = 0.09$], with higher rates of “both study and test” responses for trained associations (F1-S1 $M = 0.81$, $SD = 0.21$; F2-S1 $M = 0.83$, $SD = 0.19$; F1-S2 $M = 0.78$, $SD = 0.23$) compared with untrained associations (F2-S2 $M = 0.52$, $SD = 0.21$; all t 's > 6, p 's < 0.001; Bonferroni-corrected alpha level: $p < 0.0083$). No differences between trained associations reached significance (all t 's < 1.95, p 's > 0.06). There was also a significant main effect of shared parent blend level [$F_{(3, 108)} = 5.03$, $p = 0.003$, $\eta_p^2 = 0.12$]. Numerically, items with the parent face blended at 75% were most likely to be judged as having been presented during both study and test ($M = 0.83$, $SD = 0.24$), followed by those blended at 1% ($M = 0.73$, $SD = 0.23$), those blended at 25% ($M = 0.69$, $SD = 0.21$), and lastly those blended at 50% ($M = 0.69$, $SD = 0.25$). Rates of “both” responses were significantly higher for 75% blends compared with 50% blends and 25% blends (both t 's > 2.9, p 's = 0.005; Bonferroni-corrected alpha level: $p < 0.0083$), and no other difference between blend levels passed the corrected alpha level (all t 's < 2.5, p 's > 0.02).

TABLE 3 | Experiment 2 source memory responses separated by shared parent blend level and trial type.

	Response type			
	Study only	Test only	Both study and test	Never
Face 1—Scene 1				
1%	0.04 (0.14)	0.07 (0.21)	0.85 (0.33)	0.04 (0.18)
25%	0.04 (0.14)	0.12 (0.27)	0.81 (0.32)	0.03 (0.11)
50%	0.05 (0.20)	0.14 (0.30)	0.76 (0.40)	0.05 (0.20)
75%	0.00 (0.00)	0.12 (0.27)	0.84 (0.31)	0.04 (0.14)
Face 2—Scene 1				
1%	0.03 (0.11)	0.08 (0.22)	0.85 (0.31)	0.04 (0.18)
25%	0.04 (0.14)	0.12 (0.30)	0.77 (0.37)	0.07 (0.21)
50%	0.01 (0.08)	0.16 (0.26)	0.78 (0.28)	0.04 (0.14)
75%	0.04 (0.14)	0.05 (0.20)	0.91 (0.26)	0.00 (0.00)
Face 1—Scene 2				
1%	0.01 (0.08)	0.12 (0.25)	0.78 (0.32)	0.08 (0.19)
25%	0.03 (0.11)	0.10 (0.20)	0.80 (0.34)	0.08 (0.22)
50%	0.04 (0.18)	0.16 (0.33)	0.69 (0.43)	0.11 (0.27)
75%	0.01 (0.08)	0.11 (0.29)	0.84 (0.35)	0.04 (0.14)
Face 2—Scene 2				
1%	0.04 (0.14)	0.37 (0.42)	0.42 (0.40)	0.18 (0.32)
25%	0.04 (0.14)	0.45 (0.40)	0.39 (0.38)	0.12 (0.27)
50%	0.04 (0.14)	0.30 (0.40)	0.53 (0.41)	0.14 (0.25)
75%	0.03 (0.11)	0.16 (0.29)	0.76 (0.33)	0.05 (0.16)
Recombined face	0.05 (0.15)	0.14 (0.22)	0.22 (0.33)	0.59 (0.36)
Recombined scene	0.05 (0.11)	0.11 (0.15)	0.21 (0.25)	0.63 (0.34)

Mean (SD). Bold font = correct response for a given trial type.

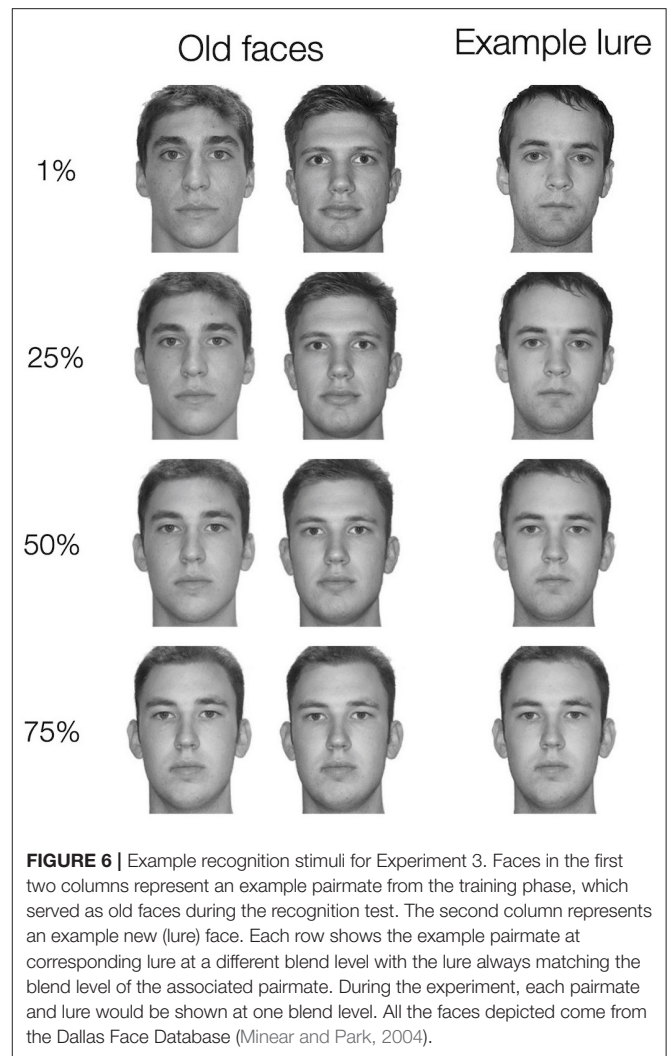
Lastly, there was a significant test item type $\times$ blend level interaction effect [$F_{(9, 324)} = 2.49$, $p = 0.009$, $\eta_p^2 = 0.07$]. To understand the nature of this interaction, we computed a separate one-way ANOVA for each test item type, looking for an effect of blend level on rates of “both” responses. The effect of blend level did not reach significance for any of the trained association types (all F 's < 2, p 's > 0.12), but there was a significant effect of blend level for the untrained association [$F_{(3, 108)} = 7.59$, $p < 0.001$, $\eta_p^2 = 0.17$; Bonferroni-corrected alpha level: $p < 0.0125$]. The participants were most likely to falsely remember having seen the untrained associations during the study when they were blended at 75% shared parent ($M = 0.76$, $SD = 0.33$), and this false memory rate was higher than for all the other blend levels (50% $M = 0.53$, $SD = 0.41$; 25% $M = 0.39$, $SD = 0.38$; 1% $M = 0.42$, $SD = 0.4$; all t 's > 2.9, p 's < 0.007; Bonferroni-corrected alpha level: $p < 0.0083$). There were no significant differences among the remaining blend levels (all t 's < 1.7, p 's > 0.1). Thus, the similarity between pair-mate faces did not affect source memory for the trained associations, but the highest level of similarity led the participants to falsely remember having seen the untrained association during the study.

Comparing Physical Resemblance Effects in Generalization and False Source Memories

As in Experiment 1, we were interested in whether the effect of physical resemblance was of a similar magnitude in acquired equivalence and source false memories. Qualitatively, the pattern in Experiment 2 was different for 50% blends than it was in Experiment 1 shared parent condition as increased generalization in that condition was not accompanied by increased false memory when compared with the low similarity condition. To test for potential differential effects of similarity on the two measures formally, we computed a test (acquired equivalence, source memory) $\times$ shared parent blend level (1, 25, 50, and 75%) ANOVA on rates of generalization and false memory for inference pairs (F2-S2). As in Experiment 1, the interaction effect was not significant [$F_{(3, 108)} = 0.88, p = 0.46, \eta_p^2 = 0.02$]. Thus, physical resemblance had a similar effect on generalization and source memory when all blend levels were considered.

Discussion

In Experiment 2, we tested whether parametrically increasing similarity across pair-mates in an acquired equivalence paradigm would lead to corresponding increases in rates of generalization and false memories for the source of inferred associations. We reasoned that resemblance can serve as a cue to reactivate prior experiences, leading to their integration. Integration across related episodes could then facilitate generalization but could lead the participants to confuse the source of inferred associations. The results were mostly but not fully consistent with this idea. Consistent with the hypothesis and with Experiment 1, we found that higher degrees of pair-mate resemblance were associated with higher rates of acquired equivalence as well as a greater tendency to falsely remember having seen untrained associations during the training phase. The effect of similarity level on generalization and false memory emerged for the untrained associations despite similarly high performance across different levels of pair-mate similarity for the trained associations. However, unlike in Experiment 1, we did not find increased source memory confusion in the 50% blend condition compared with the 1% blend condition, even though we found the levels of generalization in that condition were well above chance and significantly higher than in the 1% blend condition. Thus, whether similarity level affects generalization and source memory equally was somewhat inconclusive in Experiment 2. Finally, we noted that in the 75% blend condition, the participants endorsed untrained F2-S2 pairs as being seen at both study and test to a high level that was approaching the rate of endorsing the actually studied F1-S2 pairs (proportion of 0.78 vs. 0.84). Thus, we wanted to test an alternative explanation of the results from the 75% blend condition: that the participants simply could not tell pair-mate faces apart at the higher levels of physical resemblance. If this was the case, it might appear that the participants were actively generalizing across the two faces when, in fact, they simply did not notice that there were two faces rather than only one. We, thus, ran a follow-up experiment to determine the degree to which pair-mate faces at each level of resemblance were distinguishable from one another as well as gather additional data



to further test whether the effect of similarity on generalization and source memory confusion go hand in hand.

EXPERIMENT 3

Experiment 3 followed the overall design of Experiment 2 except that we included a face recognition test that immediately followed the training phase. After the recognition test, the participants completed the acquired equivalence test and the source memory test as in prior experiments. The recognition test included the old training faces as well as lures that were blends of the shared parent faces from training with new parent faces that were not used to generate the training sets (see **Figure 6**). We tested recognition prior to acquired equivalence and source memory, so that we would have the best possible measure of the ability to discriminate faces during learning, as there is evidence that generalization itself can reduce subsequent memory specificity (Carpenter and Schacter, 2017, 2018). However, this design feature makes the acquired equivalence and source memory data less comparable with Experiment 2 since performing a

recognition task could also affect these other types of memory judgments. Nonetheless, it served the primary goals of measuring the discriminability of pair-mate faces at each blend level, as well as testing the common or differential effects of similarity level on generalization and source memory.

Materials and Method

Participants

Prior to data collection, the sample size was determined based on the effect of blend level on generalization from Experiment 2 ($\eta_p^2 = 0.16$), which was the smallest of all observed effect sizes for the main effects of interest (similarity level effects on generalization and source memory) across Experiments 1 and 2. Using the repeated measures ANOVA protocol from GPower (Faul et al., 2007), we determined that $N = 33$ was a sufficient sample to obtain 80% power for an effect of this size. We recruited subjects from the University of Oregon who completed the experiment online for course credit. A total of 37 participants were recruited in order to have 33 subjects retained in all analyses. Four subjects were excluded and replaced because of poor learning during the training phase (<60% overall accuracy in the last block of training). Demographic data for six participants were lost because of experimenter error, leaving reportable demographics for 27 participants (13 females, mean age = 20.2 years, SD age = 3.1 years, age range = 18–34 years). Although demographics were missing for some subjects, we retained those subjects for the experimental task, leaving data from 33 subjects reported in all analyses. All the participants were provided with the consent information and asked to affirm it with a button press. The Institutional Review Board of the University of Oregon approved all procedures.

Materials

The stimulus sets were constructed as in Experiment 2 with the exception that, rather than creating a new randomization of parent faces and scene assignment for each participant, three randomized sets were created and counterbalanced across the participants.

Procedure

Data for Experiment 3 were collected online through pavlov.org. The overall procedure followed that of Experiment 2 except that the recognition test followed immediately after training rather than following the source memory test. This change ensured that any effects of blend level on recognition were not due to demand-driven integration during the generalization test (Carpenter and Schacter, 2017, 2018). While there were just three possible stimulus sets that were counterbalanced across the participants, the presentation order within each phase was randomized for each participant. Deviations from the Experiment 2 procedure for specific tasks are noted below.

Initial Exposure

Faces were shown in a random order for each of the three repetitions. It was possible for a face to be shown twice in a row only if a given face was the last one shown in one repetition and the first one shown in the following repetition.

Training

The question cues were indicated as in Experiment 1: “Where does he live?” or “Where does he vacation?”. The feedback for trials in which the participants did not respond within the 3 s allotted was “Too slow.”

Face Recognition Test

Immediately following training, the participants completed an old/new face recognition test. **Figure 6** depicts an example face recognition set. The old faces were the eight ones presented during the training phase—two pair-mate faces at each blend level. We generated lure faces by taking the eight unblended parent faces that remained after generating the training set and assigning each of them to be blended with the shared parent face of one set of pair-mates. This led to two lures at each blend level for a total of eight new items at recognition. The new unblended faces were blended with the shared parent face at the same blend level as the original pair-mates. For example, the lures for the pair-mates blended with 1% shared parent were the new unblended faces blended with the shared parent also at 1%. This allowed us to test whether the participants could distinguish between faces made up of a given shared parent at the blend level experienced during training. If they were able to do so for new lures, then they were likely able to distinguish between the two pair-mate faces from the training set as well.

Each face was presented once during the recognition phase. Each face was presented for 3 s during which time the participants could make their old/new response. This timing matched the length of face presentation and response time from the training phase. Each face was followed by a 1-s fixation cross.

Acquired Equivalence Test

The acquired equivalence test was as in Experiment 2, except that the question cues at the top of the screen were as in the training phase of this experiment.

Source Memory Test

The source memory test was as in Experiment 2 with two exceptions. First, associations were tested only once. Second, the recombined face condition consisted of two scenes that had been together as alternative choices but with a different face. This condition was always distinct from the F2-S2 condition. The recombined scene condition consisted of a face with its target scene but with a new distractor. The distractor always came from the same category (city scene, nature scene) as the target.

Design and Statistical Analysis

The design and analytical approach were as in Experiment 2, except that there was an additional dependent variable of interest: recognition scores that indicated the ability to distinguish training faces from new faces using the same shared parent face at the same blend level.

Results

Training

Mean accuracies for each type of trained association are presented in **Table 4**. Consistent with Experiments 1 and 2, there was a significant main effect of training block [$F_{(3,1, 99,1)} =$

TABLE 4 | Experiment 3 training accuracy separated by block, shared parent blend level, and trained association type.

	Training block					
	1	2	3	4	5	6
Face 1—Scene 1						
1%	0.52 (0.40)	0.68 (0.39)	0.79 (0.35)	0.78 (0.35)	0.84 (0.31)	0.91 (0.20)
25%	0.55 (0.38)	0.64 (0.31)	0.58 (0.40)	0.71 (0.36)	0.77 (0.40)	0.81 (0.33)
50%	0.50 (0.33)	0.68 (0.35)	0.65 (0.42)	0.71 (0.40)	0.72 (0.37)	0.91 (0.23)
75%	0.49 (0.39)	0.62 (0.42)	0.82 (0.30)	0.74 (0.36)	0.86 (0.31)	0.95 (0.19)
Face 2—Scene 1						
1%	0.55 (0.36)	0.74 (0.38)	0.68 (0.41)	0.75 (0.35)	0.87 (0.31)	0.88 (0.22)
25%	0.58 (0.38)	0.65 (0.40)	0.73 (0.33)	0.71 (0.33)	0.79 (0.31)	0.85 (0.30)
50%	0.50 (0.40)	0.70 (0.33)	0.67 (0.37)	0.80 (0.28)	0.82 (0.33)	0.82 (0.33)
75%	0.56 (0.32)	0.62 (0.42)	0.73 (0.36)	0.85 (0.26)	0.78 (0.36)	0.91 (0.26)
Face 1—Scene 2						
1%	0.59 (0.44)	0.68 (0.33)	0.73 (0.38)	0.82 (0.32)	0.82 (0.32)	0.96 (0.15)
25%	0.42 (0.40)	0.73 (0.36)	0.86 (0.29)	0.82 (0.30)	0.87 (0.28)	0.99 (0.09)
50%	0.64 (0.31)	0.74 (0.33)	0.76 (0.28)	0.77 (0.33)	0.92 (0.22)	0.85 (0.31)
75%	0.53 (0.35)	0.49 (0.34)	0.61 (0.45)	0.70 (0.35)	0.72 (0.39)	0.86 (0.29)

Mean (SD).

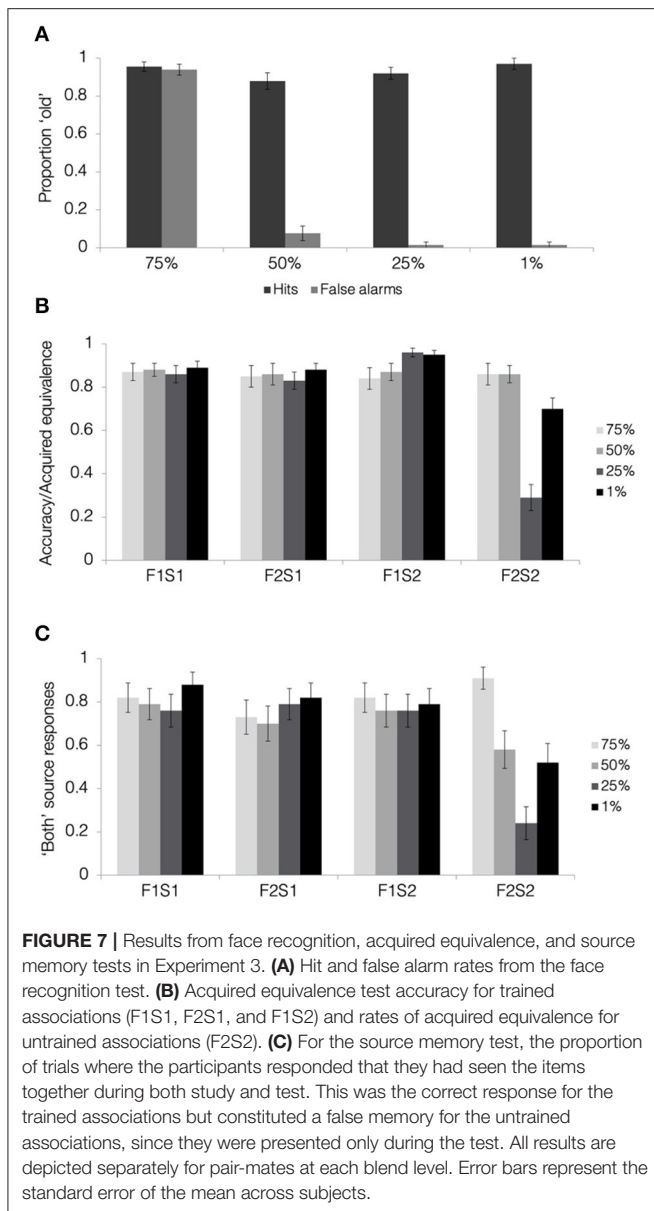
54.85, $p < 0.001$, $\eta_p^2 = 0.63$, GG] accompanied by a significant linear effect [$F_{(1, 32)} = 214.52$, $p < 0.001$, $\eta_p^2 = 0.87$] showing increasing accuracy across training. Neither the main effect of trained association type nor the main effect of blend level was significant (both F 's < 1.4 , p 's > 0.27). There was, however, a significant trained association type $\times$ blend level interaction effect [$F_{(6, 192)} = 3.16$, $p = 0.006$, $\eta_p^2 = 0.09$]. To better understand the nature of this interaction, we computed separate one-way ANOVAs for each type of trained association collapsed across training blocks, testing whether there was an overall effect of the blend level for each. The effect of blend level was only significant for F1-S2 pairs [$F_{(3, 96)} = 4.61$, $p = 0.005$, $\eta_p^2 = 0.13$; other F 's < 1.9 , p 's > 0.14]. Training accuracy for the pair-mates blended with 75% shared parent was poorer than all the other pairs (t 's > 3 , p 's < 0.005), although the difference from the 1% blend level did not pass a correction for multiple comparisons [$t_{(32)} = 2.56$, $p = 0.016$, $d = 0.26$; Bonferroni corrected alpha level: $p < 0.0083$]. No other pairwise difference was significant (all t 's < 0.5 , p 's > 0.7). Thus, blend level did not strongly affect learning for F1-S1 or F2-S1 associations, but there was some evidence of poorer learning for F1-S2 associations when there was a 75% shared parent. No other interaction effect reached significance (all F 's < 1.5 , p 's > 0.06).

Face Recognition

Hit and false alarm rates separated by blend level are presented in **Figure 7A**. Face recognition scores were calculated as the corrected hit rate (hit rate—false alarm rate) separately for each blend level. We first tested whether the participants were able to discriminate old faces from lure faces at above-chance levels for each of the four blend levels. One-sample t -tests comparing corrected hit rates with zero (i.e., no old/new discrimination) showed above-chance performance for 1, 25, and 50% blends

(all t 's > 14 , p 's < 0.001 ; Bonferroni corrected alpha level: $p < 0.0125$), with all of these conditions showing hit rates above 85% and false alarm rates below 10%. In contrast, corrected hit rates for the 75% blends were not different from chance [$t_{(32)} = 0.44$, $p = 0.66$, $d = 1.33$], driven by a false alarm rate of over 90%. Comparing the corrected hit rates with a one-way, repeated measures ANOVA, we found a significant effect of blend level [$F_{(1.9, 60.3)} = 140.51$, $p < 0.001$, $\eta_p^2 = 0.82$]. This overall effect was qualified by significant linear [$F_{(1, 32)} = 642.22$, $p < 0.001$, $\eta_p^2 = 0.95$], quadratic [$F_{(1, 32)} = 52.63$, $p < 0.001$, $\eta_p^2 = 0.62$], and cubic effects [$F_{(1, 32)} = 22.01$, $p < 0.001$, $\eta_p^2 = 0.41$]. Pairwise comparisons revealed significantly poorer recognition scores for 75% blends compared with all others (all t 's > 11 , p 's < 0.001). Recognition scores for 50% blends were poorer than those for 1% blends [$t_{(32)} = 2.39$, $p = 0.023$, $d = 0.37$] and 25% blends [$t_{(32)} = 2.51$, $p = 0.017$, $d = 0.25$], but these differences did not pass a correction for multiple comparisons (Bonferroni corrected alpha level: $p < 0.0083$).

To summarize, we see clear evidence that the participants were not able to discriminate between faces at the highest degree of physical similarity at 75% blend level. This indicates that the high generalization and high false memory found in Experiment 2 for 75% blends was likely driven by a failure to discriminate between pair-mate faces, simply confusing Face 2 for Face 1. For the rest of this report, we will report results from all the four blend levels in tables and figures for completeness, but we will not consider the 75% blend condition in further analyses. Importantly, we also see clear evidence that the participants were able to distinguish between faces at the other three of the blend levels, 1, 25, and 50%. Thus, increased generalization between pair-mates at 50% blend level observed in Experiments 1 and 2 was unlikely driven by a lack of discrimination between them.



Acquired Equivalence

Figure 7B presents rates results from the acquired equivalence test in terms of accuracy for trained associations (F1-S1, F2-S1, and F1-S2) and the proportion of trials showing acquired equivalence for untrained associations (F2-S2). First, we tested whether rates of acquired equivalence differed significantly from chance (0.5) using a one-sample t -test at each blend level. Results revealed that rates of acquired equivalence were significantly higher than chance for pair-mates blended at 1 and 50% (all t 's > 3.5 , p 's < 0.002 ; Bonferroni-corrected alpha level: $p < 0.0167$). Those blended at 25% showed significantly *below chance* acquired equivalence [$t_{(32)} = -3.46$, $p = 0.002$, $d = 0.36$]. In other words, the participants were more likely to select a scene for Face 2 that was *not* associated with their pair-mate when the pair-mate was at 25% shared level of similarity.

Comparing across association types and blend levels, there was a significant main effect of association type [$F_{(2.5, 79.7)} = 44.43$, $p < 0.001$, $\eta_p^2 = 0.58$, GG]. Rates of acquired equivalence for the F2-S2 pairs were lower than rates of correct associative memory for all the trained pairs (all t 's > 7.1 , p 's < 0.001 ; Bonferroni-corrected alpha level: $p < 0.0083$). Pairwise differences among trained associations (F1-S1, F2-S1, and F1-S2) did not reach significance following correction for multiple comparisons (all t 's < 2.5 , p 's > 0.02). There was also a main effect of blend level [$F_{(2, 64)} = 9.94$, $p < 0.001$, $\eta_p^2 = 0.24$]. Scores were lower for pair-mates blended with 25% shared parent compared with all other blend levels (both t 's > 3.7 , p 's < 0.002 ; Bonferroni-corrected alpha level: $p < 0.0167$). The difference between 1 and 50% blends was not significant [$t_{(32)} = 0.41$, $p = 0.68$, $d = 0.2$].

Critically, we found a significant association type $\times$ blend level interaction effect [$F_{(3.6, 114)} = 19.15$, $p < 0.001$, $\eta_p^2 = 0.37$, GG]. To understand the nature of this interaction, we computed separate one-way ANOVAs for each association type, looking for an effect of blend level. The effect of blend level was not significant for any of the trained associations following a correction for multiple comparisons (all F 's < 3.4 , p 's > 0.04 ; Bonferroni-corrected alpha level: $p < 0.0125$). There was, however, a significant effect of blend level in acquired equivalence (i.e., F2-S2 pairs) [$F_{(1.8, 57.2)} = 28.23$, $p < 0.001$, $\eta_p^2 = 0.47$, GG]. Rates of acquired equivalence were higher for 50% blends compared with 1% blends [$t_{(32)} = 2.55$, $p = 0.016$, $d = 0.37$; Bonferroni-corrected alpha level: $p < 0.0167$] and 25% blends [$t_{(32)} = 6.7$, $p < 0.001$, $d = 0.5$]. Rates of acquired equivalence were also higher for 1% blends compared with 25% blends [$t_{(32)} = 4.85$, $p < 0.001$, $d = 0.49$]. The low rates of acquired equivalence for 25% blends was unexpected and not seen in Experiment 2. The exact reason for the low generalization in this condition are not immediately clear. Nonetheless, we replicate the key finding from Experiment 2 that rates of acquired equivalence are highest for 50% compared with the lower 1 and 25% blend levels.

Another way to think about the association type $\times$ blend level interaction is to ask whether rates of acquired equivalence matched accuracy rates for trained associations at each blend level. We know that, overall, rates of acquired equivalence were lower than rates of memory for trained associations, but is that true across levels of physical resemblance? Comparing across association types for each blend level, we find lower rates of acquired equivalence compared with trained accuracy for 1% blends and 25% blends (both F 's > 11.4 , p 's < 0.001). However, there was no effect of association type for the 50% blends [$F_{(2.5, 80.4)} = 0.16$, $p = 0.89$, $\eta_p^2 = 0.005$, GG], indicating that the participants chose the acquired equivalence response for untrained pairs to a comparable degree as they remembered the trained associations. This further demonstrates the very robust generalization for the 50% blends.

Lastly, we compared rates of acquired equivalence to rates of false face recognition across blend levels, testing the degree to which differences in generalization track differences in false recognition of faces. We were interested specifically in the test $\times$ blend level interaction effect, which was significant [$F_{(2, 64)} =$

22.25, $p < 0.001$, $\eta_p^2 = 0.41$]. This pattern was driven by rates of acquired equivalence that increased from 25 to 1 to 50% blends (see stats above), whereas mean false alarm rates for lure faces were the same for 1 and 25% blends ($M = 1.5\%$, $SD = 8.7\%$) and only slightly higher for 50% blends ($M = 7.6\%$, $SD = 22\%$). Thus, the patterns in generalization and false recognition of faces did not mirror one another.

Source Memory

Mean response rates for all source memory response options are presented in **Table 5**. Mean proportion of “both study and test” responses for each test item type and each blend level is presented in **Figure 7C**. Comparing rates of “both” responses across test item types (F1-S1, F2-S1, F1-S2, and F2-S2), and blend levels (1, 25, and 50%) using repeated measures ANOVA, there was a significant main effect of test item type [$F_{(2.2, 69.5)} = 21.03$, $p < 0.001$, $\eta_p^2 = 0.4$, GG]. Rates of “both study and test” responses were higher for all trained associations (for which it was the correct response) compared with untrained associations (all t 's > 5 , p 's < 0.001 ; Bonferroni-corrected alpha level: $p < 0.0083$) when collapsed across blend levels. Pairwise differences among trained associations were not significant (all t 's < 1.2 , p 's > 0.2). The main effect of blend level was not significant [$F_{(1.6, 51.7)} = 2$, $p = 0.14$, $\eta_p^2 = 0.06$, GG]. The test item type $\times$ blend level interaction was also not significant [$F_{(4, 129.5)} = 2.15$, $p = 0.078$, $\eta_p^2 = 0.06$]. Although this interaction was significant in Experiment 2, it was driven by higher rates of source errors for 75% blends compared with all others. Ignoring that condition in which the faces were not discriminable from one another in memory, Experiments 2–3 show little effect of the blend level on source memory judgments.

Comparing Physical Resemblance Effects in Generalization and False Source Memories

As in Experiments 1–2, we compared blend level effects for generalization and source memory to determine the degree to which physical resemblance had a similar or different effect in generalization and source memory. As in prior experiments, the test $\times$ blend level interaction effect was not significant [$F_{(2, 64)} = 1.46$, $p = 0.24$, $\eta_p^2 = 0.04$]. Thus, although there was a significant effect of blend level on generalization but not source memory errors, the overall patterns did not differ reliably across tests.

Discussion

The aim of Experiment 3 was to test whether the effect of physical resemblance on generalization and source memory could be explained by the poor discriminability of pair-mate faces at higher blend levels. Poor discriminability was clear for 75% blends, which were not reliably discriminable from one another in memory. Thus, source memory and generalization results for this condition were likely driven purely by the confusability of the pair-mate faces. In contrast, we replicated high rates of generalization for 50% blends while demonstrating that faces blended at 50% shared parent retained good discriminability. Interestingly, generalization rates for 50% blends in Experiment 3 were comparable with memory for trained relationships in this condition (86% generalization vs. 87% trained accuracy).

TABLE 5 | Experiment 3 source memory responses separated by shared parent blend level and trial type.

	Response type			
	Study only	Test only	Both study and test	Never
Face 1—Scene 1				
1%	0.03 (0.17)	0.03 (0.17)	0.88 (0.33)	0.06 (0.24)
25%	0.12 (0.33)	0.03 (0.17)	0.76 (0.44)	0.09 (0.29)
50%	0.06 (0.24)	0.06 (0.24)	0.79 (0.42)	0.09 (0.29)
75%	0.06 (0.24)	0.06 (0.24)	0.82 (0.39)	0.06 (0.24)
Face 2—Scene 1				
1%	0.06 (0.24)	0.09 (0.29)	0.82 (0.39)	0.03 (0.17)
25%	0.06 (0.24)	0.06 (0.24)	0.79 (0.42)	0.09 (0.29)
50%	0.03 (0.17)	0.18 (0.39)	0.70 (0.47)	0.09 (0.29)
75%	0.09 (0.29)	0.09 (0.29)	0.73 (0.45)	0.09 (0.29)
Face 1—Scene 2				
1%	0.06 (0.24)	0.12 (0.33)	0.79 (0.42)	0.03 (0.17)
25%	0.03 (0.17)	0.09 (0.29)	0.76 (0.44)	0.12 (0.33)
50%	0.03 (0.17)	0.09 (0.29)	0.76 (0.44)	0.12 (0.33)
75%	0.00 (0.00)	0.09 (0.29)	0.82 (0.39)	0.09 (0.29)
Face 2—Scene 2				
1%	0.09 (0.29)	0.21 (0.42)	0.52 (0.51)	0.18 (0.39)
25%	0.12 (0.33)	0.48 (0.51)	0.24 (0.44)	0.15 (0.36)
50%	0.03 (0.17)	0.30 (0.47)	0.58 (0.50)	0.09 (0.29)
75%	0.03 (0.17)	0.00 (0.00)	0.91 (0.29)	0.06 (0.24)
Recombined face	0.06 (0.11)	0.13 (0.13)	0.23 (0.26)	0.58 (0.35)
Recombined scene	0.10 (0.18)	0.06 (0.09)	0.28 (0.22)	0.56 (0.32)

Mean (SD) Bold font = correct response for a given trial type.

Furthermore, the results indicate that increased similarity does not have to lead to source memory confusion, at least not at the 50% blend level where discriminability is maintained. Overall, results from Experiment 3 bolster the idea that physical resemblance between related experiences can foster generalization. The results also indicate that better generalization is not always associated with tradeoffs in terms of increased false memories.

GENERAL DISCUSSION

In this study, we manipulated the degree of resemblance between items constituting related experiences and tested both the tendency to generalize across those experiences and the specificity of memory for separate experiences. We predicted that higher degrees of overlap would make it more likely that the participants would reactivate the prior related episode when encountering new, related information and, therefore, increase the likelihood that representations of those episodes would become integrated. As a signature of memory integration, we expected higher rates of generalization accompanied by a source memory confusion, such as mistaking the inferred information for a directly observed one. We found partial support for this hypothesis. Rates of generalization (acquired equivalence)

were higher when there was more physical resemblance across related episodes, suggesting that physical resemblance helped the participants make links across experiences. However, results were somewhat equivocal as to whether the same pattern was present for source memory errors once the discriminability of pair-mate faces was taken into account.

Across three experiments, we showed that higher levels of physical resemblance across episodes are associated with higher subsequent generalization across those episodes. This effect was present even when discarding the 75% blends that Experiment 3 indicated were not discriminable from one another. In fact, the 50% blends showed comparable levels of generalization with the 75% blends without the issue of poor discriminability. That is, participants generalized across the 50% blends at a rate similar to the condition where the participants could not even tell the pair-mate faces apart. Further, rates of generalization for 50% blends in Experiment 3 were comparable with memory for their respective trained associations, indicating that the participants generalized as well as could be expected given their memory for the premises. Together, these findings provide strong evidence that physical resemblance can help individuals make connections between related experiences, even when similar items are clearly discriminable from one another.

Prior studies of acquired equivalence have typically used stimuli without systematic resemblance between related experiences, akin to the no-shared and 1% conditions. Some of these studies have, nonetheless, found robust generalization (Edwards et al., 1982; Shohamy and Wagner, 2008; Meeter et al., 2009), while other studies found rates of generalization that were above chance but relatively low (Duncan et al., 2012; de Araujo Sanchez and Zeithamova, 2020). For instance, a study from our laboratory has previously shown in a large sample ($N = 190$) that rates of acquired equivalence can be quite modest: around 55% (when chance is 50%), with many subjects not showing any hint of generalization (de Araujo Sanchez and Zeithamova, 2020). Interestingly, we rarely saw above-chance rates of acquired equivalence at lower similarity levels (for the no-shared parent blends, 1% blends, or 25% blends), with the only exception being the 1% blend condition in Experiment 3. One possible explanation for lower generalization rates in conditions with low resemblance is that there may be generalization tradeoffs across quadruplets. If making connections between related experiences is cognitively demanding, then the participants may have been selective about when they made such links. In this way, physical resemblance may have served as a cue as to which experiences to link, promoting generalization for some pairs but inhibiting generalization for others. Future studies could manipulate pair-mate similarity between subjects to test whether the effect of physical resemblance is driven in whole or in part by the contrast between high and low resemblance within the same set.

Across the experiments, we found that generalization rates were consistently higher for the 50%/shared parent blends than the 1%/no shared parent blends. However, we did not find that rates of generalization for the 25% blends were intermediate between the 1 and 50% blends. Instead, generalization rates for the 25% blends were either comparable with the rate for the 1% blends (Experiment 2) or significantly lower than the

1% blends (Experiment 3). Matched generalization rates for 1 and 25% blends can be explained by the 25% blend level not generating a level of similarity that the participants could detect, like a mirror of the lack of discriminability we found between faces at the 75% blend level. However, the lower generalization rates for 25% blends compared with 1% blends were unexpected and more difficult to explain. Future studies will be needed to determine if this was simply due to sampling error or whether it is a real, replicable pattern, perhaps reflecting some form of pattern separation or repulsion effect (Chanales et al., 2017).

While we found that physical resemblance led to increases in generalization, research on reducing memory interference often shows that representations of highly similar items are orthogonalized to make them discriminable from one another (for reviews see Colgin et al., 2008; Yassa and Stark, 2011; Leal and Yassa, 2018). For example, Favila et al. (2016) showed that hippocampal representations for two highly similar scene images became more dissimilar from one another when paired with a common face, presumably to aid in discriminating between the scenes despite their shared perceptual details and shared association. The degree of integration vs. separation is driven to some extent by whether learning demands emphasize commonalities across related items or discrimination between them (Ashby et al., 2020; Chanales et al., 2020). During acquired equivalence learning in this study, instructions did not explicitly emphasize either discrimination between similar faces or generalization across them, and the participants were not aware of upcoming generalization or source memory tests. Under these conditions, it seems that the participants defaulted to linking across similar experiences in service of generalization.

While the data support the role of physical resemblance in promoting generalization, the support for the proposed mechanism—that similarity increases reactivation of related memories and leads to integration across the related experiences—is less clear. Prior studies of episodic inference, such as those using the acquired equivalence paradigm, have often focused on the role of memory integration in supporting generalization (Shohamy and Wagner, 2008; Zeithamova et al., 2012a; Schlichting et al., 2015). Such studies have shown that integrating representations of related episodes at the time of learning can facilitate later generalization, which tends to be faster and more accurate when based on integrated representations compared with separate representations (Shohamy and Wagner, 2008; Schlichting et al., 2014). Prior study has also identified several contextual factors that can increase integrative encoding of related episodes, such as explicit instructions to integrate (Richter et al., 2016), the temporal proximity of related episodes (Zeithamova and Preston, 2017), and whether related associations are studied in a blocked vs. intermixed manner (Schlichting et al., 2015). Here, we tested another such factor, physical resemblance, and found that rates of generalization were higher for faces with increased resemblance, consistent with the idea that similarity promotes integrative encoding, which in turn supports subsequent generalization. A similar effect of physical resemblance was recently reported in another inference paradigm, associative inference, where participants are explicitly asked to relate A and C items after

separately learning A-B and B-C associations (Molitor et al., 2021). Thus, this aspect of the findings is consistent with the proposal that physical resemblance serves as a pattern completion cue to reactivate prior related episodes, which then becomes integrated with current experience.

However, not all aspects of the data point clearly to an integration mechanism. Prior study has shown that successful inference can lead to poorer source memory (Carpenter and Schacter, 2017, 2018). Anecdotal evidence from Shohamy and Wagner (2008) also suggested that successful generalization was accompanied by mistaking inferred face preferences for actually observed ones. These findings have been taken as evidence that integrating across episodes causes unique aspects of related experiences to be discarded, and assumptions that generalization and false memory may be two sides of the same coin (Zeithamova et al., 2012b; Varga et al., 2019). We thus expected to find increases in source memory errors with increasing rates of generalization, indicating that integration caused the loss of contextual information or that the participants mistakenly attributed internal reactivation of a pair-mate for an external presentation of the untrained association during encoding. When only considering blends where pair-mate faces were discriminable from one another, results for this aspect of the hypothesis were mixed. In Experiment 1, there was a significant increase in false source memories between the no shared parent and the shared parent pair-mates, and the increase mirrored the increase in generalization. This finding is in line with integration as the mechanism driving the effects of physical similarity. However, in Experiments 2 and 3, there was no effect of the blend level on source memory errors across the 1, 25, and 50% blends. Instead, rates of “both” responses were consistently lower for untrained associations compared with trained associations across blend levels. This contrasts with the generalization scores for 50% blends, which matched or nearly matched accuracy for trained associations, indicating that high generalization was not always accompanied by source memory confusion. Yet, complicating the story further, it is also not possible to conclude that the effect of physical similarity was different for generalization and source memory, as the difference across tests was not significant in any experiment. Instead, we are left with a clear effect of physical similarity in generalization and equivocal findings from source memory.

What might cause these mixed findings? Importantly, integrative encoding is but one mechanism proposed to support generalization. Other mechanisms of generalization have been postulated, with potentially distinct predictions about memory for individual events, memory generalization, and their relationship (Zeithamova and Bowman, 2020). Some argue that integrated representations are not needed; instead, generalization can be achieved on-demand, based on flexible retrieval of separate episodes (Kumaran, 2012; Kumaran and McClelland, 2012). The degree of physical similarity across pair-mates could potentially affect on-the-fly generalization from separate memories if the overlap across experiences increased the probability of successfully chaining all relevant memories. In this case, the generalization would not need to come at the expense of detailed memory for individual experiences. Indeed,

our prior study that did not include a similarity manipulation showed that rates of generalization and source memory scores were not related either across subjects or on a trial-to-trial basis (de Araujo Sanchez and Zeithamova, 2020). Still, others have found that generalization can be positively related to the fidelity of individual memories (Banino et al., 2016). Thus, the chained retrieval of separate representations may have contributed to generalization in this study to some degree, explaining why high levels of generalization were not always accompanied by high levels of false memory.

Alternatively, similar predictions would stem from recent proposals that people may represent the same events at multiple levels of specificity, forming integrated representations alongside separate ones rather than at their expense (Collin et al., 2015; Schapiro et al., 2017; Brunec et al., 2018; Bowman et al., 2020; Zeithamova and Bowman, 2020). Different representations may be differentially susceptible to the effects of physical similarity, in which case the benefit of increased similarity for generalization (more likely relying on integrated memories) may not be accompanied by a corresponding effect on source memory (more likely relying on separate memories of individual events). While the current data are inconclusive with respect to whether generalization and source memory judgments were based on integrated memories, separated memories or both, they highlight the benefit of considering multiple measures in interpreting results rather than drawing conclusions about underlying mechanisms (such as integration) based on the generalization score alone. Importantly, the data clearly show that physical similarity promotes generalization across episodes, whether it is through promoting the formation of integrated representations and/or through enhancing flexible retrieval for on-the-fly generalization.

SUMMARY

In the three experiments, we manipulated the degree of overlap across related episodes in an acquired equivalence paradigm and tested the tendency to generalize across experiences and the ability to remember the source of generalized information. All three experiments showed increases in generalization for experiences with greater overlap but differed in whether errors in source memory accompanied increases in generalization. These results suggest a clear facilitatory effect of resemblance across episodes in the generalization that may sometimes, but not always, come with a loss of memory specificity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://osf.io/2ubgv/>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Compliance

Services, University of Oregon. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All the authors contributed to the conceptualization, design of the experiments, and edited the manuscript. M-AA, SR, and WH programmed the experiments

and oversaw data collection. CB, M-AA, SR, and WH contributed to data analysis. CB wrote the first draft of the manuscript.

FUNDING

This study was supported by the National Institutes of Health [Grant Nos: R01-NS112366 (DZ) and F32-AG054204 (CB)].

REFERENCES

- Arndt, J., and Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *J. Memory Lang.* 48, 1–15. doi: 10.1016/S0749-596X(02)00518-1
- Ashby, S. R., Bowman, C. R., and Zeithamova, D. (2020). Perceived similarity ratings predict generalization success after traditional category learning and a new paired-associate learning task. *Psychonomic Bull. Rev.* 27, 791–800. doi: 10.3758/s13423-020-01754-3
- Banino, A., Koster, R., Hassabis, D., and Kumaran, D. (2016). Retrieval-based model accounts for striking profile of episodic memory and generalization. *Sci. Rep.* 6, 1–15. doi: 10.1038/srep31330
- Bowman, C. R., Chamberlain, J. D., and Dennis, N. A. (2019). Sensory representations supporting memory specificity: age effects on behavioral and neural discriminability. *J. Neurosci.* 39, 2265–2275. doi: 10.1523/JNEUROSCI.2022-18.2019
- Bowman, C. R., and Dennis, N. A. (2015). The neural correlates of correctly rejecting lures during memory retrieval: the role of item relatedness. *Exp. Brain Res.* 233, 1963–1975. doi: 10.1007/s00221-015-4268-y
- Bowman, C. R., Iwashita, T., and Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. *eLife*. 9:e59360. doi: 10.7554/eLife.59360.sa2
- Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z. X., et al. (2018). Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Curr. Biol.* 28, 2129–2135.e6. doi: 10.1016/j.cub.2018.05.016
- Carpenter, A. C., and Schacter, D. L. (2017). Flexible retrieval: when true inferences produce false memories. *J. Exp. Psychol. Learn. Memory Cogn.* 43, 335–349. doi: 10.1037/xlm0000340
- Carpenter, A. C., and Schacter, D. L. (2018). False memories, false preferences: flexible retrieval mechanisms supporting successful inference bias novel decisions. *J. Exp. Psychol. General* 147, 988–1004. doi: 10.1037/xge0000391
- Chanales, A. J. H., Oza, A., Favila, S. E., and Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Curr. Biol.* 27, 2307–2317. doi: 10.1016/j.cub.2017.06.057
- Chanales, A. J. H., Tremblay-McGaw, A. G., and Kuhl, B. A. (2020). Adaptive repulsion of long-term memory representations is triggered by event similarity. *bioRxiv*. doi: 10.1101/2020.01.14.900381
- Colgin, L. L., Moser, E. I., and Moser, M. B. (2008). Understanding memory through hippocampal remapping. *Trends Neurosci.* 31, 469–477. doi: 10.1016/j.tins.2008.06.008
- Collin, S. H. P., Milivojevic, B., and Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci.* 18, 1562–1564. doi: 10.1038/nn.4138
- de Araujo Sanchez, M. A., and Zeithamova, D. (2020). Generalization and source memory in acquired equivalence. *PsyArXiv*. doi: 10.31234/osf.io/txjsh
- Duncan, K., Sadanand, A., and Davachi, L. (2012). Memory's penumbra: episodic memory decisions induce lingering mnemonic biases. *Science* 337, 485–488. doi: 10.1126/science.1221936
- Edwards, C. A., Jagielo, J. A., Zentall, T. R., and Hogan, D. E. (1982). Acquired equivalence and distinctiveness in matching to sample by pigeons: mediation by reinforcer-specific expectancies. *J. Exp. Psychol. Anim. Behav. Processes* 8:244. doi: 10.1037/0097-7403.8.3.244
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Favila, S. E., Chanales, A. J. H., and Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat. Commun.* 7, 1–10. doi: 10.1038/ncomms11066
- Gutchess, A. H., and Schacter, D. L. (2012). The neural correlates of gist-based true and false recognition. *NeuroImage* 59, 3418–3426. doi: 10.1016/j.neuroimage.2011.11.078
- Guzowski, J. F., Knierim, J. J., and Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron* 44, 581–584. doi: 10.1016/j.neuron.2004.11.003
- Honey, R. C., and Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *J. Exp. Psychol. Anim. Behav. Processes* 15:338. doi: 10.1037/0097-7403.15.4.338
- Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., and Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat. Commun.* 6, 1–11. doi: 10.1038/ncomms8462
- Hunsaker, M. R., and Kesner, R. P. (2013). The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neurosci. Biobehav. Rev.* 37, 36–58. doi: 10.1016/j.neubiorev.2012.09.014
- Kumaran, D. (2012). What representations and computations underpin the contribution of the hippocampus to generalization and inference? *Front. Human Neurosci.* 6:157. doi: 10.3389/fnhum.2012.00157
- Kumaran, D., and McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychol. Rev.* 119, 573–616. doi: 10.1037/a0028681
- Leal, S. L., and Yassa, M. A. (2018). Integrating new findings and examining clinical applications of pattern separation. *Nat. Neurosci.* 21, 163–173. doi: 10.1038/s41593-017-0065-1
- Lee, I., Yoganarasimha, D., Rao, G., and Knierim, J. J. (2004). Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. *Nature* 430, 456–459. doi: 10.1038/nature02739
- Liu, K. Y., Gould, R. L., Coulson, M. C., Ward, E. V., and Howard, R. J. (2016). Tests of pattern separation and pattern completion in humans—a systematic review. *Hippocampus* 26, 705–717. doi: 10.1002/hipo.22561
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* 262, 23–81. doi: 10.1098/rstb.1971.0078
- McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295X.102.3.419
- Meeter, M., Shohamy, D., and Myers, C. (2009). Acquired equivalence changes stimulus representations. *J. Exp. Anal. Behav.* 91, 127–141. doi: 10.1901/jeab.2009.91.127
- Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods. Instrum. Comput.* 36, 630–633.
- Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A., and Preston, A. R. (2021). Memory reactivation during learning simultaneously promotes dentate Gyrus/CA 2,3 pattern differentiation and CA 1 memory integration. *J. Neurosci.* 41, 726–738. doi: 10.1523/JNEUROSCI.0394-20.2020
- Richter, F. R., Chanales, A. J. H., and Kuhl, B. A. (2016). Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage* 124, 323–335. doi: 10.1016/j.neuroimage.2015.08.051

- Roediger, H. L., McDermott, K. B., Hintzman, D. L., Lindsay, S., Rajaram, S., and Tulving, E. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Memory Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., and Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160049. doi: 10.1098/rstb.2016.0049
- Schlichting, M. L., Mumford, J. A., and Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nat. Commun.* 6:8151. doi: 10.1038/ncomms9151
- Schlichting, M. L., Zeithamova, D., and Preston, A. R. (2014). CA1 subfield contributions to memory integration and inference. *Hippocampus* 24, 1248–1260. doi: 10.1002/hipo.22310
- Shohamy, D., and Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* 60, 378–389. doi: 10.1016/j.neuron.2008.09.023
- Toner, C. K., Pirogovsky, E., Kirwan, C. B., and Gilbert, P. E. (2009). Visual object pattern separation deficits in nondemented older adults. *Learn. Memory*. doi: 10.1101/lm.1315109
- Turney, I. C., and Dennis, N. A. (2017). Elucidating the neural correlates of related false memories using a systematic measure of perceptual relatedness. *NeuroImage* 146, 940–950. doi: 10.1016/j.neuroimage.2016.09.005
- Varga, N. L., Gaugler, T., and Talarico, J. (2019). Are mnemonic failures and benefits two sides of the same coin?: investigating the real-world consequences of individual differences in memory integration. *Memory Cogn.* 47, 496–510. doi: 10.3758/s13421-018-0887-4
- Vazdarjanova, A., and Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *J. Neurosci.* 24, 6489–6496. doi: 10.1523/JNEUROSCI.0350-04.2004
- Vieweg, P., Stangl, M., Howard, L. R., and Wolbers, T. (2015). Changes in pattern completion - a key mechanism to explain age-related recognition memory deficits? *Cortex*. doi: 10.1016/j.cortex.2014.12.007
- Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., and Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus* 21, 968–979. doi: 10.1002/hipo.20808
- Yassa, M. A., and Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends Neurosci.* 34, 515–525. doi: 10.1016/j.tins.2011.06.006
- Zeithamova, D., and Bowman, C. R. (2020). Generalization and the hippocampus: more than one story? *Neurobiol. Learn. Memory* 175. doi: 10.1016/j.nlm.2020.107317
- Zeithamova, D., Dominick, A. L. L., and Preston, A. R. R. (2012a). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* 75, 168–179. doi: 10.1016/j.neuron.2012.05.010
- Zeithamova, D., and Preston, A. R. (2010). Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *J. Neurosci.* 30, 14676–14684. doi: 10.1523/JNEUROSCI.3250-10.2010
- Zeithamova, D., and Preston, A. R. (2017). Temporal proximity promotes integration of overlapping events. *J. Cogn. Neurosci.* 29, 1311–1323. doi: 10.1162/jocn_a_01116
- Zeithamova, D., Schlichting, M. L., and Preston, A. R. (2012b). The hippocampus and inferential reasoning: building memories to navigate future decisions. *Front. Human Neurosci.* 6:70. doi: 10.3389/fnhum.2012.00070

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bowman, de Araujo Sanchez, Hou, Rubin and Zeithamova. 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) and the copyright owner(s) 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.



Feelings of Contrast at Test Reduce False Memory in the Deese/Roediger-McDermott Paradigm

Sara Cadavid^{1†}, Maria Soledad Beato^{2†}, Mar Suarez^{2†} and Pedro B. Albuquerque^{3†}

¹ School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia, ² Faculty of Psychology, University of Salamanca, Salamanca, Spain, ³ School of Psychology, University of Minho, Braga, Portugal

OPEN ACCESS

Edited by:

Eddy J. Davelaar,
Birkbeck, University of London,
United Kingdom

Reviewed by:

Mark J. Huff,
University of Southern Mississippi,
United States
Naveen Kashyap,
Indian Institute of Technology
Guwahati, India

*Correspondence:

Sara Cadavid
sara.cadavid@urosario.edu.co

†ORCID:

Sara Cadavid
orcid.org/0000-0002-2606-3742
Maria Soledad Beato
orcid.org/0000-0001-6735-8398
Mar Suarez
orcid.org/0000-0002-2393-2228
Pedro B. Albuquerque
orcid.org/0000-0002-5874-4497

Specialty section:

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

Received: 26 March 2021

Accepted: 16 August 2021

Published: 13 September 2021

Citation:

Cadavid S, Beato MS, Suarez M
and Albuquerque PB (2021) Feelings
of Contrast at Test Reduce False
Memory in the
Deese/Roediger-McDermott
Paradigm.
Front. Psychol. 12:686390.
doi: 10.3389/fpsyg.2021.686390

False memories in the Deese/Roediger-McDermott (DRM) paradigm are explained in terms of the interplay between error-inflating and error-editing (e.g., monitoring) mechanisms. In this study, we focused on disqualifying monitoring, a decision process that helps to reject false memories through the recollection of collateral information (i.e., recall-to-reject strategies). Participants engage in recall-to-reject strategies using one or two metacognitive processes: (1) applying the logic of mutual exclusivity or (2) experiencing feelings of contrast between studied items and unstudied lures. We aimed to provide, for the first time in the DRM literature, evidence favorable to the existence of a recall-to-reject strategy based on the experience of feelings of contrast. One hundred and forty participants studied six-word DRM lists (e.g., spy, hell, fist, fight, abduction, mortal), simultaneously associated with three critical lures (e.g., WAR, BAD, FEAR). Lists differed in their ease to identify their critical lures (extremely low-BAS lists vs. high-BAS lists). At recognition test, participants saw either one or the three critical lures of the lists. Participants in the three-critical-lure condition were expected to increase their monitoring, as they would experience stronger feelings of contrast than the participants in the one-critical-lure condition. Results supported our hypothesis, showing lower false recognition in the three-critical-lure condition than in the one-critical-lure condition. Critically, in the three-critical-lure condition, participants reduced even more false memory when they could also resort to another monitoring strategy (i.e., identify-to-reject). These findings suggest that, in the DRM context, disqualifying monitoring could be guided by experiencing feelings of contrast between different types of words.

Keywords: false memories, false recognition, DRM paradigm, disqualifying monitoring, memory error-editing processes, multiple critical lures per list, backward associative strength

INTRODUCTION

During the last decades, memory researchers have intensely explored the underlying mechanisms of memory distortions, and have shown a particular interest in false memories (Gallo, 2006, 2010). One of the most widely employed procedures to induce false memories in controlled settings is the Deese/Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger and McDermott, 1995). In this paradigm, participants study lists of words (e.g., sour, candy, sugar, etc.), all of them associated with a non-presented critical lure (e.g., SWEET). In a subsequent memory task, participants often claim to recall or recognize the critical lure (false memories) along with the studied items

(true memories) (Roediger and McDermott, 1995). Numerous experimental manipulations have revealed that falsely remembered critical lures present highly compelling memorial evidence of the occurrence of the event (e.g., Beato et al., 2013; Boldini et al., 2013; Thakral et al., 2019; Brainerd et al., 2020; Howe and Akhtar, 2020; H. Liu et al., 2020; Beato and Arndt, 2021; Huff et al., 2021; Z. Liu et al., 2021; Wang et al., 2021).

In order to explain false memories, the two main theories are the fuzzy-trace theory or FTT (e.g., Brainerd and Reyna, 1990; Reyna and Brainerd, 1995; Brainerd et al., 2008) and the activation-monitoring framework or AMF (e.g., Roediger et al., 2001a,b). Despite differences between FTT and AMF, both agree to propose the interplay of two types of processes: error-inflating and error-editing.¹ These processes would work together to increase true memories, but they would operate in opposite directions in false memory (Arndt and Gould, 2006). Thus, whereas error-inflating processes would increase the likelihood to produce false memories, error-editing processes would reduce it.

The main aim of this research was to study error-editing processes in associative false memories and, in particular, the monitoring process, which has been identified as key to reduce false memories (e.g., Roediger et al., 2001b; Gallo, 2004; Gallo et al., 2006; Gallo and Lampinen, 2016; Coane et al., 2020; Huff et al., 2020). The monitoring process can be generally described as a decision process that helps participants allocate the source of mentally activated information, eventually reducing false memory (e.g., Gray, 2016; Roediger et al., 2001b; cf. Jou et al., 2018, who proposed that the ability to establish an appropriate decision criterion to monitor is based on what the observer monitors against). This monitoring process can be classified into diagnostic and disqualifying monitoring (e.g., Gallo, 2004, 2006; Gallo et al., 2006; Gallo and Lampinen, 2016; Nieznański et al., 2018; Moore et al., 2020). This division is based on the decisional processes around the avoidance of false memory.

On the one hand, diagnostic monitoring relies on the expectations generated around the decision making and happens when the critical lure is rejected due to an absence of recollection. In those cases, the dubious event (i.e., the critical lure) is rejected following a reasoning such as “if I had studied that item (e.g., my favorite fruit), I would recall it; as I do not remember it, it must not have been presented” (Gallo and Lampinen, 2016). On the other hand, disqualifying monitoring involves deciding whether the questionable event (i.e., the critical lure) was studied or not is made based on the recollection of collateral evidence. In these cases, certain information is recalled, and the recollection of that memory eliminates the possible occurrence of the questionable event. The recollection of collateral information to reject a dubious event is called “recall-to-reject” (e.g., Gallo, 2004; Gallo and Lampinen, 2016; Moore et al., 2020). Research has suggested that recall-to-reject can work through two different metacognitive processes: (1) participants apply the logic of mutual exclusivity, or (2) participants experience a feeling of contrast between studied items and unstudied critical lures (Gallo and Lampinen, 2016).

Examples of how recall-to-reject can occur following a logic of mutual exclusivity come from studies in which DRM lists are very short (e.g., three items). In that case, participants could reject the critical lures at test by remembering all the studied words of each list (i.e., exhaustive recall/recognition), which leads to extremely low false recognition rates (Gallo, 2004). Another example of the use of the logic of mutual exclusivity to avoid false memories is the finding that highly identifiable critical lures are more likely to be rejected (Carneiro et al., 2009, 2012). In this case, participants would apply a particular type of logic of mutual exclusivity that has been called “identify-to-reject,” and that would follow such a reasoning: “I did not encode A because, first, I remember to notice that A was the theme of the list, and second, I realized that A was not presented”. In addition, as noted above, participants can also engage in a type of recall-to-reject based on a feeling of contrast between the studied items and the unstudied critical lures (Gallo and Lampinen, 2016; Moore et al., 2020). In these cases, a sort of automatic attributional process might be intervening to reject the false memory, which would follow a reasoning like this: “This word (i.e., A, the critical lure) seems familiar to me, but I probably have this feeling just because another related word was actually presented (i.e., B, a studied item), so I will reject A” (Brainerd et al., 2003; Lampinen et al., 2005). Evidence for this strategy came from a study in which participants rated words for pleasantness and then completed an old/new recognition test in which they also had to think out loud and say whatever came to their minds during the retrieval process (Lampinen et al., 2005). This study found that participants sometimes noticed differences between the recollection experience of studied and unstudied items (“Cup—is new. I don’t remember seeing cup, but I remember seeing mug”). These results suggest that participants sometimes experience a feeling of contrast between items when performing a recognition test. In this context, it is worth noting that, even though the theoretical explanation of the feeling-of-contrast strategy is strongly related to the DRM associative illusion, to date, there is no empirical evidence that supports the idea that participants engage in this type of monitoring process in the DRM paradigm. To fill this gap, we examined whether, in the DRM paradigm, participants apply a recall-to-reject strategy based on feelings of contrast between items. As far as we know, ours is the first attempt in the literature to tackle this particular question directly.

To test the existence of the feeling-of-contrast strategy, we need a procedure that precludes the possibility to apply the logic of mutual exclusivity and allows the feeling-of-contrast strategy. As mentioned above, there are two main examples of how recall-to-reject can occur following a logic of mutual exclusivity: exhaustive recall/recognition and identify-to-reject strategies. One could argue that, since DRM lists are usually long, participants cannot resort to exhaustive recall/recognition because it is tough to remember all the studied items. However, this argument is not sufficient to affirm that participants are not engaging a mutual exclusivity logic as we also need to consider the possibility that participants apply an identify-to-reject strategy. In this regard, it should be noted that a typical DRM list includes one critical lure and words with high backward

¹ Other theoretical explanations of false memories, such as global matching models (Arndt and Hirshman, 1998), also agree on proposing these two types of processes.

associative strength (BAS, the association from studied items to the critical lure). When using DRM lists with these characteristics (i.e., high BAS and one critical lure), participants are prone to engage in an identify-to-reject strategy, a subtype of the logic of mutual exclusivity. Two mechanisms could trigger identify-to-reject in these lists. First, when there is only one critical lure per list, participants could identify this word as the theme of the list, they could be aware of the absence of this word in the study list, and, therefore, they might reject it at test (i.e., identify-to-reject strategy). Instead, we think that if the DRM lists included multiples critical lures, a condition used in the present study, it would be more difficult to engage in an identify-to-reject strategy. Second, in previous studies, a positive correlation has been found between BAS and identifiability indexes of the critical lures (e.g., Beato and Cadavid, 2016). That is, critical lures were more easily identifiable as the theme in high-BAS lists than in low-BAS lists. Therefore, participants are less likely to engage in an identify-to-reject strategy in DRM lists with lower backward associative strength.²

To prevent the use of mutual exclusivity logic (both via exhaustive recognition and identify-to-reject strategies) and analyze whether experiencing feelings of contrast could guide error-editing processes in DRM studies, we used DRM lists with multiple critical lures and extremely low levels of backward associative strength. Specifically, we manipulated two independent variables in this study: the number of critical lures per DRM list presented at test (one vs. three) and BAS (high vs. low). Including three actual critical lures per list at the recognition test would increase the likelihood that participants engage in a feeling-of-contrast strategy (i.e., more critical lures would increase the chances that participants have a feeling of contrast). Also, it seems likely that including all the three critical lures of our lists in the recognition test could diminish the probability of engaging in an identify-to-reject strategy (i.e., participants would not be able to explicitly identify the three critical lures of each list to reject them). Furthermore, to make it even less likely that participants engage in an identify-to-reject strategy, we included lists with the minimum possible BAS levels. These extremely low-BAS levels served as a proxy for low-identifiability levels. Therefore, lists with three critical lures and extremely low BAS would constitute the experimental condition in which we prevented the use of mutual exclusivity logic and, instead, foster a feeling-of-contrast strategy. Hereunder, these ideas are explained in more detail.

With respect to the number of critical lures per list, previous studies have reported that two- or three-critical-lure DRM lists produce robust false memories (e.g., Beato and Díez, 2011; Beato et al., 2012; Cadavid et al., 2012; Beato and Arndt, 2014, 2017;

Beato and Cadavid, 2016; Cadavid and Beato, 2016, 2017; Arndt and Beato, 2017; Pitarque et al., 2018). In these previous studies, it is unlikely that participants would have used the logic of mutual exclusivity to reject the critical lures because they would not be able to (1) remember all the studied items or (2) explicitly identify the two or three themes of the list (i.e., critical lures) to reject them. In contrast, we expected that participants would engage more often in a feeling-of-contrast strategy following a reasoning like this: “This word (i.e., A, one of the critical lures) seems familiar to me, but I probably have this feeling just because another related word was actually presented (e.g., studied word B or critical lure C), so I will reject A”. In other words, the presence of more critical lures per list at the recognition test would increase the chance of feeling the contrast between a lure and the rest of the words (i.e., studied and other lures from the same list). Hence, just as in other manipulations that facilitate error-editing processes, overall false recognition rates were expected to be lower in the three-critical-lure condition than in the one-critical-lure condition.

Regarding the BAS manipulation, we used DRM lists from a previous normative study (Cadavid and Beato, 2017) where the low-BAS lists had the lowest possible BAS levels. In this case, as referred above, we expected that extremely low levels of associative strength would serve as a proxy for low-identifiability levels. That is, including extremely low-BAS lists allows us to virtually eliminate the possibility that participants engage in an identify-to-reject strategy.

Hence, as previously mentioned, the low-BAS/three-critical-lure condition would be the condition where it seems less likely that monitoring can occur via mutual exclusivity processes as (1) it is not likely that participants remember all the studied words (i.e., exhaustive recognition), (2) the lists had extremely low-BAS levels, which hinders the engagement of an identify-to-reject strategy, and (3) more critical lures at test increase the likelihood of experiencing feelings of contrast between the actually studied words and the critical lures.

We predicted that false recognition would be lower in the three-critical-lure condition than in the one-critical-lure condition, showing evidence toward the presence of feelings of contrast in the DRM paradigm. Furthermore, we were interested in analyzing two specific comparisons. First, we compared false recognition in the one-critical-lure vs. the three-critical-lure conditions for the low-BAS lists, as the extremely low-BAS levels prevent the use of an identify-to-reject strategy. We anticipated a lower false recognition rate in the low-BAS/three-critical-lure condition than in the low-BAS/one-critical-lure condition. This finding would show that, in the absence of an identify-to-reject strategy (low-BAS lists), participants whose feelings of contrast were not triggered (one-critical-lure condition) would show higher false memory than the participants who experienced feelings of contrast (three-critical-lure condition).

Our second specific comparison referred to BAS levels in the three-critical-lure condition. It is worth reminding that, in this study, BAS levels were used as a proxy for identifiability levels. We expected higher false recognition levels in the extremely low-BAS lists than in the high-BAS lists when tested with all its three critical lures. From our monitoring process perspective, as

²The backward associative strength or BAS has been identified as a reliable predictor of false memory and is one of the most commonly cited factors that facilitates error-inflating processes, as research has shown that false recall and false recognition rates were higher in high-BAS than in low-BAS lists (e.g., McEvoy et al., 1999; Roediger et al., 2001b; Gallo and Roediger, 2003; Arndt and Gould, 2006; Beato and Arndt, 2017). However, some studies have also reported that BAS did not affect false recognition rates (e.g., Cadavid et al., 2012; Brainerd et al., 2020). Although these results are of interest, the aim of the present study is not to shed light on the role of BAS in false memory, as the extremely low-BAS levels employed here are only used as a proxy for low-identifiability levels.

it was previously mentioned, in the experimental condition in which low-BAS lists are studied and three critical lures per list are included at test, participants could engage in a feeling-of-contrast strategy, but not in monitoring processes following the logic of mutual exclusivity (i.e., identify-to-reject strategy). However, when participants included in the three-critical-lure condition study high-BAS lists, it might be the case that they could use both feelings of contrast and an identify-to-reject strategy to monitor their memory (i.e., high-BAS lists could be used as a proxy for high-identifiability levels). Therefore, when using high-BAS lists and presenting three critical lures per list at test, participants could engage in more than one type of editing processes. If we assumed that there could be an additive effect on their ability to reduce false memories, this would lead to a greater reduction of false recognition in high-BAS lists than in low-BAS lists when presenting three critical lures per list at test.

MATERIALS AND METHODS

Participants

A total of 140 undergraduate students, who were native Spanish speakers, voluntarily participated in this experiment (69.29% woman; $M_{age} = 21.35$, $SD = 3.91$). A power analysis using G*Power 3.1 (Faul et al., 2007), with a power of 0.80 and an alpha of 0.05, showed that a total of 128 participants were enough to detect a medium effect size ($f = 0.25$) in the 2 (Number of critical lures per list at test) $\times$ 2 (BAS) interaction of our interest. We increased the sample size from 32 to 35 participants per group for a total of 140 participants. All participants signed an informed consent form and received course credit. The study was approved by the Bioethics Committee of the University of Salamanca.

Design

The experiment followed a 2×2 between-subjects design. The two independent variables were Number of critical lures per list at test (one, three) and BAS (high, low).

Materials

We used 32 six-word DRM lists from Cadavid and Beato (2017) normative study (see Table 1). Specifically, lists were constructed from Spanish free-association norms (Fernandez et al., 2011). Lists were built to ensure that all the three critical lures (e.g., WATER, BOAT, and SEA) were produced by the same study items (*marine, lifejacket, dyke, castaway, island, and exportation*) in free association. That is, the six study items were simultaneously related via backward associative strength or BAS to each of the three critical lures.

The BAS values for each critical lure (hereafter, *critical lure BAS*) were computed as the mean of the associative strengths between each of the six associated words and the particular critical lure, just as in previous research (Robinson and Roediger, 1997). For its part, the BAS values of each list (hereafter, *BAS list strength*) were calculated by averaging the BAS values for the three critical lures within each list. For example, a low-BAS list included the critical lures WAR (BAS = 0.013), BAD (BAS = 0.010), and FEAR (BAS = 0.010), each of which

had backward associations to the study items *spy, hell, fist, fight, abduction, and mortal*. An example of a high-BAS list included the critical lures FOREST (BAS = 0.070), FIELD (BAS = 0.068), and HILL (BAS = 0.073), all of them associated with the study items *excursion, mushroom, cottage, deer, green, and meadow*.

Furthermore, ten DRM lists were selected from Alonso et al.'s (2004) normative study, from which the distractors were extracted. We used this normative study to select distractors from DRM lists that did not include, or were related to, our study items or critical lures. Lists included a critical lure (e.g., TELEPHONE) and fifteen associated words (e.g., *call, home, communication, mobile, dream, numbers, speak, invoice, conversation, guide, distance, cable, noise, chat, and prefix*). Specifically, we selected the lists of critical lures BOX, GLASSES, COMB, TRAVEL, KEY, FORK, LAMP, TELEPHONE, COW, and PENGUIN. Unrelated critical-distractors were the critical lure of the lists, whereas unrelated distractors were selected from its associates.

For the present study, we selected 16 high-BAS lists ($M_{BAS} = 0.44$, $SD = 0.08$, $range_{per\ lure} = 0.21-0.62$) and 16 low-BAS lists ($M_{BAS} = 0.13$, $SD = 0.04$, $range_{per\ lure} = 0.06-0.21$). It is important to note that the low-BAS lists included extremely low-BAS values, values never used before in the DRM paradigm, covering the lowest end of the entire spectrum of associative strength, as they included the minimum associative strength theoretically possible (Cadavid and Beato, 2017). We confirmed that the associative strengths of high- and low-BAS lists differed significantly, Welch's $t_{per\ lure}(70.24) = 22.38$, $p < 0.001$, $d = 4.57$. Lists were audio-recorded with a male voice, and during the study phase, stimuli were presented auditorily by using speakers.

The recognition test was administered as a pen and paper task. This memory test included a total of 192 words (studied words, critical lures, unrelated critical-distractors, and unrelated distractors) that varied between the experimental conditions. For the one-critical-lure condition, the recognition test consisted of 96 studied words, 16 critical lures (i.e., one critical lure per study list) and 80 distractors (10 unrelated critical-distractors and 70 unrelated distractors). We ensured that all the three critical lures of each list were tested in the one-critical-lure condition. For the three-critical-lure condition, the recognition test included 96 studied words, 48 critical lures (i.e., three critical lures per study list), and 48 distractors (8 unrelated critical-distractors and 40 unrelated distractors). Thus, the recognition memory test included the same number of studied and unstudied words (96 studied and 96 unstudied words) in both experimental conditions.

The items of the recognition test were pseudorandomized according to criteria proposed in previous research (Gallo and Roediger, 2002; Graham, 2007). Concretely, we made sure that two or more items separated words from the same list. Besides, we assured critical lures were separated from each other for at least two items. There were six versions of the recognition test, which was included at the end of a booklet that also contained sixteen pages of unsolved arithmetic operations series that were solved in-between the study of the lists.

TABLE 1 | Thirty-Two Six-Word Lists with Three Critical Lures (Approximated English Translation), Backward Associative Strength (BAS) Condition, and Critical Lure BAS.

CRITICAL LURES: Associated words (Approximated English translation)	BAS condition	BAS Lure 1	BAS Lure 2	BAS Lure 3
GUERRA, MALO, MIEDO: espía, infierno, puño, pelea, rapto, mortal (WAR, BAD, FEAR: spy, hell, fist, fight, abduction, mortal)	Low	0.013	0.010	0.010
EXAMEN, FÁCIL, TRABAJO: ejercicio, introducción, aplicación, exigencia, memoria, importante (EXAM, EASY, WORK: exercise, introduction, application, demand, memory, important)	Low	0.015	0.012	0.013
ANGUSTIA, LLORAR, PENA: llanto, afligido, deprimido, desazón, alivio, victimismo (ANGUISH, TO CRY, SORROW: crying, mournful, depressed, unease, relief, sense of victimization)	Low	0.012	0.012	0.017
HONOR, NOBLEZA, PERSONA: lealtad, nobiliario, integridad, orgullo, solemnidad, duque (HONOR, NOBILITY, PERSON: loyalty, nobility, integrity, pride, solemnity, duke)	Low	0.018	0.013	0.017
DOLOR, MUERTE, TRISTEZA: odio, hambre, inanición, morir, huérfano, consolado (PAIN, DEATH, SADNESS: hatred, hunger, starvation, to die, orphan, comforted)	Low	0.015	0.021	0.015
LIMPIEZA, SUCIEDAD, SUCIO: limpiar, gérmenes, basura, bastoncillo, fregadero, servilleta (CLEANLINESS, DIRT, DIRTY: to clean, germs, trash, cotton swab, sink, napkin)	Low	0.013	0.021	0.018
BOSQUE, CAMPO, MONTE: natural, conejo, valle, liebre, roble, refugio (FOREST, FIELD, HILL: natural, rabbit, valley, hare, oak, refuge)	Low	0.019	0.020	0.017
ALEGRÍA, CONTENTO, SONRISA: carcajada, jubiloso, animado, agrado, agradecer, esperanzado (JOY, PLEASED, SMILE: guffaw, jubilant, cheerful, kindness, to thank, hopeful)	Low	0.028	0.019	0.018
BEBÉ, CARIÑO, NIÑO: dulzura, hijo, tierno, protegido, acurrucarse, peluche (BABY, FONDNESS, CHILD: gentleness, son, tender, protected, to cuddle, teddy)	Low	0.018	0.025	0.027
CAMISA, PANTALÓN, ROPA: chaqueta, jersey, suéter, roto, rayas, arrugado (SHIRT, TROUSERS, CLOTHING: jacket, jersey, sweater, torn, stripes, wrinkled)	Low	0.024	0.024	0.028
DIVERSIÓN, FIESTA, NOCHE: club, marcha, droga, alcohol, concierto, cantar (FUN, PARTY, NIGHT: club, going out, drug, alcohol, concert, to sing)	Low	0.019	0.031	0.028
JUEZ, JUICIO, LEY: juramento, enmienda, justo, defensor, penal, defendido (JUDGE, TRIAL, LAW: oath, amendment, fair, defender, criminal, defendant)	Low	0.028	0.026	0.028
ENFERMO, HOSPITAL, MÉDICO: medicina, salud, dolencia, visita, virus, interno (SICK, HOSPITAL, DOCTOR: medicine, health, ailment, visit, virus, internal)	Low	0.028	0.023	0.033
INTELIGENCIA, LISTO, SABIO: erudición, genio, inculto, tenacidad, científico, elocuencia (INTELLIGENCE, SMART, WISE: erudition, genius, uncultured, tenacity, scientific, eloquence)	Low	0.023	0.032	0.032
CURA, IGLESIA, RELIGIÓN: papa, doctrina, blasfemia, reverencia, místico, súplica (CLERGYMAN, CHURCH, RELIGION: pope, doctrine, blasphemy, reverence, mystic, plea)	Low	0.035	0.027	0.030
CLASE, COLEGIO, ESCUELA: primaria, lección, aprender, academia, punzón, promoción (CLASS, SCHOOL, SCHOOL: elementary school, lesson, to learn, academy, punch, class)	Low	0.033	0.032	0.033
DESASTRE, HORROR, MUERTE: masacre, fatalidad, catástrofe, terremoto, tragedia, barbarie (DISASTER, HORROR, DEATH: massacre, fatality, catastrophe, earthquake, tragedy, brutality)	High	0.055	0.067	0.035
FOLIO, HOJA, PAPEL: doblar, margen, grapa, copia, clip, arrugado (FOLIO, SHEET, PAPER: to fold, margin, staple, copy, clip, crumpled)	High	0.053	0.045	0.062
DIOS, IGLESIA, MISA: mandamiento, oración, bendecir, devoción, comunión, gloria (GOD, CHURCH, MASS: commandment, prayer, to bless, devotion, communion, glory)	High	0.043	0.063	0.055
LÁGRIMA, LLORAR, TRISTEZA: lacrimal, llanto, despedida, emoción, infeliz, llover (TEAR, TO CRY, SADNESS: lachrymal, crying, farewell, emotion, unhappy, to rain)	High	0.061	0.055	0.068
ABRIGO, CHAQUETA, ROPA: cuero, gabardina, botón, colgar, chaleco, corchetes (COAT, JACKET, CLOTHING: leather, gabardine, button, to hang, vest, snap fastener)	High	0.066	0.074	0.049
ALCOHOL, BEBER, BEBIDA: ron, cerveza, tomar, botella, bar, copa (ALCOHOL, TO DRINK, DRINK: rum, beer, to drink, bottle, bar, drink)	High	0.066	0.065	0.072
BOSQUE, CAMPO, MONTE: excursión, seta, cabaña, ciervo, verde, pradera (FOREST, FIELD, HILL: excursion, mushroom, cottage, deer, green, meadow)	High	0.070	0.068	0.073
AMOR, CARIÑO, MADRE: ternura, dulzura, hijo, apreciación, consuelo, comprensión (LOVE, FONDNESS, MOTHER: tenderness, gentleness, son, fondness, comfort, comprehension)	High	0.092	0.069	0.065
CINE, PELÍCULA, TEATRO: actor, actriz, estrena, actuar, comedia, reparto (CINEMA, FILM, THEATER: actor, actress, premiere, to act, comedy, cast)	High	0.069	0.086	0.073
INTELIGENCIA, LISTO, SABIO: astucia, sabiduría, ingenio, erudición, genio, inculto (INTELLIGENCE, SMART, WISE: astuteness, wisdom, ingenuity, erudition, genius, uncultured)	High	0.077	0.080	0.075
CÁRCEL, LADRÓN, POLICÍA: detención, robo, mazmorra, delito, persecutoria, vigilancia (JAIL, THIEF, POLICE: detention, robbery, dungeon, crime, relative to persecution, vigilance)	High	0.088	0.060	0.089
FUMAR, HUMO, TABACO: pipa, puro, cenicero, pulmones, mechero, habano (TO SMOKE, SMOKE, TOBACCO: pipe, cigar, ashtray, lungs, lighter, havana cigar)	High	0.099	0.073	0.072

(Continued)

TABLE 1 | (Continued)

CRITICAL LURES: Associated words (Approximated English translation)	BAS condition	BAS Lure 1	BAS Lure 2	BAS Lure 3
AGUA, BARCO, MAR: marina, salvavidas, dique, naufrago, isla, exportación (WATER, BOAT, SEA: marine, lifejacket, dyke, castaway, island, exportation)	High	0.072	0.088	0.090
MÚSICA, RUIDO, SONIDO: acústica, tambor, tono, cascabel, sonar, grillos (MUSIC, NOISE, SOUND: acoustic, drum, tone, rattle, to sound, crickets)	High	0.087	0.090	0.075
ENFERMEDAD, HOSPITAL, MÉDICO: clínica, sanidad, paciente, sarampión, dolencia, curar (DISEASE, HOSPITAL, DOCTOR: clinic, health service, patient, measles, ailment, to heal)	High	0.093	0.079	0.083
DINERO, SUELDO, TRABAJO: empleo, jornal, aumento, ganancias, jefe, mensual (MONEY, WAGE, WORK: job, day wage, raise, profits, boss, monthly)	High	0.085	0.078	0.104

Lists Appear in Increasing Order of BAS per List.

Procedure

Participants were randomly assigned to one of the four experimental conditions defined by two between-subjects variables: Number of critical lures per list at test (one, three) and BAS (high, low). This experiment was run in about 60-min group sessions. Before starting the study phase, participants were presented with a practice list. The study phase instructions indicated that words should be remembered for a later memory test. No mention was made about the associative nature of the study lists.

Participants studied 16 DRM lists randomly presented. The items within each list were presented in decreasing order of BAS values, at a rate of one word every 2000 ms. Lists were alternated with 20 s series of simple arithmetic operations that had to be solved in a booklet. The self-paced recognition test was administrated in the same booklet as the arithmetic operations. Participants had to judge 192 words and decide whether each word was presented in the study phase or not by circling “YES” or “NO” on the response sheet. As mentioned above, the number of critical lures included in the memory test varied between the one- and the three-critical-lure conditions (16 vs. 48, respectively).

RESULTS

Across all analyses, Greenhouse-Geisser correction of degrees of freedom was applied where appropriate in the repeated measures ANOVAs, the alpha level was set at 0.05, and effect sizes are reported with Cohen’s d and omega squared (ω^2). All analyses were performed using JASP Team (2020).

True Recognition

A 2 (Number of critical lures per list at test: one, three) $\times$ 2 (BAS: high, low) between-subjects ANOVA was ran on true recognition rates. No significant main effects, $F_{\text{number of lures}}(1, 136) = 0.04$, $p = 0.848$, $\omega^2 < 0.001$, $F_{\text{BAS}}(1, 136) = 0.62$, $p = 0.432$, $\omega^2 < 0.001$, nor interaction were found, $F(1, 136) = 0.35$, $p = 0.557$, $\omega^2 < 0.001$ (see Table 2 for descriptive statistics).

False Recognition Effect

A one-way repeated measures ANOVA (Type of word: studied, critical lure, unrelated critical-distractor, and unrelated distractor) was conducted, $F(2.42, 336.06) = 717.37$, $p < 0.001$,

$\omega^2 = 0.74$ (see Table 2 for descriptive statistics).³ We computed six comparisons and applied Bonferroni correction. Hence, the new alpha was set at 0.008. True recognition ($M = 63.47$, $SD = 12.40$) was significantly higher than false recognition to critical lures ($M = 36.62$, $SD = 17.10$), $t(139) = 16.79$, $p < 0.001$, Cohen’s $d = 1.42$; than false alarms to unrelated critical-distractors ($M = 9.74$, $SD = 13.11$), $t(139) = 34.20$, $p < 0.001$, Cohen’s $d = 2.89$; and also higher than false alarms to unrelated distractors ($M = 9.09$, $SD = 9.75$), $t(139) = 39.58$, $p < 0.001$, Cohen’s $d = 3.35$. Furthermore, false recognition was significantly higher than false alarms to unrelated critical-distractors, $t(139) = 18.96$, $p < 0.001$, Cohen’s $d = 1.60$; and unrelated distractors, $t(139) = 21.69$, $p < 0.001$, Cohen’s $d = 1.83$. This result confirmed that critical lures produced above-baseline levels of false recognition. No significant differences were found between unrelated critical-distractors and unrelated distractors, $t(139) = 0.78$, $p = 0.435$, Cohen’s $d = 0.07$.

False Recognition and Critical Lures’ Position at Test

Besides using extremely low-BAS lists, the ease of engaging a feeling-of-contrast strategy during the recognition test was manipulated by including all the three critical lures of each list or including just one of them. According to the logic of the feeling of contrast, participants would experience more feelings of contrast when they encounter several critical lures. In order to check whether the feeling-of-contrast strategy was actually favored in the three-critical-lure condition, we analyzed false recognition of any of the three critical lures in the first, second or third position at test. Specifically, in the three-critical-lure condition, 35.24% ($SD = 12.39$) of the total false recognition occurred in the first critical lure of the lists, 35.91% ($SD = 11.67$) appeared in the second critical lure, and, finally, 28.84% ($SD = 10.72$) of the total false recognition happened in the third critical lure. We conducted a repeated measures ANOVA (Position at test of the

³No correction was made on recognition rates because false alarms to unrelated critical-distractors did not show significant differences between the two experimental conditions, one critical lure ($M = 10.02$, $SD = 12.40$) and three critical lures per list at test ($M = 9.46$, $SD = 13.86$), $t(138) = 0.25$, $p = 0.804$, Cohen’s $d = 0.04$. Furthermore, false alarms to unrelated distractors did not show significant differences between the two experimental conditions either, $t(138) = 1.07$, $p = 0.287$, Cohen’s $d = 0.18$ (one-critical-lure condition, $M = 9.97$, $SD = 9.62$; three-critical-lure condition, $M = 8.21$, $SD = 9.87$).

TABLE 2 | Mean Percentage (SD) of True and False Recognition and False Alarms to Unrelated Critical-Distractors and Unrelated Distractors as a function of Number of Critical Lures per List at Test (one vs. three) and BAS (high vs. low).

	One lure		Three lures	
	High BAS	Low BAS	High BAS	Low BAS
True recognition	63.75 (14.83)	63.80 (9.86)	61.73 (12.04)	64.61 (12.73)
False recognition	41.25 (19.19)	42.14 (15.48)	26.19 (14.92)	36.90 (14.13)
False alarms to unrelated critical-distractors	8.86 (12.31)	11.18 (12.55)	6.79 (10.65)	12.14 (16.18)
False alarms to unrelated distractors	6.80 (6.36)	13.15 (11.26)	7.71 (10.58)	8.71 (9.22)

critical lures of each list: first, second, and third) on the false recognition rates, $F(2, 138) = 5.28$, $p = 0.007$, $\omega^2 = 0.06$. We computed three comparisons and applied Bonferroni correction. Hence, the new alpha was set at 0.016. False recognition for the first and second critical lures did not show significant differences, $t(69) = 0.26$, $p = 0.795$, Cohen's $d = 0.03$. However, false recognition for the third critical lure was significantly lower than false recognition for the first, $t(69) = 2.68$, $p = 0.009$, Cohen's $d = 0.32$, and the second critical lure, $t(69) = 3.17$, $p = 0.002$, Cohen's $d = 0.38$. These results supported the idea that the feeling-of-contrast strategy was favored in the three-critical-lure condition.

Error-Editing Processes: The Effect of the Number of Critical Lures per List at Test and BAS

The mean percentage of false recognition as a function of the number of critical lures per list at test and BAS are presented in **Table 2**. In both one-critical-lure condition and three-critical-lure condition, false recognition was calculated as the mean of the false recognition of all the critical words included in the recognition test.

We used a 2 (Number of critical lures per list at test: one, three) $\times$ 2 (BAS: high, low) between-subjects ANOVA to examine the effects of the number of critical lures per list at test and BAS on the error-editing processes in false recognition. Results showed a significant main effect of BAS, $F(1, 136) = 4.58$, $p = 0.034$, $\omega^2 = 0.02$, indicating that false recognition rates were significantly higher in low-BAS lists ($M = 39.52$, $SD = 14.95$) than in high-BAS lists ($M = 33.72$, $SD = 18.67$). As expected, the number of critical lures per list at test also showed a significant main effect, $F(1, 136) = 14.00$, $p < 0.001$, $\omega^2 = 0.09$. Specifically, false recognition rates were lower in the three-critical-lure condition ($M = 31.55$, $SD = 15.40$) than in the one-critical-lure condition ($M = 41.70$, $SD = 17.31$). The interaction was not significant, $F(1, 136) = 3.28$, $p = 0.072$, $\omega^2 = 0.02$.

Since our goal was to analyze whether experiencing feelings of contrast could guide error-editing processes in DRM studies, we needed to eliminate the possibility that participants apply the mutual exclusivity logic (both via exhaustive recognition and identify-to-reject strategies). We assumed that the identify-to-reject strategy is mitigated with extremely low-BAS lists. However, no experimental manipulation was made to eliminate the possibility that participants engage in an exhaustive recognition strategy (i.e., recognize all the six study items).

Therefore, we ran an additional conditioned analysis removing, for each participant, false recognition for lists that had perfect correct recognition for list items. This approach provided a better estimate of the feelings of contrast effects on false recognition, as we removed the lists in which mutual exclusivity could be occurring via exhaustive recognition. Participants in the low-BAS condition had 13.57% of lists removed ($M_{one\ critical\ lure} = 12.68$, $SD = 13.93$; $M_{three\ critical\ lures} = 14.46$, $SD = 10.59$), whereas 16.96% of the lists were eliminated in the high-BAS condition ($M_{one\ critical\ lure} = 18.04$, $SD = 13.96$; $M_{three\ critical\ lures} = 15.89$, $SD = 12.62$). Turning to the number of critical lures per list at test, in the one-critical-lure condition, 15.36% of the lists were removed, and 15.18% of the lists for the three-critical-lure condition were not included in the conditionalized analysis.

Just as in the previous ANOVA, in this conditioned analysis the main effect of BAS showed that false recognition was higher in low- ($M = 38.42$, $SD = 14.76$) than in high-BAS lists ($M = 31.25$, $SD = 18.86$), $F(1, 136) = 6.65$, $p = 0.01$, $\omega^2 = 0.04$. Again, the main effect of the number of critical lures per list at test was also significant, $F(1, 136) = 8.09$, $p = 0.005$, $\omega^2 = 0.05$, showing that false recognition was lower in the three-critical-lure condition ($M = 30.88$, $SD = 15.95$) than in the one-critical-lure condition ($M = 38.80$, $SD = 17.70$). The interaction was not significant, $F(1, 136) = 1.73$, $p = 0.191$, $\omega^2 = 0.005$ (see **Table 3**).

Taking into account our hypotheses stated in the introduction, we computed two planned comparisons on the conditioned false recognition rates (alpha was adjusted to 0.025). First, in the low-BAS condition, no differences were found between one and three critical lures, $t(68) = 1.21$, $p = 0.12$, Cohen's $d = 0.289$. Considering the limitations of the null hypothesis significance testing (Dienes, 2011), Bayesian analyses were run (H_0 = no differences between the means, H_1 = differences between the means). A Bayesian independent samples t -test indicated that the H_0 is 2.17 times more likely than the H_1 , which represent anecdotal evidence in favor of the H_0 . Second, in the three-critical-lure condition, false recognition was higher when participants could not resort to an identify-to-reject strategy (low-BAS lists) than when they could use such a strategy (high-BAS lists), $t(68) = 3.00$, $p = 0.004$, Cohen's $d = 0.718$. A Bayesian independent samples t -test indicated that the H_1 is 10.21 times more likely than the H_0 , which corresponds to moderate to strong evidence in favor of H_1 .

In sum, since false recognition was lower in the three-critical-lure condition than in the one-critical-lure condition, it seems that feelings of contrast could guide monitoring processes in the DRM paradigm. However, in the absence of an identify-to-reject

TABLE 3 | Mean Percentage (SD) of True and False Recognition After Removing Lists with 100% of True Recognition, as a Function of Number of Critical Lures per List at Test (one vs. three) and BAS (high vs. low).

	One lure		Three lures	
	High BAS	Low BAS	High BAS	Low BAS
True recognition	56.14 (12.08)	58.99 (7.50)	55.23 (10.66)	59.34 (11.62)
False recognition	37.04 (19.91)	40.56 (15.28)	25.46 (16.01)	36.30 (14.13)

strategy (low-BAS conditions), participants who experienced feelings of contrast (three-critical-lure condition) did not show lower levels of false memory than participants in the one-critical-lure condition. Instead, evidence was favorable to the existence of an additive effect of the two monitoring strategies examined in this study (i.e., identify-to-reject and feelings of contrast), as false recognition was significantly lower when both strategies were allowed (high-BAS/three-critical-lure condition).

DISCUSSION

The current study focused on disqualifying monitoring, a type of decision process that helps participants to reject false memories through the recollection of collateral information. The recollection of collateral information is achieved by the engagement of recall-to-reject strategies. Participants engage in recall-to-reject strategies employing one or two metacognitive processes: (1) applying the logic of mutual exclusivity, or (2) experiencing a feeling of contrast between studied items and unstudied critical lures (Gallo and Lampinen, 2016). As, to our knowledge, no previous studies have attempted to directly examine whether participants can apply a recall-to-reject strategy based on feelings of contrast in the DRM paradigm context, our goal was to fill this gap.

In the present study, we used a new experimental design where participants would find it difficult to use the logic of mutual exclusivity (neither exhaustive recall nor identify-to-reject strategy) to edit false memories, and instead, they could only experience feelings of contrast between words. Specifically, we employed DRM lists with three critical lures each of them (words not included in the study list) and extremely low levels of backward associative strength (the minimum possible association levels). In the recognition test, we manipulated the number of critical lures per list presented at test (only one vs. all the three critical lures per list).

On the one hand, regarding the number of critical lures per list, participants studied all the associates of the lists following the same instructions at the encoding phase. Importantly, at the recognition test, participants were presented with either only one or all the three critical lures of the lists. We expected that participants who were presented with three critical lures per list at test would increase their ability to edit false memories, as they would experience stronger feelings of contrast than the participants who were presented with only one critical lure per list at test. Participants in the three-critical-lure condition would be more prone to think something like: “This word (i.e., A, one of the critical lures) seems familiar to me, but I

probably have this feeling just because another related word was actually presented (e.g., critical lure B or studied word C), so I will reject A.” This type of reasoning, repeated all across the recognition test, would lead to lower false recognition levels in the three-critical-lure condition than in the one-critical-lure condition. In other words, when exposed to three critical lures per list at test, participants would have fewer false memories. The results supported our hypothesis, meaning that, in the DRM context, disqualifying monitoring could be guided by the experience of feelings of contrast between different types of words.

On the other hand, we manipulated the level of BAS, including extremely low-BAS lists, that is, lists with the lowest possible BAS level. We anticipated that such extremely low-BAS levels would make it difficult for participants to guess the critical lures. In fact, previous DRM studies with three-critical-lure lists have found a significant correlation between BAS and identifiability of the critical lures (Beato and Cadavid, 2016). Therefore, we expected that the exceptionally low-BAS levels would also make it difficult to resort to recall-to-reject strategies guided by the logic of mutual exclusivity (e.g., identify-to-reject). The results supported our hypotheses, showing that lists with extremely low-BAS levels produced higher false recognition rates than lists with high-BAS levels. As expected, this difference was specially important in the three-critical-lure condition, where participants were better at avoiding false recognition. Indeed, participants committed fewer mistakes in the three-critical-lure condition, and, within this condition, they were particularly efficient at reducing false memory in the high-BAS condition. These results are in favor of the possibility that, when combined together, the two monitoring strategies explored in this study (i.e., identify to reject and feelings of contrast) could trigger an additive effect to reduce false memory (high-BAS/three-critical-lure condition). This outcome needs to be explored further in future research.

The current findings help us to gain knowledge of the mechanisms that underlie memory accuracy. Memory editing is a complex phenomenon that comprises multiple sub-processes and different strategies. In this research, while the encoding instructions at the study phase were not manipulated, we did manipulate the BAS level of the lists and the number of critical lures per list available at the recognition test. We found that, when presented with more critical lures per list at test, participants are better at avoiding false memories than when they are presented just with one critical lure per list at test (48 vs. 16 critical lures, respectively). Therefore, we provided evidence that the amount of competing information available at test is determinant to trigger different memory editing mechanisms. Our data are

consistent with Gallo and collaborators' classification of the decisional processes that guide memory distortion avoidance (e.g., Gallo, 2004; Gallo et al., 2006; Gallo and Lampinen, 2016; Moore et al., 2020) and leave the door open to new questions. One possible future line of continuing this research would come from manipulations both at encoding and at test. For example, would different types of instructions trigger the sort of attributional process associated with experiencing feelings of contrast? Would these feelings always decrease false memory? Could explicitly drawing participants' attention to their subjective memory experience reduce false memory? One potential limitation of the current study is that participants were never directly queried regarding the monitoring of individual test items. Future studies with DRM lists with three critical lures could benefit greatly from adopting think-aloud protocols like those used by Lampinen et al. (2005). These and other questions remain to be explored in future research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Alonso, M. A., Fernandez, A., Díez, E., and Beato, M. S. (2004). Índices de producción de falso recuerdo y falso reconocimiento para 55 listas de palabras en castellano [False recall and recognition indexes for 55 lists of words in Spanish]. *Psicothema* 16, 357–362.
- Arndt, J., and Beato, M. S. (2017). The role of language proficiency in producing false memories. *J. Memory Lang.* 95, 146–158. doi: 10.1016/j.jml.2017.03.004
- Arndt, J., and Gould, C. (2006). An examination of two-process theories of false recognition. *Memory* 14, 814–833. doi: 10.1080/09658210600680749
- Arndt, J., and Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *J. Memory Lang.* 39, 371–391. doi: 10.1006/jmla.1998.2581
- Beato, M. S., and Arndt, J. (2014). False recognition production indexes in forward associative strength (FAS) lists with three critical words. *Psicothema* 26, 457–463. doi: 10.7334/psicothema2014.79
- Beato, M. S., and Arndt, J. (2017). The role of backward associative strength in false recognition of DRM lists with multiple critical words. *Psicothema* 29, 358–363. doi: 10.7334/psicothema2016.248
- Beato, M. S., and Arndt, J. (2021). The effect of language proficiency and associative strength on false memory. *Psychol. Res.* 2021:1449. doi: 10.1007/s00426-020-01449-3
- Beato, M. S., Boldini, A., and Cadavid, S. (2012). False memory and level of processing effect: An event-related potential study. *NeuroReport* 23, 804–808. doi: 10.1097/WNR.0b013e32835734de
- Beato, M. S., and Cadavid, S. (2016). Normative study of theme identifiability: Instructions with and without explanation of the false memory effect. *Behav. Res. Methods* 48, 1252–1265. doi: 10.3758/s13428-015-0652-6
- Beato, M. S., Cadavid, S., Pulido, R. F., and Pinho, M. S. (2013). No effect of stress on false recognition. *Psicothema* 25, 25–30. doi: 10.7334/psicothema2012.158
- Beato, M. S., and Díez, E. (2011). False recognition production indexes in Spanish for 60 DRM lists with three critical words. *Behav. Res. Methods* 43, 499–507. doi: 10.3758/s13428-010-0045-9
- Boldini, A., Beato, M. S., and Cadavid, S. (2013). Modality-match effect in false recognition: An event-related potential study. *NeuroReport* 24, 108–113. doi: 10.1097/WNR.0b013e32835c93e3
- Brainerd, C. J., Chang, M., and Bialer, D. M. (2020). From association to gist. *J. Exp. Psychol. Learn. Memory Cogn.* 46, 2106–2127. doi: 10.1037/xlm0000938

ETHICS STATEMENT

This study involving human participants was reviewed and approved by the Bioethics Committee of the University of Salamanca. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC and MSB conceived and planned the experiment, analyzed data, and wrote the original draft of the manuscript. MS and PBA reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by grant SA052G18, Junta de Castilla y León, by the University of Salamanca (Spain), and by Support scholarships for doctoral students (Becas de apoyo para estudiantes doctorales) from the Universidad del Rosario (Colombia).

- Brainerd, C. J., and Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Dev. Rev.* 10, 3–47. doi: 10.1016/0273-2297(90)90003-M
- Brainerd, C. J., Reyna, V. F., and Ceci, S. J. (2008). Developmental reversals in false memory: a review of data and theory. *Psychol. Bull.* 134, 343–382. doi: 10.1037/0033-2909.134.3.343
- Brainerd, C. J., Reyna, V. F., Wright, R., and Mojardin, A. H. (2003). Recollection rejection: false-memory editing in children and adults. *Psychol. Rev.* 110, 762–784. doi: 10.1037/0033-295X.110.4.762
- Cadavid, S., and Beato, M. S. (2016). Memory distortion and its avoidance: An event-related potentials study on false recognition and correct rejection. *PLoS One* 11:e0164024. doi: 10.1371/journal.pone.0164024
- Cadavid, S., and Beato, M. S. (2017). False recognition in DRM lists with extremely low association: A normative study. *Psicologica* 38, 133–147.
- Cadavid, S., Beato, M. S., and Fernandez, A. (2012). Falso reconocimiento en listas DRM con tres palabras críticas: Asociación directa vs. inversa [False recognition in DRM lists with three critical words: Forward vs. backward association]. *Psicologica* 33, 39–58.
- Carneiro, P., Fernandez, A., and Dias, A. R. (2009). The influence of theme identifiability on false memories?: Evidence for age-dependent opposite effects. *Memory Cogn.* 37, 115–129. doi: 10.3758/MC.37.2.115
- Carneiro, P., Fernandez, A., Díez, E., García-Marques, L., Ramos, T., and Ferreira, M. B. (2012). “Identify-to-reject”: A specific strategy to avoid false memories in the DRM paradigm. *Memory Cogn.* 40, 252–265. doi: 10.3758/s13421-011-0152-6
- Coane, J. H., McBride, D. M., and Xu, S. (2020). The feature boost in false memory: The roles of monitoring and critical item identifiability. *Memory* 28, 481–493. doi: 10.1080/09658211.2020.1735445
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspect. Psychol. Sci.* 6, 274–290. doi: 10.1177/1745691611406920
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fernandez, A., Díez, E., and Alonso, M. A. (2011). *Materiales normativos en castellano: Normas de asociación libre* [Normative materials in Castilian: Free association norms]. Available online at: <http://campus.usal.es/gimc/nalc>

- Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *J. Exp. Psychol. Learn. Memory Cogn.* 30, 120–128. doi: 10.1037/0278-7393.30.1.120
- Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. London: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory Cogn.* 38, 833–848. doi: 10.3758/MC.38.7.833
- Gallo, D. A., Bell, D. M., Beier, J. S., and Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory* 14, 730–741. doi: 10.1080/09658210600648506
- Gallo, D. A., and Lampinen, J. M. (2016). “Three pillars of false memory prevention: Orientation, evaluation, and corroboration,” in *The Oxford handbook of metamemory*, Chap. Oxford, eds J. Dunlosky, S. Uma, and K. Tauber (Oxford: Oxford University Press), 387–403. doi: 10.1093/oxfordhb/9780199336746.013.11
- Gallo, D. A., and Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *J. Memory Lang.* 47, 469–497. doi: 10.1016/S0749-596X(02)00013-X
- Gallo, D. A., and Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *J. Memory Lang.* 47, 469–497. doi: 10.3758/BF03196124
- Graham, L. M. (2007). Need for cognition and false memory in the Deese–Roediger–McDermott paradigm. *Personal. Individ. Diff.* 42, 409–418. doi: 10.1016/j.paid.2006.07.012
- Gray, S. G. (2016). *Transcranial direct current stimulation and episodic memory retrieval* [Doctoral dissertation, Department of Psychology. Chicago: University of Chicago, doi: 10.6082/M13F4MHX
- Howe, M. L., and Akhtar, S. (2020). Priming older adults and people with mild to moderate Alzheimer's disease problem-solving with false memories. *Cortex* 125, 318–331. doi: 10.1016/j.cortex.2020.01.014
- Huff, M. J., Bodner, G. E., and Gretz, M. R. (2020). Reducing false recognition in the Deese–Roediger/McDermott paradigm: Related lures reveal how distinctive encoding improves encoding and monitoring processes. *Front. Psychol.* 11:602347. doi: 10.3389/fpsyg.2020.602347
- Huff, M. J., Bodner, G. E., and Gretz, M. R. (2021). Distinctive encoding of a subset of DRM lists yields not only benefits, but also costs and spillovers. *Psycholog. Res.* 85, 280–290. doi: 10.1007/s00426-019-01241-y
- JASP Team (2020). *JASP* (Version 0.14.1) [Computer software].
- Jou, J., Escamilla, E. E., Torres, A. U., Ortiz, A., Perez, M., and Zuniga, R. (2018). Why seemingly more difficult test conditions produce more accurate recognition of semantic prototype words: A recognition memory paradox? *Conscious. Cogn.* 63, 239–253. doi: 10.1016/j.concog.2018.06.003
- Lampinen, J. M., Meier, C. R., Arnal, J. D., and Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *J. Exp. Psychol. Learn. Memory Cogn.* 31, 954–963. doi: 10.1037/0278-7393.31.5.954
- Liu, H., Gao, Q., Zheng, L., Wu, Y., Wang, C., Weng, X., et al. (2020). The neural correlates of context retrieval in false recognition. *NeuroReport* 31, 966–970. doi: 10.1097/WNR.0000000000001502
- Liu, Z., Liu, T., and Li, Y. (2021). How does social competition affect true and false recognition? *Psychon. Bull. Rev.* 28, 292–303. doi: 10.3758/s13423-020-01807-7
- McEvoy, C. L., Nelson, D. L., and Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *J. Exp. Psychol.* 25, 1177–1194. doi: 10.1037/0278-7393.25.5.1177
- Moore, K. N., Lampinen, J. M., Bridges, A. J., and Gallo, D. A. (2020). Developmental trends in children's use of different monitoring processes to avoid false memories. *Cogn. Dev.* 55:100911. doi: 10.1016/j.cogdev.2020.100911
- Nieznanski, M., Obidziński, M., Niedziałkowska, D., and Zyskowska, E. (2018). Context recollection and false memory of critical lures in the Deese/Roediger–McDermott paradigm: The role of encoding-and retrieval-based mechanisms. *Psycholog. Teme* 27, 365–384. doi: 10.31820/pt.27.3.2
- Pitarque, A., Satorres, E., Sales, A., Escudero, J., and Meléndez, J. C. (2018). Effects of stimuli repetition and age in false recognition. *Psycholog. Rep.* 121, 1106–1119. doi: 10.1177/0033294117747284
- Reyna, V. F., and Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learn. Individ. Diff.* 7, 1–75. doi: 10.1016/1041-6080(95)90031-4
- Robinson, K. J., and Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psycholog. Sci.* 8, 231–237. doi: 10.1111/j.1467-9280.1997.tb00417.x
- Roediger, H. L., Balota, D. A., and Watson, J. M. (2001a). “Spreading activation and arousal of false memories,” in *Science conference series. The nature of remembering: Essays in honor of Robert G. Crowder*, eds H. L. Roediger, J. S. Nairne, I. Neath, and A. M. Surprenant (Washington, D.C: American Psychological Association), 95–115. doi: 10.1037/10394-006
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *J. Exp. Psychol.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, H. L., Watson, J. M., McDermott, K. B., and Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407. doi: 10.3758/BF03196177
- Thakral, P. P., Madore, K. P., Devitt, A. L., and Schacter, D. L. (2019). Adaptive constructive processes: An episodic specificity induction impacts false recall in the Deese–Roediger–McDermott paradigm. *J. Exp. Psychol. Gen.* 148, 1480–1493. doi: 10.1037/xge0000577
- Wang, J., Otgaar, H., Santtila, P., Shen, X., and Zhou, C. (2021). How culture shapes constructive false memory. *J. Appl. Res. Memory Cogn.* 2021:002. doi: 10.1016/j.jarmac.2020.12.002.12.002

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Cadavid, Beato, Suarez and Albuquerque. 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) and the copyright owner(s) 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.



Questioning the Role of Forward Associative Strength in False Memories: Evidence From Deese-Roediger-McDermott Lists With Three Critical Lures

Maria Soledad Beato^{1*} and Jason Arndt²

¹Faculty of Psychology, University of Salamanca, Salamanca, Spain, ²Department of Psychology, Middlebury College, Middlebury, VT, United States

OPEN ACCESS

Edited by:

Sjoerd J. H. Ebisch,
University of Studies G. d'Annunzio
Chieti and Pescara, Italy

Reviewed by:

Marek Nieznański,
Cardinal Stefan Wyszyński University,
Poland
Enmanuelle Pardilla-Delgado,
Brigham and Women's Hospital,
Harvard Medical School,
United States
Dawn McBride,
Illinois State University,
United States

*Correspondence:

Maria Soledad Beato
msol@usal.es

Specialty section:

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

Received: 13 June 2021

Accepted: 16 August 2021

Published: 13 September 2021

Citation:

Beato MS and Arndt J (2021)
Questioning the Role of Forward
Associative Strength in False
Memories: Evidence From Deese-
Roediger-McDermott Lists With
Three Critical Lures.
Front. Psychol. 12:724594.
doi: 10.3389/fpsyg.2021.724594

We report an experiment examining the factors that produce false recognition in the Deese-Roediger-McDermott (DRM) paradigm. We selectively manipulated the probability that critical lures produce study items in free association, known as forward associative strength (FAS), while controlling the probability that study items produce critical lures in free association, known as backward associative strength (BAS). Results showed that false recognition of critical lures failed to differ between strong and weak FAS conditions. Follow-up correlational analyses further supported this outcome, showing that FAS was not correlated with false recognition, despite substantial variability in both variables across our stimulus sets. However, these correlational analyses did produce a significant and strong relationship between BAS and false recognition. These results support views that propose false memory is produced by activation spreading from study items to critical lures during encoding, which leads critical lures to be confused with episodically-experienced events.

Keywords: false memory, Deese-Roediger-McDermott paradigm, multiple critical lures, forward associative strength, backward associative strength

INTRODUCTION

False memory has been studied extensively using the Deese/Roediger-McDermott paradigm (DRM; Deese, 1959; Roediger and McDermott, 1995). In this paradigm, people study a list of words, which are all related to the same non-studied word known as the critical lure. On a subsequent memory test, critical lures are often mistakenly recalled or recognized (e.g., Arndt and Beato, 2017; Pitarque et al., 2018; Huff et al., 2020; Beato and Arndt, 2021). While this paradigm produces robust false memory, there is substantial variability in the false recognition that occurs across DRM lists (e.g., Gallo and Roediger, 2002; Beato and Díez, 2011; Cadavid and Beato, 2017; Coane et al., 2021).

In order to understand this variability, many studies have focused on the roles played by the probability that list items produce the critical lure in free association (referred to as backward associative strength or BAS) and the probability that the critical lure produces list

items in free association (referred to as forward associative strength or FAS). This work has shown that BAS influences false memory reliably (e.g., McEvoy et al., 1999; Roediger et al., 2001; Gallo and Roediger, 2002; Arndt, 2006; Arndt and Gould, 2006; Howe et al., 2009a; Knott et al., 2012). On the contrary, prior work examining the effect of FAS on false memory has produced inconsistent results. Some studies did not find significant correlations between FAS and false recall/recognition (e.g., Roediger et al., 2001; Gallo and Roediger, 2002; Beato and Arndt, 2014), while other studies found correlations between FAS and false memory (e.g., Brainerd and Wright, 2005; Howe et al., 2009b; Arndt, 2012b, 2015). Brainerd and Wright (2005) argued that the lack of correlation found between FAS and false memory by, for example, Roediger et al. (2001) was due to the restricted range of FAS values used in their lists. Indeed, this same criticism applies to most studies showing that FAS fails to predict false memory (see Beato and Arndt, 2014 for an exception).

Beyond the empirical considerations highlighted above, the question of whether FAS impacts false memory is important theoretically. One class of theories, associative activation theories (Roediger et al., 2001; Howe et al., 2009b), posit that false memory in the DRM paradigm is caused by the activation spreading from study items' representations in semantic memory to critical lures' representations. As a result, these theories propose that BAS, but not FAS, should impact false memory. In contrast, other theories suggest that featural similarity between study items and critical lures (Arndt and Hirshman, 1998; Brainerd et al., 2008; Arndt, 2012a; Brainerd et al., 2020) increases false memory. In the view of these theories, both BAS and FAS should impact false memory, because both variables can be interpreted to index the extent to which study items and critical lures share features. Thus, investigating whether FAS influences false memory will help to distinguish between theoretical views of false memory's genesis.

In the present study, we sought to determine whether FAS was related to false memory using lists that had substantial variability in FAS. In order to understand the unique role that FAS plays in producing false memories, it is important to separate the contributions of FAS and BAS to false memory, given that they are correlated (Brainerd et al., 2008). For this purpose, we built DRM lists that varied widely in FAS, while controlling BAS and employed DRM lists that were related to multiple critical lures (e.g., Beato et al., 2012; Cadavid et al., 2012; Beato and Arndt, 2014). Constructing lists that were related to multiple critical lures allowed us to evaluate correlations between FAS and false recognition using both the variability due to list-wide characteristics (the approach used in prior work, where lists are related to a single critical lure) and the variability due to individual critical lures' characteristics as the basis for analyses. To illustrate, consider the lists that were related to the general theme "School," we constructed one list of associates (homework, work, school, book, class, and test) that had strong FAS with three critical lures (assignment, lesson, and study), and a second list of associates (school, book, work, boring, long, and test) that had weak FAS with three critical

lures (essay, homework, and study).¹ Within a list that had a strong FAS-based relationship with its critical lures, there was variation in the summed FAS values of the three critical lures (e.g., assignment=0.755, lesson=0.294, and study=0.341). Importantly, lists that had a weak FAS-based relationship with its critical lures also varied in the summed FAS values of the three critical lures (e.g., essay=0.158, homework=0.265, and study=0.321).

The benefit of constructing lists this way is that it enabled us to examine the effects of study item characteristics and critical lure characteristics on false recognition separately. In contrast, the standard DRM paradigm, where each list of words is related to a single critical lure, only allows evaluation of these characteristics simultaneously, making it impossible to understand their unique impact on false memory. Thus, if the results of correlational analyses using study item and critical lure characteristics converge, it is appropriate to infer the variables correlated with false recognition reflect a general property of DRM lists, as well as the factors that drive false memory. On the other hand, if the results of correlational analyses produce different results using study item and critical lure characteristics as the unit of analysis, one may have less confidence that they reflect a general property of the factors that drive false memory, and instead may reflect specific item characteristics. As a consequence, factors that show the same effects for both sets of analyses are more likely to be key drivers of false memory effects, making them carry greater theoretical importance.

MATERIALS AND METHODS

Participants

The sample comprised 40 native English speakers (70% female) who participated as part of a course research appreciation requirement. Participants' ages ranged from 18 to 20 years ($M_{age}=18.60$; $SD=0.78$).

Materials

A total of 28 six-word DRM lists were constructed as stimuli (see **Table 1**). Lists were built to ensure that three critical lures produced the same six study items in free association (i.e., they were related *via* FAS) based upon the University of South Florida free-association norms (Nelson et al., 1998). This list length was chosen because it allowed us the best opportunity to construct lists with multiple critical lures that manipulate FAS while controlling BAS across levels of FAS. While there is a tendency for critical lure false alarms to be lower with shorter lists, studying six associates of critical lures produces robust false memory (e.g., Robinson and Roediger, 1997).

¹The astute reader will notice that some words are repeated across the two sets of critical lures and study items. As is clarified in the method section, participants studied either the strong or weak FAS lists for a given general theme. Thus, the overlap in specific stimulus items is not problematic for assessing the effects of FAS on false recognition.

TABLE 1 | The 28 six-word study lists with three critical lures, the general theme for each pair of high- and low-FAS list, and the forward associative strength (FAS) condition for each list.

List	CRITICAL LURES: Associated words	General theme	FAS condition
List 1	ADVIL, EXCEDRIN, TYLENOL: headache, aspirin, medicine, pain, pill, drug	Medicine	High
List 2	OPERATION, SURGERY, TRAUMA: pain, hospital, blood, heart, hurt, sick	Medicine	Low
List 3	BOURBON, BRANDY, VODKA: drink, alcohol, liquor, drunk, whiskey, rum	Alcohol	High
List 4	BOTTLE, DRINK, WHISKEY: beer, drunk, liquor, wine, glass, alcohol	Alcohol	Low
List 5	COLONEL, CORPORAL, LIEUTENANT: army, sergeant, military, officer, captain, general	Army	High
List 6	COLONEL, COMMANDER, CORPORAL: sergeant, lieutenant, military, general, captain, officer	Army	Low
List 7	SOCCER, SOFTBALL, VOLLEYBALL: ball, game, sport, football, team, play	Sports	High
List 8	PLAYER, SPORTS, VOLLEYBALL: football, game, basketball, team, tennis, soccer	Sports	Low
List 9	BURIAL, BURY, GRAVE: dead, death, funeral, cemetery, die, ground	Death	High
List 10	DEAD, DEATH, DIE: sad, end, funeral, bury, heaven, grave	Death	Low
List 11	SHOWER, SPONGE, WASHCLOTH: clean, water, bath, soap, wet, wash	Soap	High
List 12	SPONGE, TOWEL, WASHCLOTH: water, bath, soap, wet, rag, wash	Soap	Low
List 13	CROPS, FARMER, HARVEST: corn, farm, food, wheat, field, vegetables	Food	High
List 14	CAKE, DESSERT, PASTRY: sweet, food, pie, fat, good, yummy	Food	Low
List 15	PURCHASE, SALE, SHOP: buy, store, clothes, money, mall, sell	Buy	High
List 16	PRODUCT, PURCHASE, SALE: buy, item, store, sell, car, price	Buy	Low
List 17	BURGLAR, THEFT, THIEF: steal, robber, crook, crime, criminal, bad	Crime	High
List 18	BURGLAR, FRAUD, ROBBERY: thief, steal, crime, money, crook, criminal	Crime	Low
List 19	JUPITER, NEPTUNE, URANUS: planet, mars, pluto, saturn, venus, space	Planet	High
List 20	JUPITER, NEPTUNE, PLUTO: mars, saturn, venus, uranus, moon, space	Planet	Low
List 21	ASSIGNMENT, LESSON, STUDY: homework, work, school, book, class, test	School	High
List 22	ESSAY, HOMEWORK, STUDY: school, book, work, boring, long, test	School	Low
List 23	NORM, NORMAL, ROUTINE: abnormal, average, same, regular, usual, boring	Normal	High
List 24	COMMON, ROUTINE, STANDARD: same, everyday, usual, average, ordinary, regular	Normal	Low
List 25	COMMENT, REMARK, RESPOND: answer, talk, say, speak, reply, tell	Talk	High
List 26	COMMENT, REMARK, SUGGEST: talk, answer, opinion, say, tell, speak	Talk	Low
List 27	CLARINET, TRUMPET, TUBA: instrument, music, horn, flute, band, blow	Music	High
List 28	TROMBONE, TRUMPET, TUBA: band, loud, flute, brass, clarinet, blow	Music	Low

The FAS values for each critical lure word (*lure word FAS* hereafter) were determined by the sum of the probabilities that each critical lure produced its six associated words in free association. Similarly, the FAS values for each list (*FAS list strength* hereafter) were calculated as the sum of the FAS values for the three critical lures (similar to Robinson and Roediger, 1997; Beato and Díez, 2011; Beato and Arndt, 2014). BAS values, measured as the probability that study items produced critical lures in free association, were similarly calculated for both critical lures and lists.

There were 14 “general themes” in the lists (e.g., *music*). For each theme, we built two different six-word study lists. One of the two study lists per general theme included six associates that had relatively stronger FAS relations to critical lures (*high-FAS lists* hereafter) and the other included six associates that had relatively weaker FAS relations to critical lures (*low-FAS lists* hereafter). For example, for the general theme “music” the high-FAS critical lures were CLARINET (FAS=0.660), TRUMPET (FAS=0.727), and TUBA (FAS=0.576), each of which had forward associations to the study items *instrument, music, horn, flute, band, and blow*. The low-FAS critical lures for the same theme were TROMBONE (FAS=0.275), TRUMPET (FAS=0.202), and TUBA (FAS=0.210), each of which had forward associations to the study items *band, loud, flute, brass, clarinet, and blow*. Lists constructed with high-FAS list strength ($M=1.641$; $SD=0.367$; ranged from 1.073 to 2.301), and low-FAS list strength ($M=0.787$; $SD=0.183$; ranged from 0.454 to 1.112) reliably differed from one another, $t(26)=7.79$; $p<0.001$, $d=2.94$, 95% CI [0.63, 1.08].

Further, there was a wide variability in FAS, such that FAS list strength ranged from 0.45 to 2.30² and lure word FAS ranged from 0.127 to 0.865. Finally, we constructed lists in a way that controlled mean levels of BAS across the stimulus sets that varied in FAS ($M=0.34$ and $M=0.29$, for the high- and low-FAS lists, respectively), $t(26)=0.37$; $p>0.05$, although BAS still varied substantially across stimulus sets and for the critical lures within a stimulus set. **Table 2** reports FAS and BAS values per critical lure.

Finally, we built five additional six-word lists, each with three critical lures (see **Table 3**), to be used as *unrelated distractors* and *unrelated critical-lure distractors* on the memory tests. Distractor lists were constructed to ensure that they were associatively unrelated to study items (Nelson et al., 1998) following a procedure similar to that used to construct study lists, such that FAS list strength ranged from 0.93 to 1.46. The recognition memory test included 168 words randomly intermixed: the 84 studied words, the 42 critical lures, and 42 distractors (15 unrelated critical-lure distractors, 27 unrelated distractors).

Procedure

First, participants were informed about the nature and procedure of the study and signed a consent form. Participants were

²The range of mean FAS values was between 0.025 and 0.128. The use of the sum or mean of FAS values to determine FAS list strength and lure word FAS does not alter in any way the conclusions obtained in the different analyses as they are linear transformations of one another.

TABLE 2 | Forward associative strength (FAS) condition, mean percentages of true recognition (TR) and false recognition (FR) per list, mean percentages of false recognition per critical lure, and FAS and BAS values per critical lure were included.

List	FAS condition	TR list	FR list	FR Lure 1	FR Lure 2	FR Lure 3	FAS Lure 1	FAS Lure 2	FAS Lure 3	BAS Lure 1	BAS Lure 2	BAS Lure 3
List 1	High	78.33	5.00	15	0	0	0.626	0.865	0.810	0.015	0.000	0.107
List 2	Low	77.50	8.33	5	20	0	0.271	0.219	0.168	0.000	0.000	0.000
List 3	High	94.17	15.00	15	15	15	0.664	0.627	0.544	0.051	0.000	0.184
List 4	Low	75.00	21.67	10	50	5	0.337	0.288	0.354	0.032	0.686	0.000
List 5	High	75.00	16.67	30	10	10	0.544	0.421	0.684	0.014	0.000	0.054
List 6	Low	79.17	28.33	30	35	20	0.275	0.148	0.204	0.082	0.014	0.021
List 7	High	72.50	3.33	5	5	0	0.848	0.400	0.307	0.124	0.000	0.000
List 8	Low	73.33	10.00	15	10	5	0.368	0.446	0.229	0.116	0.000	0.000
List 9	High	81.67	38.33	20	40	55	0.658	0.355	0.516	0.067	0.056	0.196
List 10	Low	86.67	36.67	40	40	30	0.137	0.190	0.127	0.553	0.786	0.067
List 11	High	70.83	10.00	20	10	0	0.539	0.480	0.370	0.355	0.000	0.000
List 12	Low	71.67	8.33	10	10	5	0.393	0.210	0.322	0.000	0.162	0.000
List 13	High	70.00	21.67	25	25	15	0.753	0.246	0.391	0.024	0.015	0.000
List 14	Low	72.50	8.33	10	10	5	0.223	0.258	0.261	0.168	0.016	0.000
List 15	High	76.67	38.33	0	40	75	0.758	0.407	0.746	0.155	0.050	0.519
List 16	Low	64.17	25.00	15	15	45	0.194	0.676	0.242	0.000	0.155	0.092
List 17	High	81.67	45.00	45	20	70	0.265	0.476	0.739	0.090	0.000	0.988
List 18	Low	73.33	18.33	30	10	15	0.421	0.140	0.311	0.126	0.000	0.028
List 19	High	75.00	15.00	25	10	10	0.829	0.700	0.732	0.183	0.039	0.091
List 20	Low	83.33	28.33	40	20	25	0.375	0.236	0.220	0.170	0.047	0.100
List 21	High	64.17	25.00	5	15	55	0.755	0.294	0.341	0.015	0.000	0.356
List 22	Low	70.00	35.00	5	35	65	0.158	0.265	0.321	0.000	0.036	0.163
List 23	High	63.33	21.67	0	60	5	0.385	0.576	0.283	0.020	0.723	0.034
List 24	Low	69.17	10.00	15	10	5	0.212	0.254	0.145	0.220	0.092	0.000
List 25	High	55.83	10.00	5	0	25	0.276	0.247	0.550	0.000	0.000	0.000
List 26	Low	49.17	5.00	10	0	5	0.322	0.240	0.171	0.000	0.000	0.000
List 27	High	78.33	16.67	0	45	5	0.660	0.727	0.576	0.097	0.143	0.013
List 28	Low	75.00	16.67	20	30	0	0.275	0.202	0.210	0.014	0.165	0.000

tested individually and were instructed that their task was to remember the words as best they could, because they would be given a memory test later in the experiment.

Each participant was presented with the study items from 14 lists, one list per general theme (seven high-FAS lists and seven low-FAS lists). General themes' high-FAS lists and low-FAS lists were presented equally often across participants. Further, we confirmed that no associates or critical lures were repeated within the stimuli experienced by a participant. Study items were presented individually on a computer screen for 2,000 ms with a 500-ms ISI blocked by DRM list. The associates within each list were arranged in decreasing order of FAS. The order of list presentation was randomized for each participant. At the conclusion of the study phase, participants completed a self-paced recognition memory test, where they were asked to determine whether each word was previously studied by pressing the "O" key to indicate it was OLD, and the "N" key to indicate it was NEW.

Power Analysis

We evaluated the power to detect the effects of FAS on false recognition using the three strategies highlighted above using G*Power (Faul et al., 2007, 2009). When participants were used as the unit of analysis, our sample size of 40 was sufficient to detect a large effect ($d_z=0.5$), with power=0.869. However, the power to detect a medium-sized effect ($d_z=0.3$) was considerably smaller, 0.457. When study lists ($N=28$) were

used as the unit of analysis for correlations, the power to detect a large effect ($\rho=0.5$) was 0.799, which is near the conventionally-preferred level of 0.80 (Cohen, 1988). However, the power to detect a medium-sized effect ($\rho=0.3$) was considerably lower, 0.348. Finally, when critical lures ($N=84$) were used as the unit of analysis for correlations, the power to detect a large effect ($\rho=0.5$) and a medium-sized effect ($\rho=0.3$) were both sufficient by conventional standards, with power of 0.999 and 0.800, respectively.

Data Analysis

Given the relatively modest levels of power to detect medium-sized effects in most of our analyses, we chose to analyze our data using both standard null-hypothesis tests and Bayesian analysis (Kass and Raftery, 1995), which allows quantification of the strength of the evidence for the null and alternative hypothesis. We conducted standard analyses that treat FAS as a categorical variable (e.g., t -tests) and as a continuous variable (correlation) to assess its relationship to false recognition. We also conducted correlational analyses on a variety of other characteristics of our stimuli (Nelson et al., 1998) to evaluate how well semantic memory variables other than FAS predicted false recognition (see, e.g., Brainerd et al., 2008). As highlighted above, these analyses were conducted using study lists' characteristics as the unit of analysis and using critical lures' characteristics as the unit of analysis. Finally, we conducted Bayesian analyses using JASP (Version 0.14.1; JASP Team, 2020)

TABLE 3 | Five six-word distractor lists with three critical lures, general theme, and FAS were included.

List	DISTRACTOR LIST CRITICAL LURES: Associated words	General theme	FAS distractor 1	FAS distractor 2	FAS distractor 3	FAS list distractors
List 1	ATTRACTIVE, GORGEOUS, SEXY: pretty, beautiful, girl, man, handsome, woman	Pretty	0.609	0.420	0.226	1.255
List 2	COUNTY, LOCAL, PROVIDENCE: city, state, place, country, area, town	Place	0.454	0.139	0.335	0.928
List 3	DRAWING, PAINTER, PAINTING: art, picture, artist, paint, color, canvas	Paint	0.497	0.422	0.538	1.457
List 4	CRUDE, REPULSIVE, VULGAR: rude, gross, disgusting, mean, ugly, nasty	Disgusting	0.510	0.546	0.393	1.449
List 5	AX, CHISEL, HATCHET: hammer, knife, chop, tool, cut, saw	Tools	0.263	0.466	0.240	0.969

to quantify the strength of the evidence for the observed statistical outcomes using BF_{10} . $BF_{10} > 1$ supports the alternate hypothesis, and a $BF_{10} < 1$ supports the null hypothesis. Importantly, BF_{10} between 3 and 20 signifies positive evidence for the alternate hypothesis, BF_{10} between 20 and 150 signifies strong evidence for the alternate hypothesis, and BF_{10} greater than 150 signifies very strong evidence for the alternate hypothesis (Kass and Raftery, 1995). Similarly, BF_{10} between 0.33 and 0.05 signifies positive evidence for the null hypothesis, BF_{10} between 0.05 and 0.0067 signifies strong evidence for the null hypothesis, and BF_{10} below 0.0067 signifies very strong evidence for the null hypothesis.³

RESULTS

Table 2 reports the mean percentage of true recognition per list and false recognition per critical lure and list, while **Table 4** reports the mean percentage of true and false recognition as a function of FAS and whether items were studied, critical lures, unrelated critical-lure distractors, or unrelated distractors.

False Memory Effect

A one-way repeated-measures analysis of variance (ANOVA) was used to compare the percentage of old judgments to studied words, critical lures, unrelated critical-lure distractors, and unrelated distractors. This analysis revealed a significant difference, $F(3, 117) = 521.535$; $p < 0.001$, $\eta^2_p = 0.930$. Bonferroni *post-hoc* tests showed that hits to studied words (true recognition; $M = 73.48$, $SD = 13.06$) were higher than false alarms to critical lures (false recognition; $M = 19.35$, $SD = 10.64$), unrelated critical-lure distractors ($M = 5.67$, $SD = 7.78$) and unrelated distractors ($M = 5.92$, $SD = 7.45$; $p < 0.001$ for all comparisons). There were also significant differences between false alarms to critical lures and both unrelated critical-lure distractor and unrelated distractor items ($p < 0.001$). There was not a reliable difference between the two types of unrelated distractors ($p > 0.05$). Thus, the stimuli we constructed for this study produced the typical DRM false memory effect.

True Recognition, False Recognition, and FAS

The percentage of hits (true recognition) and false alarms to critical lures (false recognition) as a function of FAS is presented

in **Table 4**. FAS did not impact hits, $t(39) = 0.630$; $p = 0.532$, $d = 0.09$, $BF_{10} = 0.206$, or false alarms to critical lures, $t(39) = 0.868$; $p = 0.391$, $d = 0.13$, $BF_{10} = 0.242$. The BF_{10} values for hits and false alarms to critical lures indicate positive support for the conclusion that FAS failed to impact true and false recognition.

We also examined the relationship between FAS and false recognition using correlation. Our DRM lists included three critical lures per list, which allowed us to correlate FAS list strength and lure word FAS with false recognition separately. Neither of these analyses produced a significant correlation [$r(26) = -0.026$, $p = 0.895$, $BF_{10} = 0.237$ for FAS list strength; $r(82) = 0.047$, $p = 0.668$, and $BF_{10} = 0.149$ for lure word FAS]. The BF_{10} values for these correlations again indicate positive evidence that FAS was unrelated to false recognition. It is important to note that these null correlations occurred despite there being substantial variability in false recognition across lists and critical lures. For example, some high-FAS lists yielded high levels of false recognition, such as the list with the critical lures BURGLAR, THEFT, and THIEF (45%), whereas other high-FAS lists produced very low levels of false recognition (e.g., SOCCER, SOFTBALL, and VOLLEYBALL list, 3%). In low-FAS lists, we also found wide differences in false recognition, ranging between 37% (DEAD, DEATH, and DIE list) and 5% (COMMENT, REMARK, and SUGGEST list).

Although FAS was unrelated to false recognition, we sought to explore whether other stimulus characteristics were related to false recognition. Thus, we correlated the characteristics of the study words included in the lists with the overall level of false recognition produced by that list (i.e., averaged across the three critical lures). The variables examined in these analyses were BAS, interconnectivity among the associates included in the lists (sum of the FAS and BAS of all possible pairings of study items and critical lures), associates' set size, associates' concreteness, the mean connectivity among lists' associates, the probability of a resonant connection, and resonant connection strength (Nelson et al., 1998). The only reliable correlation found in these analyses was between false recognition and BAS, $r(26) = 0.643$, $p < 0.001$, $BF_{10} = 152.01$, indicating very strong evidence that false recognition and BAS were related.⁴

We also conducted correlational analyses between false recognition and critical lure characteristics. We computed each

³The BF_{10} values that quantify support for the null hypothesis are computed as the multiplicative inverse of the BF_{10} values that quantify support for the alternate hypothesis.

⁴While it may seem unusual that BAS was related to false recognition because we controlled BAS across the high- and low-FAS conditions, the method we used to equate BAS across FAS conditions only ensured that the *mean* BAS values were comparable across high- and low-FAS lists. Thus, this method left substantial variability in BAS across lists, which allowed the possibility that correlational analyses would reveal the impact of BAS on false recognition.

TABLE 4 | Mean percentage of true recognition and false recognition as a function of FAS, as well as baseline false alarm rates to unrelated critical-lure distractors and unrelated distractors.

	True recognition	False recognition
High FAS	74.11 (9.36)	20.12 (12.72)
Low FAS	72.86 (8.91)	18.57 (10.70)
Unrelated critical-lure distractors		5.67 (7.78)
Unrelated distractors		5.92 (7.45)

Standard deviations are reported in parenthesis.

critical lure's average BAS with study items, frequency of occurrence, concreteness, set size (i.e., number of different words produced by a critical lure), density (i.e., mean connectivity among all critical lure associates), accessibility index (i.e., number of word that produced the critical lure as a response), resonant connections (i.e., number of critical lure's associates that produced it as an associate), and resonant strength (i.e., associative strength from all the words produced by the critical lure to the critical lure; Nelson et al., 1998). This analysis showed there were significant correlations between false recognition and critical lures' BAS, $r(82)=0.662$, $p<0.001$, $BF_{10}>1,422,000,000$, frequency, $r(82)=0.388$, $p<0.001$, $BF_{10}=95.27$, resonant connections, $r(82)=0.463$, $p<0.001$, $BF_{10}=2099.64$, resonant strength, $r(82)=0.470$, $p<0.001$, $BF_{10}=2982.68$, and accessibility, $r(82)=0.479$, $p<0.001$, $BF_{10}=4735.24$. For each correlation, BF_{10} indicated strong or very strong evidence each variable was positively related to false recognition.

DISCUSSION

The empirical and theoretical aim of this research was to analyze the effect of FAS on false recognition. In order to do this, we constructed stimulus sets that varied widely in FAS. The results of this study showed that false recognition was robust. Moreover, there was wide variability in false recognition rates per list, ranging from 3 to 45%. Thus, there was substantial variability in both false recognition and FAS, which is critical for assessing if there is a relationship between FAS and false recognition.

Despite empirical conditions that were conducive to observing a relationship between FAS and false recognition, no such relationship was found. This finding replicates previous research that has failed to find a relationship between FAS and false recognition (e.g., Roediger et al., 2001; Gallo and Roediger, 2002; Beato and Arndt, 2014), but stands in contrast to research that has found a reliable relationship between FAS and false recognition (e.g., Brainerd and Wright, 2005; Arndt, 2012b, 2015). Importantly, interpretation of the present results is not complicated by restricted range in FAS, a concern that has been advanced to explain the finding of Roediger et al. (2001) that FAS was not correlated with false recognition (see Brainerd and Wright, 2005). Finally, the present results extend prior findings of a null correlation between FAS and false recognition to DRM lists related to multiple critical lures.

Although FAS failed to predict false memory, our correlational analyses produced several notable results. Most importantly, BAS was associated with false recognition. This association is particularly notable because we sought to control the mean levels of this variable across the high- and low-FAS conditions. Despite this constraint, BAS was strongly correlated with false recognition, replicating extensive evidence that BAS is a reliable predictor of false memory (e.g., Roediger et al., 2001; Gallo and Roediger, 2002; Arndt, 2012b, 2015; Beato and Arndt, 2017). Beyond BAS, our correlational analyses found that the factors that were correlated with greater false recognition generally measured the extent to which a critical lure is activated by the study of its associates, such as resonant connections and resonant strength. Thus, these measures may reflect, like BAS, how active a critical lure's representation is following study of its associates (Roediger et al., 2001).

At a theoretical level, the present results fit most naturally with associative activation views of false memory (Roediger et al., 2001; Howe et al., 2009b). These views posit that spreading activation from study item representations to critical lure representations plays a key role in producing false memory. Two cardinal predictions from these theories are upheld by the present data. First, that the primary driver of false memory is the extent to which study items activate lure representations in semantic memory, and thus the extent to which lure items can be confused with episodically-experienced items. This activation is most directly measured by BAS in word association norms. Second, that associative variables, which are unrelated to how much study items activate critical lures' representations, such as FAS, will not affect false memory. Both of these predictions were supported in the present study, despite the fact that we implemented a strong manipulation of FAS between lists and sought to control BAS across levels of that manipulation.

In addition to favoring associative-activation theories of false memory, the present results are puzzling from the perspective of theories highlighting the role that the similarity between study items and critical lures in semantic memory plays in producing false memory (Arndt and Hirshman, 1998; Brainerd et al., 2008; Arndt, 2012a; Brainerd et al., 2020). In particular, these views suggest that false memory increases with the similarity between study items and lure items, as well as the extent to which study lists' gist is encoded during study (Brainerd et al., 2020). Thus, FAS, BAS, and other measures of semantic memory activation should increase critical lure false memory. In contrast to this expectation, FAS failed to produce differences in false memory in this study, despite our intentional and substantial manipulation of this variable. In addition, FAS failed to correlate with false recognition, both when measured based upon study list characteristics and when measured based upon critical lure characteristics. Finally, in our analysis of list-wide semantic memory variables with false recognition as well as critical lures' semantic memory characteristics, the only correlation we found in *both* sets of analyses was between BAS and false recognition.

One set of outcomes from the present study may be taken as partial evidence favoring the view that similarity among study items enhances gist encoding, which is hypothesized

to play a role in false memory production (Brainerd et al., 2008, 2020). In our analyses of semantic memory variables associated with false memory, several semantic memory variables, such as critical lures' word frequency, resonant connections, resonant strength, and accessibility were all correlated with false recognition. While this broader set of semantic memory variables associating with false recognition is consistent with general semantic memory activation underlying false recognition (Brainerd et al., 2008), it is critically important that other key semantic variables, such as connectivity, failed to correlate with false recognition (Brainerd et al., 2020).⁵ Indeed, it has been suggested that connectivity can serve as a proxy measure for a study lists' gist, since it assesses inter-relationships among studied items, which can be viewed as assessing, in part, the semantic relationships among studied items that are thought to underlie a study list's overall gist (Brainerd et al., 2020). Thus, while views proposing that non-associative semantic memory activation underlies false memory are consistent with some aspects of the present data, the correlations observed in our results were (1) not as wide-ranging as would be expected if semantic memory activation is the primary basis for false memory and (2) not reliable for key variables thought to be proxy-measures of gist processing during encoding.

In closing, we wish to emphasize four key points. First, we failed to observe a correlation between FAS and false recognition, despite using conditions that provide an excellent opportunity for such a relationship to be found. Second, we observed a positive relationship between BAS and false recognition, despite not directly attempting to manipulate BAS in this study. Third, both of these results occurred when

we assessed the relationship between list-wide associative strength and false recognition as well as when we assessed the relationship at the level of individual critical lures. Importantly, because the FAS and BAS results occurred regardless of the method we used to calculate FAS, BAS, and false recognition, it suggests that the relationships we observed in this study are products of the nature of the associations between study lists and critical lures. Fourth, and finally, these results favor activation-based explanations of false memory (Roediger et al., 2001; Howe et al., 2009b) over similarity-based explanations (Arndt and Hirshman, 1998; Brainerd et al., 2008; Arndt, 2012a; Brainerd et al., 2020). Thus, these results best support the view that study items in the DRM paradigm activate critical lures' representations during encoding, which leads critical lures to be falsely recognized on a subsequent memory test.

DATA AVAILABILITY STATEMENT

The data collected for this study will be made available by the authors on request.

ETHICS STATEMENT

The research reported in this paper was reviewed and approved by the Middlebury College Institutional Review Board. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The authors contributed equally to this work.

FUNDING

This research was partially supported by the University of Salamanca and by grant SA052G18, awarded to MSB by Junta de Castilla y León.

⁵Connectivity was not correlated with false recognition when study lists were used as the unit of analysis [$r(26) = 0.025$, $p = 0.901$, $BF_{10} = 0.236$] and was negatively correlated with false recognition when critical lures were used as the unit of analysis, albeit not significantly so [$r(82) = -0.180$, $p = 0.102$, $BF_{10} = 0.507$]. While BF_{10} for this latter correlation falls in the range, where it fails to provide support for the null hypothesis, evaluating the statistical hypothesis that connectivity and false recognition were positively correlated, as gist-based perspectives predict, produces a value of p of 0.949 and a $BF_{10} = 0.054$, which falls in the range of positive evidence that the two variables are unrelated, and is close to the range where the evidence is considered "strong."

REFERENCES

- Arndt, J. (2006). Distinctive information and false recognition: the role of encoding and retrieval processes. *J. Mem. Lang.* 54, 113–130. doi: 10.1016/j.jml.2005.08.003
- Arndt, J. (2012a). "False recollection: empirical findings and their theoretical implications" in *Psychology of Learning and Motivation*. ed. B. H. Ross. Vol. 56 (Amsterdam, The Netherlands: Academic Press), 81–124.
- Arndt, J. (2012b). The influence of forward and backward associative strength on false recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 747–756. doi: 10.1037/a0026375
- Arndt, J. (2015). The influence of forward and backward associative strength on false memories for encoding context. *Memory* 23, 1093–1111. doi: 10.1080/09658211.2014.959527
- Arndt, J., and Beato, M. S. (2017). The role of language proficiency in producing false memories. *J. Mem. Lang.* 95, 146–158. doi: 10.1016/j.jml.2017.03.004
- Arndt, J., and Gould, C. (2006). An examination of two process theories of false recognition. *Memory* 14, 814–833. doi: 10.1080/09658210600680749
- Arndt, J., and Hirshman, E. (1998). True and false recognition in MINERVA2: explanations from a global-matching perspective. *J. Mem. Lang.* 39, 371–391. doi: 10.1006/jmla.1998.2581
- Beato, M. S., and Arndt, J. (2014). False recognition production indexes in forward associative strength (FAS) lists with three critical words. *Psicothema* 26, 457–463. doi: 10.7334/psicothema2014.79
- Beato, M. S., and Arndt, J. (2017). The role of backward associative strength in false recognition of DRM lists with multiple critical lures. *Psicothema* 29, 358–363. doi: 10.7334/psicothema2016.248
- Beato, M. S., and Arndt, J. (2021). The effect of language proficiency and associative strength on false memory. *Psychol. Res.* doi: 10.1007/s00426-020-01449-3
- Beato, M. S., and Díez, E. (2011). False recognition production indexes in Spanish for 60 DRM lists with three critical words. *Behav. Res. Methods* 43, 499–507. doi: 10.3758/s13428-010-0045-9

- Beato, M. S., Boldini, A., and Cadavid, S. (2012). False memory and level of processing effect: an event-related potential study. *Neuroreport* 23, 804–808. doi: 10.1097/WNR.0b013e32835734de
- Brainerd, C. J., and Wright, R. (2005). Forward association, backward association, and the false-memory illusion. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 554–567. doi: 10.1037/0278-7393.31.3.554
- Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., and Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychon. Bull. Rev.* 15, 1035–1053. doi: 10.3758/PBR.15.6.1035
- Brainerd, C. J., Chang, M., and Bialer, D. M. (2020). From association to gist. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 2106–2127. doi: 10.1037/xlm0000938
- Cadavid, S., and Beato, M. S. (2017). False recognition in DRM lists with low association: A normative study. *Psicológica* 38, 133–147.
- Cadavid, S., Beato, M. S., and Fernández, A. (2012). Falso reconocimiento en listas DRM con tres palabras críticas: Asociación directa vs. inversa [false recognition in DRM lists with three critical words: forward vs. backward association]. *Psicológica* 33, 39–58.
- Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., and Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: A review of how structural and conceptual similarity affect false memory. *Front. Psychol.* 12:668550. doi: 10.3389/fpsyg.2021.668550
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Edn). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Gallo, D. A., and Roediger, H. L. III (2002). Variability among word lists in eliciting memory illusions: evidence for associative activation and monitoring. *J. Mem. Lang.* 47, 469–497. doi: 10.1016/S0749-596X(02)00013-X
- Howe, M. L., Wimmer, M. C., and Blease, K. (2009a). The role of associative strength in children's false memory illusions. *Memory* 17, 8–16. doi: 10.1080/09658210802438474
- Howe, M. L., Wimmer, M. C., Gagnon, N., and Plumpton, S. (2009b). An associative-activation theory of children's and adults' memory illusions. *J. Mem. Lang.* 60, 229–251. doi: 10.1016/j.jml.2008.10.002
- Huff, M. J., Bodner, G. E., and Gretz, M. R. (2020). Reducing false recognition in the Deese-Roediger/McDermott paradigm: related lures reveal how distinctive encoding improves encoding and monitoring processes. *Front. Psychol.* 11, 602347. doi: 10.3389/fpsyg.2020.602347
- JASP Team (2020). JASP (Version 0.14.1) [Computer software].
- Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.1080/01621459.1995.10476572
- Knott, L. M., Dewhurst, S. A., and Howe, M. L. (2012). What factors underlie associative and categorical memory illusions? The roles of backward associative strength and inter-item connectivity. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 229–239. doi: 10.1037/a0025201
- McEvoy, C. L., Nelson, D. L., and Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1177–1194. doi: 10.1037//0278-7393.25.5.1177
- Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available at: <http://w3.usf.edu/FreeAssociation> (Accessed October 19, 2005).
- Pitarque, A., Satorres, E., Sales, A., Escudero, J., and Meléndez, J. C. (2018). Effects of stimuli repetition and age in false recognition. *Psychol. Rep.* 121, 1106–1119. doi: 10.1177/0033294117747284
- Robinson, K. J., and Roediger, H. L. III (1997). Associative processes in false recall and false recognition. *Psychol. Sci.* 8, 231–237. doi: 10.1111/j.1467-9280.1997.tb00417.x
- Roediger, H. L. III, and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, H. L. III, Watson, J. M., McDermott, K. B., and Gallo, D. A. (2001). Factors that determine false recall: a multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407. doi: 10.3758/BF03196177

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Beato and Arndt. 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) and the copyright owner(s) 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.



Transcranial Direct Current Stimulation Over the Right Anterior Temporal Lobe Does Not Modulate False Recognition

María Angeles Alonso^{1,2*}, Antonio M. Díez-Álamo², Carlos J. Gómez-Ariza³, Emiliano Díez^{1,2} and Angel Fernandez^{1,2}

¹ Instituto Universitario de Neurociencia (IUNE), University of La Laguna, San Cristóbal de La Laguna, Spain, ² Institute on Community Integration (INICO), University of Salamanca, Salamanca, Spain, ³ Department of Psychology, University of Jaén, Jaén, Spain

OPEN ACCESS

Edited by:

Beatriz Martín-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

Jovana Bjekic,
University of Belgrade, Serbia
Kilian Abellaneda-Pérez,
University of Barcelona, Spain

*Correspondence:

María Angeles Alonso
maalonso@ull.edu.es

Specialty section:

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

Received: 31 May 2021

Accepted: 20 August 2021

Published: 17 September 2021

Citation:

Alonso MA, Díez-Álamo AM, Gómez-Ariza CJ, Díez E and Fernandez A (2021) Transcranial Direct Current Stimulation Over the Right Anterior Temporal Lobe Does Not Modulate False Recognition. *Front. Psychol.* 12:718118. doi: 10.3389/fpsyg.2021.718118

Non-invasive transcranial direct current stimulation (tDCS) over the left anterior temporal lobe (ATL) has been shown to cause a reduction in the rate of false memories with semantically related words. Such a reduction seems to be specific to false memories induced by the study of associative lists, but is not observed when the studied lists are categorical in nature. These findings are interpreted as evidence that the left ATL functions as an integration hub that is crucial for the binding of semantic information into coherent representations of concepts. In order to investigate whether the right ATL might also contribute to semantic integration in the processing of verbal associative material, a follow-up tDCS study was conducted with the stimulation at study lateralized on the right ATL. A sample of 75 undergraduate students participated in an experiment in which they studied 8 associative lists and 8 categorical lists. One third of the participants studied all their word lists under anodal stimulation, another third studied under cathodal stimulation and the other third under sham stimulation. Results showed that stimulation of the right ATL by tDCS does not modulate false recognition for either association-related critical words or category-related critical words. These results provide preliminary support to views positing asymmetric connectivity between the anterior temporal lobes and the semantic representational network, and provide evidence for understanding bilateral brain dynamics and the nature of semantically induced memory distortions.

Keywords: false memory, DRM paradigm, right anterior temporal lobe, semantic memory, brain stimulation

INTRODUCTION

Research on memory distortions using the Deese, Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger and McDermott, 1995) has consistently shown that presenting a list of words associated with a critical word not presented for study produces high levels of false recall and false recognition of that critical word (Gallo, 2006, 2010). There is strong evidence of a relationship between the memory illusion typically obtained with the DRM paradigm and aspects of semantic representation and processing (Gallo, 2010; Roediger and Gallo, 2016). A number of studies have demonstrated that this kind of memory distortions in list-learning experiments is critically

modulated by the nature of lexical, semantic or structural similarity between to-be-remembered items and their related critical words (see Coane et al., 2021 for a recent review). And it has also been shown that many experimental manipulations that favor the processing of semantic characteristics of the studied words (e.g., meaning-oriented processing, relational processing, presentation of the material in meaning-consistent blocks, etc.) can cause an increase in false memories (Tussing and Greene, 1997; Thapar and McDermott, 2001; McCabe et al., 2004). Convergently, patient studies describe false memory effects that are modulated by the involvement of damaged semantic brain networks. As an example, patients with semantic dementia or fronto-temporal dementia, characterized by damage to the anterior temporal lobe (ATL), tend to show a reduction in false memories with DRM or similar tasks that require the construction of the general meaning or representation of the “gist” that summarizes the semantic characteristics common to studied list items (Simons et al., 2005; de Boysson et al., 2011).

In line with these findings, experiments with neuroimaging and non-invasive brain stimulation techniques in brain-intact participants have also shown the involvement of the ATL in the formation and modulation of false memory production with the DRM paradigm. Using fMRI, Chadwick et al. (2016) showed that the pattern of activation in the ATL while reading DRM lists predicted false recognition of the critical words associated to those lists. Going beyond correlational evidence, Gallate et al. (2009) found that altering the normal functioning of the left ATL using low-frequency repetitive transcranial magnetic stimulation (rTMS) reduced the probability of false recognition of the critical word without altering correct recognition. Consistently, Boggio et al. (2009) found a similar decrease in false recognition with anodal stimulation using transcranial direct current stimulation (tDCS) on the ATL, again with no stimulation effects on veridical memories. And more recently, Díez et al. (2017) found that the involvement of the ATL in this type of memory illusions depended on the kind of semantic relationship between the words in the list and the unstudied critical word. In their study, they applied transcranial direct current stimulation (anodal/cathodal/sham) in the left ATL and manipulated the type of semantic relationship (associative vs. categorical) between the words in the list and the critical items. The results of this study showed a significant reduction in false recognition with anodal stimulation in the left ATL, but only for those lists that had an associative relationship with the corresponding critical word. Although including only a small subset of the available evidence, the previous examples implicate that associative false memories are byproducts of relatively high-order semantic processes and that the ATL is a critical brain area for the representation of conceptual meaning.

One way in which the involvement of the ATL in the production of false memories can be more thoroughly understood is proposed by the “hub-and-spoke” model, a theoretical view that assumes that experiences (verbal and non-verbal) provide the basis for the formation of concepts and that this source of information is encoded in modality-specific areas distributed throughout the brain (the spokes). The model also assumes intermodal interactions for that specific information,

mediated by a transmodal hub located in the ATL (Patterson et al., 2007; Lambon Ralph and Patterson, 2008; Lambon Ralph et al., 2010; Binney and Lambon Ralph, 2015; Patterson and Lambon Ralph, 2016). The model has more recently been enriched by the addition of proposal for a semantic control network and its brain correlates (Lambon Ralph et al., 2017; Chiou et al., 2018). In this framework, the anterior temporal region of both hemispheres would function as an integration hub, specialized in integrating modality-specific information from distributed brain areas to form coherent conceptual representations (Wong and Gallate, 2012; Bonner and Price, 2013; Lambon Ralph, 2014; Lambon Ralph et al., 2017).

The hub-and-spoke model, as developed to this moment, has been rather successful in accounting for a wide range of empirical findings, involving both healthy participants and brain-compromised patients. And it has also been formally validated in computational simulations (Hoffman et al., 2018). However, further evidence-based specification is needed regarding some particular aspects, such as the extent to which structures and networks in both sides of the brain play equivalent roles in the representation of semantic cognition. And along these lines, a question remains as to whether the left and the right ATLs have the same representational functions or contribute similarly to conceptual processing. There is sufficient clinical and experimental evidence to support a bilateral involvement of the ATL in semantic processing. What is not so clear, however, is whether both structures are as symmetric in terms of semantic processing as initially assumed. An alternative position is that there is hemispheric specialization of the ATL, with the left side specialized in verbal semantic representation and the right side specializing in non-verbal semantic representation (Gainotti, 2011, 2012). Indeed, data from several studies suggest that semantic impairment could be modality-specific in the early stages of the disease, with significant asymmetries between the left and right ATLs. In these cases, a more atrophic left ATL tends to have effects on lexical-semantic knowledge, while an atrophy in the right ATL tends to affect pictorial representations (Snowden et al., 2004). The hypothesis that the ATL in both hemispheres is asymmetric in terms of semantic processing is also supported by the conclusions of a large meta-analysis of neuroimaging studies (Rice et al., 2015b), with the data pointing toward a more lateralized left ATL involvement in semantic tasks that required the processing of verbal stimuli (Rice et al., 2015a).

In an attempt to provide further evidence, the present study examined the role played by the right ATL in the conceptual processing manifested in the production of false memories upon studying word lists of semantically related items. Such memory distortions are, in large part, a consequence of higher-order semantic processing, the kind of processing in which the ATL is purportedly involved. As mentioned above, this has been shown in prior studies in which modulating neural activity in the left ATL via non-invasive stimulation caused a reduction of false recognition (Boggio et al., 2009; Gallate et al., 2009), with the reduction particularly affecting false recognition of items that had an associative relationship with the studied material (Díez et al., 2017). Following this rationale, in the present study we aimed to modulate activity in the right ATL by using tDCS.

tDCS involves the delivery of a low-level intensity current by a battery-driven stimulator. The conventional procedure requires two electrodes (anode and cathode) with at least one of them being placed on the scalp. The current passes from anode to cathode and this current has been shown to modulate the neurons' electrical activity. While this current is not sufficient to induce action potentials, research has revealed that tDCS may change the response threshold of the reached neurons (Bindman et al., 1964; Brunoni et al., 2011). Specifically, and based on findings derived from research that mainly focused on motor cortices (i.e., Nitsche and Paulus, 2001), it is usually stated that anodal tDCS increases neuronal excitability (by depolarization), and that cathodal tDCS decreases neuronal excitability (by hyperpolarization) (Nitsche and Paulus, 2000; Cambiaghi et al., 2010). Hence, and because anodal stimulation is sometimes associated with enhanced performance (i.e., Cerruti and Schlaug, 2009; Tanaka et al., 2009) and cathodal stimulation is sometimes associated with worse performance (Stagg et al., 2011; Young et al., 2013), it is frequently stated that anodal tDCS leads to facilitate brain functions whereas cathodal tDCS disrupts them (see Fertonani and Miniussi, 2017 for a critical view). However, and when these polarity-dependent effects are frequently reported, evidence accumulates to show that such effects are far from being straightforward both at the neurophysiological (i.e., Antal et al., 2007; Tanaka et al., 2020) or the behavioral level (i.e., Gómez-Ariza et al., 2017; Friedrich and Beste, 2018), with anodal tDCS sometimes giving rise to performance that is compatible with the disruption of brain functions (i.e., King et al., 2020) and cathodal tDCS sometimes producing enhanced performance (i.e., Brückner and Kammer, 2017). Despite this, and when the specific action mechanisms underlying the possible behavioral effects of tDCS in humans remain largely unknown and are thought to depend on a number of factors (i.e., brain activity prior to stimulation, current intensity, targeted brain area/network), tDCS is considered a useful technique to better understand the neural substrates of cognition (Berryhill et al., 2014; Filmer et al., 2014; Bestmann et al., 2015).

The number of tDCS studies on (long-term) memory has increased over the years, even though the variability of stimulation protocols (i.e., electrode montages, duration...), goals (i.e., applied vs. basic research), employed memory tasks (i.e., associative vs. item memory) and memory-related processes (i.e., encoding vs. retrieval) is considerable (for a systematic review and meta-analysis on episodic memory, see Galli et al., 2019). Many of these studies seek to enhance performance by stimulating specific brain areas/networks thought to play a pivotal role in either encoding or retrieval processes. In other cases, studies using tDCS aim to test theoretically-guided hypothesis on the involvement of certain brain regions in memory processes or representations. Thus, for example, Leach et al. (2019) showed, in younger adults, that anodal stimulation over the left dorsolateral prefrontal cortex during a face-name encoding task improved associative memory. And Bjekić et al. (2019) found that anodal tDCS over either the left or the right posterior parietal cortex led to better performance on two different associative memory tasks. Interestingly, tDCS has

also been used to dissociate the role of distinct memory-related brain areas. Pisoni et al. (2015), for example, showed that while tDCS over the left temporal cortex modulated recognition for studied items, stimulation over the right parietal cortex allowed participants to better identify new items (for other examples of dissociations see Pergolizzi and Chua, 2016, or Smirni et al., 2015).

Of special relevance here, tDCS has now been shown to be effective in modulating neural activity associated with representational aspects of semantic processing, sometimes contributing to hemispheric dissociations. Thus, for example, relative to sham or cathodal stimulation, anodal tDCS of the posterior superior temporal gyrus, which subsumes Wernicke's area, has been shown to lead participants to come up with associates that are more representative of the basic-level category of a presented image that worked as a cue. Similarly, anodal tDCS over the same temporal subregion was found to speed up the identification of meaningful word pairs, but not non-meaningful ones (Price et al., 2016). Interestingly, stimulation of the homologous region in the right hemisphere made participants faster at judging whether two words were semantically related by a subordinate meaning (Peretz and Lavidor, 2013). Hence, it would seem that the effects that tDCS over the temporal lobe have on semantic associations are hemisphere specific. Stimulation of the right hemisphere would seem to modulate semantic processing of subordinate and indirect associations, whereas tDCS of the left temporal lobe would modulate more semantically related concepts. Moreover, some tDCS studies have revealed laterality-dependent memory improvements, with memory for visuospatial information being modulated with right temporoparietal stimulation and memory for verbal information being modulated with left temporoparietal tDCS (i.e., Fiori et al., 2017; Antonenko et al., 2018).

Hence, we aimed to explore if the right ATL has an equivalent role to its left homologous in the production of semantic-based memory errors by modulating its neural activity via tDCS. If this was the case, tDCS over the right ATL should result in changes in the production of false memories, particularly of the associative kind. In order to test this prediction, and closely following the design employed by Díez et al. (2017), tDCS (both anodal and cathodal) was delivered over the right ATL to evaluate its effects on false memory with the DRM paradigm, using lists of words that maintained either associative or categorical relationships with their unstudied critical words. If the ATL of the two hemispheres had the same functionality in terms of semantic processing, tDCS of the right ATL should lead to a reduction in false recognition of associative critical words, without affecting either true recognition or false recognition in categorical lists. If, on the other hand, the functionality of the ATL is not equivalent in the two hemispheres, such a pattern of results should not be found following anodal stimulation of the right ATL. Because of the scarcity of studies combining DRM and non-invasive stimulation, and also because of our limited current knowledge on the neurophysiological effects of tDCS when applied outside sensory/motor cortices (see above), we were reluctant to make specific *a priori* predictions regarding type of lists and polarity effects for the case that the right ATL were

actually different in semantic functionality from the left ATL. However, the results of the experiment could still be relevant to further understand bilateral brain dynamics and the nature of semantically-driven memory distortions. With this last goal in mind, the design of the experiment was not only aimed toward a standard quantitative assessment of true and false memory performance in the different stimulation conditions, but it was also supplemented to characterize the subjective feelings of recollection and familiarity in their recognition responses and their possible dependence on the role played by the right ATL. To this end, the remember/know (R/K) testing paradigm originally devised by Tulving (1985) was implemented and included in the final memory test. In a study by Pergolizzi and Chua (2015), bilateral tDCS on the parietal cortex failed to show any effect on R/K responses, but whether subjective determinants of recognition are to some degree affected by stimulation-induced changes in the functioning of the ATL is an unexplored question.

METHODS

Participants

A total of 78 undergraduates were recruited from the student population at the University of La Laguna, Spain. All were native speakers of Spanish, with normal or corrected-to-normal vision, and they all gave written informed consent for their participation in the study. They received course credit as a basic compensation. Their ages ranged from 18 to 30 years ($M = 19.98$; $SD = 2.72$), and 82% were female.

The sample size was determined in advance to be at least the double of the most similar tDCS studies with significant reported effects (e.g., Boggio et al., 2009; $N = 10$ subjects by stimulation condition) and also considering the range of those used in standard DRM experiments, and the results of a power analysis performed with G*Power (Faul et al., 2007). We estimated the sample size to obtain an effect size similar to that obtained in the Díez et al. (2017) paper for the stimulation group $\times$ list type interaction ($\eta_p^2 = 0.10$, F effect size of 0.33, i.e., a medium effect size) in a repeated measures within-between interaction (2 -categorical and associative- and 3 -stimulation groups-, respectively) and the results revealed that to achieve 0.80 power we needed a minimal sample size of 27 subjects.

All participants were right-handed, according to the Edinburgh Handedness Inventory (Oldfield, 1971), and they were randomly assigned to one of three experimental conditions. Following standard procedures in experiments involving tDCS, the participants were screened and excluded if they reported any psychiatric, psychological or neurological disorder or if they reported brain injuries, migraines, epileptic seizures or family history of epilepsy. The institutional ethical committee of the University of La Laguna approved the protocol, and the study was conducted in compliance with the Declaration of Helsinki (World Medical Association, 2013). Data from three participants were excluded from the analysis for failing to meet an accuracy criterion set for the recognition task¹. The results obtained

using the data from the remaining 75 participants (25 in each stimulation condition: anodal, cathodal and sham; 20, 22, and 22 females, respectively) are presented in this report.

tDCS

Stimulation was delivered by a battery-driven electrical stimulator (TCT Research Ltd.) with an intensity of 2 mA. Following the stimulation protocol used by Díez et al. (2017), the current was transferred by two 5×7 cm rubber electrodes covered with saline-soaked sponges. For anodal stimulation, the anode was placed over site FT10 (BA 38/20), according to the International 10-10 System for EEG electrode placement, and the cathode was placed on the contralateral shoulder. For the cathodal stimulation, the cathode was placed over site FT10 and the anode was placed over the contralateral shoulder. Site FT10 is considered the closest electrode location to the right ATL (Acharya et al., 2016). Stimulation was applied for 20 min in both the anodal and cathodal conditions, using 10-s fade in/out ramps. For the sham stimulation condition, the electrodes were placed in the same positions as in the active stimulation conditions, with current ceasing to be applied after 60 s of stimulation. All participants completed the session without major complains or discomfort.

Stimuli

The stimuli were the same previously used to successfully induce false memories in the tDCS study by Díez et al. (2017) and consisted of 24 critical words (CW) each related with two separate word lists: one associative list and one categorical list. The associative list was constructed selecting the first 10 associates of the CW on the basis of their backward associative strength (BAS), obtained from Spanish free-association norms (Fernandez et al., 2004, 2014). The categorical list was built selecting 10 words belonging to the same category as the CW, according to normative data in Spanish (Marful et al., 2015). Thus, for the CW *book*, the associative list consisted of the words *author*, *foreword*, *chapter*, *page*, *volume*, *edition*, *reading*, *read*, *epilogue*, and *reader*. For the same CW, the words in the categorical list were *magazine*, *newspaper*, *novel*, *encyclopedia*, *article*, *story*, *comic*, *notes*, *notebook*, and *dictionary*.

Sixteen word lists (8 associative and 8 categorical) were presented to each subject. The remaining 8 CWs and their corresponding lists served as control CWs and distractors in the recognition test. A counterbalanced assignment of lists to subjects was used to ensure that all word lists were displayed in all the different study conditions.

Procedure

The experiment took place in a quiet laboratory with only one participant per session. First, participants filled out a personal data sheet and a screening questionnaire about medical and psychological conditions, and they also signed an informed consent form. As in Díez et al. (2017), after electrode placement,

bias in the total sample were excluded from the analyses. Following Díez et al. (2017), this screening procedure was aimed at excluding participants with extreme response patterns, as evidenced in their recognition responses (hit rates and false alarms) to the items.

¹ A non-parametric measure of response bias was calculated for each subject, and all the data from participants scoring above or below 1.96 SDs from the average

and coincident with the time in which the stimulation was started, participants were asked to perform a pen and paper visual-search task for an idle time of 7 min, consisting in circling with a pen the letters “n,” “p,” and “c” in words of a text written in an unfamiliar language. We decided to have participants engage in a specific task to minimize variability in brain/cognitive activity during stimulation. Because the encoding phase lasted about 8.5 min, and taking into account the time needed for reading the instructions, stimulation (20 min) started before encoding. This type of stimulation (partially offline partly online, in this case online during encoding) tends to show larger effects than entirely offline before encoding (see Galli et al., 2019).

When the visual-search task ended, the participants received the experimental instructions on a computer screen. These instructions and all subsequent tasks were displayed and controlled by a computer running E-Prime 3.0 software (Psychology Software Tools, 2016). The participants were informed that they would listen to a series of 16 lists of words, and that following the presentation of all the lists they would have to work on a set of arithmetical problems and to perform a final memory test on the words previously presented in the lists. See **Figure 1** for a schematic representation of the general procedure.

Following a standard DRM procedure, the words within each list were always presented in the same order, from higher to lower frequency in categorical production or BAS values. Words were presented aurally, one every 2 s. For each participant, a pseudorandom ordering of the 16-list sequence resulted in two subsets of 8 randomly distributed lists (4 categorical and 4 associative), with the last list followed by an on-screen distracting task that lasted 2 min. This task consisted of a series of three-term arithmetical problems presented with a solution that had to be checked for accuracy and required a yes/no response on the computer keyboard. After this task, stimulation was terminated and the participants performed a yes/no recognition memory test. The recognition test included a total of 64 words: the 16 CWs from studied lists (8 associative and 8 categorical), the 8 CWs from non-presented lists (control CWs), 32 studied words (words in position 2 and 7 in the studied lists), and 8 distracting non-presented words (words in position 2 in the non-presented lists). The words in the recognition test were displayed one by one on the center of the computer screen, preceded by a fixation point which lasted 750 ms, and were randomly presented for each participant. The participants were instructed to respond using the keyboard, indicating for each word whether they recognized it from the studied lists (old) or they thought it was a not studied (new) word. If the answer was “yes,” a remember/know judgement was required. The remember/know instructions were adapted from Guillory and Geraci (2010), who took them from Rajaram (1993). After the memory test was completed, all participants completed a questionnaire on tDCS adverse effects (Brunoni et al., 2011). Specifically, they were asked to report the degree to which they experienced a list of side-effects (Headache, Neck pain, Scalp pain, Tingling, Stinging/itching, Burning sensation, Skin redness, Drowsiness, Concentration problems, or Severe mood changes). None of them reported major complaints or discomfort associated with stimulation and, as shown in **Table 1**, all group means ranged

from 1 to 1.92 (i.e., absent to mild). Only skin redness was found to be significantly greater for participants in the Anode and Cathode groups relative to those in the Sham group. The experiment finished with a short debriefing message explaining the experiment and with the request to not reveal details of the session to other students in the participant pool.

Design

A mixed factorial design $3 \times 2 \times 4$ was used, with type of stimulation (anodal, cathodal or sham) as a between factor, and type of list (associative or categorical) and type of word (studied, CW, control CW or distractor) as within factors. The dependent variables were derived from the recognition responses to each type of word and from the remember/know judgments.

Data Analysis

Data on hit rates and critical false alarm rates were analyzed with mixed design repeated-measures ANOVAs, using η_p^2 as the effect size measure and reporting the corresponding 90% confidence intervals. *Post-hoc* comparisons were performed by way of Tukey's Honest Significant Difference tests. The standard ANOVAs were complemented with Bayesian repeated measures ANOVA analyses. Default priors were used with equal assignment of prior model probability across all models.

Non-parametric signal detection theory measures were used because of the impossibility to test parametric assumptions with yes/no recognition tasks, especially the equality of the signal and the noise standard deviations, and because in DRM tasks it is very common that some subjects have hit or false alarm rates of 1 or 0 (Donaldson, 1996; Stanislaw and Todorov, 1999). The formulas proposed by Zhang and Mueller (2005) were used to calculate non-parametric sensitivity (A) and bias (b). Since $b = 1$ denotes absence of bias, a logarithmic transformation was applied to convert the variable into a symmetrical one, with a value of 0 denoting absence of bias and negative values denoting liberal criteria and positive values denoting conservative criteria.

All statistical analysis was performed using R Statistical Software (version 4.0.2; R Core Team, 2020) and *jamovi* computer software (The jamovi project, 2021).

RESULTS

Table 2 shows recognition rates for all the experimental conditions and word types. Overall, across conditions, a strong false recognition effect was evidenced by the high recognition rates of CWs in comparison to those of other non-presented distractor words.

True Recognition

Data on hit rates (i.e., “yes” responses to studied words) were analyzed with a mixed design repeated-measures ANOVA, with tDCS (Anodal vs. Cathodal vs. Sham) as a between participants variable and with type of list (Associative vs. Categorical) as a within participants variable. There was a significant main effect of type of list, $F_{(1,72)} = 20.54$, $MSE = 0.01$ $p < 0.001$, $\eta_p^2 = 0.22$, 90% CI (0.09, 0.35). On average, the proportion of correct recognition was higher for categorical lists ($M = 0.77$; $SD =$

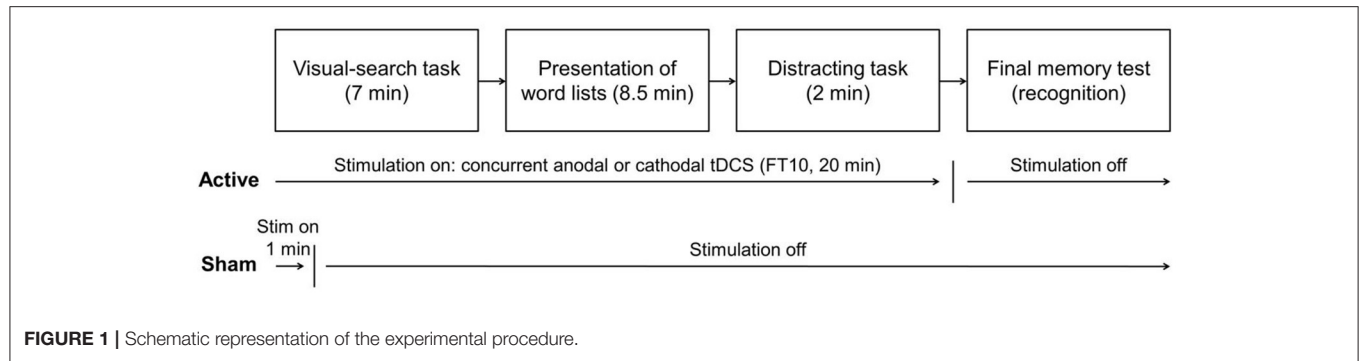


TABLE 1 | Descriptive statistics (mean and standard deviations) of the responses to the questionnaire on tDCS adverse effects.

	Anodal	Cathodal	Sham	χ^2	*p	Pairwise comparisons
Headache	1.28 (0.54)	1.31 (0.68)	1.37 (0.56)	0.76	0.684	
Neck pain	1.32 (0.63)	1.46 (0.81)	1.11 (0.32)	3.46	0.178	
Scalp pain	1.17 (0.48)	1.08 (0.27)	1.00 (0)	3.36	0.187	
Tingling	1.88 (0.80)	1.92 (0.84)	1.89 (0.80)	0.04	0.981	
Stinging/itching	1.68 (0.80)	1.92 (0.84)	1.67 (0.83)	1.81	0.405	
Burning sensation	1.52 (0.77)	1.58 (0.76)	1.41 (0.75)	0.98	0.612	
Reddening of the skin	1.68 (0.63)	1.84 (0.73)	1.11 (0.32)	19.28	<0.001	A > S; C > S
Drowsiness	1.52 (0.77)	1.54 (0.86)	1.37 (0.62)	0.47	0.791	
Concentration problems	1.68 (0.85)	1.61 (0.70)	1.70 (0.67)	0.34	0.842	
Severe mood swings	1.04 (0.20)	1.15 (0.46)	1.00 (0.00)	3.71	0.156	

Participants rated the side-effects on a scale ranging from 1 to 4 (1: absent; 2: mild; 3: moderate; 4: severe).

*p-values from Kruskal-Wallis tests and Dwass-Steel-Critchlow-Fligner pairwise comparisons for significant effects.

TABLE 2 | Mean recognition results (standard deviation) as a function of type of list and type of stimulation.

	Associative			Categorical		
	Anodal	Cathodal	Sham	Anodal	Cathodal	Sham
Studied words						
True recognition	0.66 (0.18)	0.72 (0.12)	0.68 (0.17)	0.77 (0.13)	0.78 (0.14)	0.75(0.13)
Sensitivity (A)	0.88 (0.10)	0.90 (0.07)	0.88 (0.09)	0.91 (0.06)	0.89 (0.13)	0.89 (0.10)
Bias [log(b)]	0.63 (0.60)	0.54 (0.43)	0.67 (0.37)	0.41(0.46)	0.31 (0.49)	0.51 (0.36)
Remember	0.46 (0.18)	0.46 (0.17)	0.47 (0.19)	0.58 (0.19)	0.53 (0.18)	0.51 (0.18)
Know	0.19 (0.15)	0.26 (0.16)	0.21 (0.14)	0.19 (0.17)	0.25 (0.11)	0.24 (0.15)
Critical words						
False recognition	0.53 (0.24)	0.42 (0.33)	0.53 (0.23)	0.57 (0.25)	0.54 (0.24)	0.56 (0.20)
Sensitivity (A)	0.78 (0.15)	0.76 (0.16)	0.82 (0.12)	0.84 (0.10)	0.81 (0.14)	0.84(0.12)
Bias [log(b)]	0.59 (0.63)	0.73 (0.78)	0.76 (0.55)	0.65 (0.65)	0.70 (0.62)	0.74 (0.44)
Remember	0.32 (0.21)	0.19 (0.20)	0.26 (0.13)	0.28 (0.21)	0.22(0.19)	0.26(0.18)
Know	0.19 (0.15)	0.23 (0.18)	0.27 (0.20)	0.28(0.21)	0.31 (0.18)	0.29 (0.25)
Distractors						
False alarms	0.08 (0.19)	0.07 (0.13)	0.06 (0.11)	0.07 (0.11)	0.11 (0.19)	0.07 (0.13)
Critical control	0.15 (0.16)	0.15 (0.20)	0.09 (0.12)	0.11 (0.18)	0.12 (0.23)	0.07 (0.11)

0.13) than for associative lists ($M = 0.68$; $SD = 0.16$). No other effects were statistically significant (both tDCS condition and interaction effects with $F < 1$), showing that the type of list effect was not modulated by tDCS (see **Figure 2**).

The standard ANOVA was complemented with Bayesian repeated measures ANOVA analyses, conducted with *jamovi* computer software (The jamovi project, 2021). Default priors were used with equal assignment of prior model probability across all models. The results showed extreme evidence for the type of list effect ($BF_{10} = 922.462$), and substantial evidence for H_0 compared to H_1 in the stimulation condition ($BF_{01} = 5.25$) and interaction effects ($BF_{01} = 4.44$).

The 3×2 ANOVA on A sensitivity scores did not reveal a significant effect of stimulation condition ($F < 1$; $BF_{01} = 8.60$), type of list, $F_{(1,72)} = 1.01$; $p = 0.32$; $BF_{01} = 3.42$, or interaction, $F_{(2,72)} < 1$; $BF_{01} = 4.17$. The analysis on $\log(b)$ revealed a statistically significant effect of type of list, $F_{(1,72)} = 11.63$, $MSE = 0.14$, $p = 0.001$, $\eta_p^2 = 0.14$, 90% CI (0.04, 0.26), $BF_{10} = 31.39$, with associative lists ($M = 0.62$) showing a higher conservative response bias [$\log(b)$] than categorical lists ($M = 0.41$). No other effects reached statistical significance, neither tDCS condition [$F_{(2,72)} = 1.16$, $p = 0.32$; $BF_{01} = 3.67$ and] nor interaction effects ($F < 1$; $BF_{01} = 7.82$).

The analysis on Remember/Know responses² revealed a significant effect for “remember” responses in type of list, $F_{(1,72)} = 12.82$, $MSE = 0.02$, $p = 0.001$, $\eta_p^2 = 0.15$, 90% CI (0.04, 0.27), $BF_{10} = 40.96$, with categorical lists ($M = 0.54$) showing more “remember” responses than associative lists ($M = 0.46$). No significant effects were observed for stimulation condition ($F < 1$, $BF_{01} = 5.87$) or for the interaction [$F_{(2,72)} = 1.12$; $p = 0.33$, $BF_{01} = 4.45$].

For “know” responses to true recognized words no significant effects were observed as a function of stimulation condition, $F_{(2,72)} = 1.6$; $p = 0.19$, $BF_{01} = 6.39$, type of list ($F < 1$, $BF_{01} = 0.38$) or the interaction ($F < 1$, $BF_{01} = 4.87$).

False Recognition

A 3 (tDCS condition: Anodal vs. Cathodal vs. Sham) $\times$ 2 (type of list: Associative vs. Categorical) mixed ANOVA on the false recognition rates (i.e., “yes” responses to CWs) showed a statistically significant effect of type of list, $F_{(1,72)} = 4.64$, $MSE = 0.03$, $p = 0.035$, $\eta_p^2 = 0.06$, 90% CI (0.002, 0.16), $BF_{10} = 1.37$. On average, false recognition was higher for categorical lists ($M = 0.56$; $SD = 0.23$) than for the associative lists ($M = 0.50$; $SD = 0.27$). No other source of variability reached statistical significance (both tDCS condition and interaction with $F < 1$, $BF_{01} = 3.90$ and $BF_{01} = 4.27$, respectively), which provides substantial evidence of the absence of an effect of tDCS over false recognition (see **Figure 3**).

The sensitivity analyses showed a significant effect of type of list, $F_{(1,72)} = 4.23$, $MSE = 0.02$, $p = 0.04$, $\eta_p^2 = 0.06$, 90% CI (0.0009, 0.16), $BF_{10} = 1.42$, with categorical lists ($M = 0.83$)

showing a higher sensitivity rate (A) than associative lists ($M = 0.79$). No significant effects were found for the type of stimulation condition, $F_{(2,72)} = 1.19$, $MSE = 0.03$, $p = 0.31$, $\eta_p^2 = 0.03$, $BF_{01} = 4.48$, or the interaction ($F < 1$, $BF_{01} = 6.57$). The analyses on $\log(b)$ did not reveal any statistically significant effect ($F < 1$, BF_{01} values of 5.71 for list type, 4.38 for stimulation and 7.62 for the interaction).

Finally, the proportion of Remember/Know judgments were calculated for false memories. There were no significant effects observed in “remember” responses for type of list ($F < 1$, $BF_{01} = 5.67$), type of stimulation, $F_{(2,72)} = 2.47$, $p = 0.09$, $BF_{01} = 1.42$, or interaction ($F < 1$, $BF_{01} = 6.19$). A significant effect was observed in “know” responses as a function of type of list, $F_{(1,72)} = 5.87$, $MSE = 0.03$, $p = 0.018$, $\eta_p^2 = 0.075$, 90% CI (0.0071, 0.18), $BF_{10} = 2.58$. On average, the categorical lists showed more know responses ($M = 0.29$) than associative lists ($M = 0.23$). No significant effects were found in know responses for the stimulation condition ($F < 1$, $BF_{01} = 6.43$) or the interaction ($F < 1$, $BF_{01} = 5.22$).

DISCUSSION

The present experiment aimed to examine the extent to which the right ATL played the same role as the left ATL in semantic processing leading to the generation of false memories; more specifically, it employed a standard DRM paradigm and the application of tDCS to examine the involvement of the right ATL in the production of verbal false memories after studying word lists that, either associatively or categorically, were semantically related to unrepresented critical words. On the basis of previous findings by Díez et al. (2017), it was predicted that if the ATL of both hemispheres contributed similarly to semantic processing, a decrease in false recognition in associative lists would be expected following stimulation of the right ATL.

Consistently with previous findings in the literature (Boggio et al., 2009; Gallate et al., 2009; Díez et al., 2017), true recognition was higher for categorical than for associative lists, and was unaffected by ATL stimulation. More relevant for the goal of the experiment, modulating neural activity of the right ATL using tDCS did not modify the elicitation of false memories, with Bayesian analyses showing substantial evidence in favor of the null hypothesis. Neither cathodal nor anodal stimulation altered the rates of false recognition as compared to sham stimulation, and that was true for both associative and categorical lists. A higher overall error rate for categorical lists was the only significant result, small in magnitude, and at odds with most of the reported findings in the false memory literature, with the exception of experiments in which there is a feature or thematic overlap between studied items and critical words that adds to existing associative links (e.g., Coane et al., 2016). In sum, tDCS of the right ATL failed to show effects that were comparable to previously demonstrated effects when the left ATL was stimulated. Pending replication and extension, this finding provides preliminary evidence for an asymmetrical view of the role of the temporal cortex in semantic processing

²Remember/know scores were calculated as a proportion of the total possible responses in each condition. Analyses were also performed by calculating them as a proportion of the “yes” responses. These analyses showed no significant effects of tDCS, list type or interaction for either correct or false recognition.

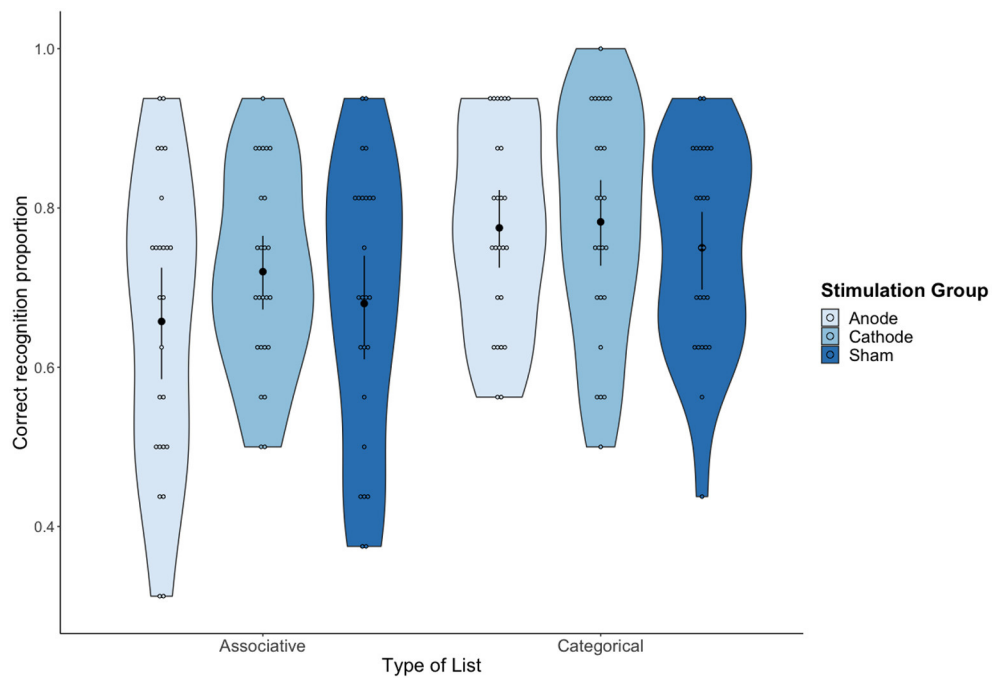


FIGURE 2 | Mean proportion of correct recognition as a function of tDCS condition and Type of list. Error bars represent 95% confidence intervals (CIs).

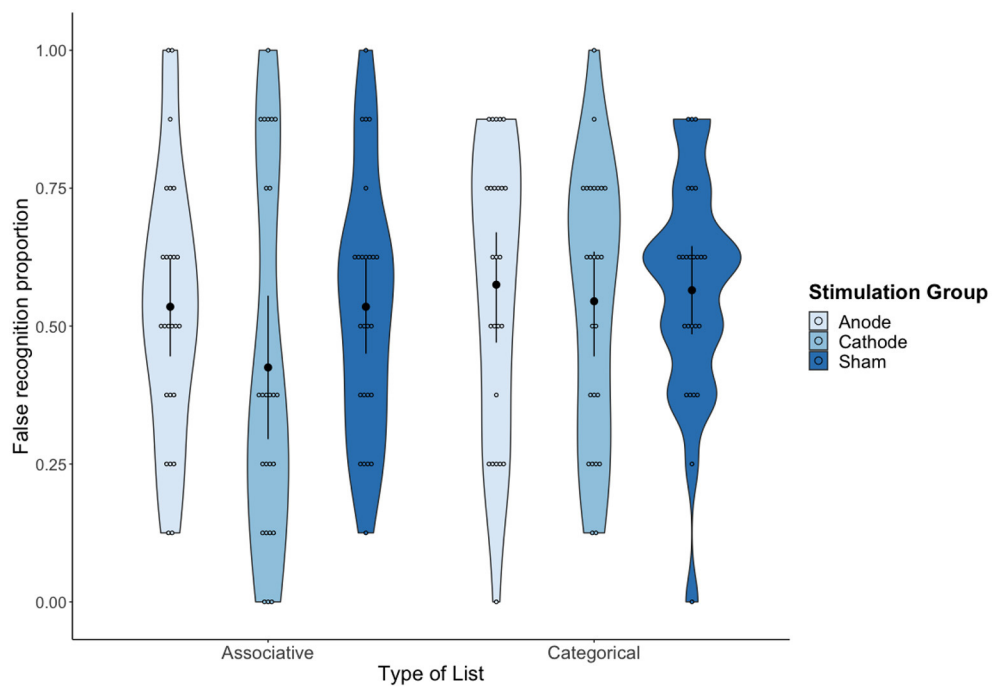


FIGURE 3 | Mean proportion of false recognition as a function of tDCS condition and Type of list. Error bars represent 95% confidence intervals (CIs).

when it comes to producing semantic memory illusions with the DRM procedure.

Studies focusing on how false memory production is affected by altered brain function (either as a consequence of brain damage or as the result of non-invasive stimulation in healthy participants) are not abundant, and even more scarce are studies focusing on the potential role of the ATL in the modulation of memory distortions. And with regard to the specific manipulation aimed at selectively modulating the activity of the right ATL with non-invasive brain stimulations techniques, the present study is, to the best of our knowledge, the first and only one. Still, the lack of evidence for hemispheric symmetry in the pattern of results is in line with recent findings in other related studies employing different methods. Thus, in a meta-analysis of fMRI studies of false memories, Kurkela and Dennis (2016) concluded that the involvement of the temporal cortex in the kind of semantic encoding that leads to false memory was lateralized to the left. More recent evidence for this kind of differential involvement of the two ATLs in verbal and non-verbal semantic processing has been provided by Woollams and Patterson (2018), working with a large sample of semantic dementia patients with lateralized lesions, and by Rice et al. (2018) in their study of a group of postsurgical temporal lobe epilepsy patients, with either left or right anterior temporal lobectomy. In sum, and regardless of which ultimately be the most likely explanation, the data from several years of studies of patients with compromised ATLs are consistent with the idea that the left ATL is prevalently involved in verbal aspects of conceptual processing, while the right ATL is more implicated in pictorial or non-verbal aspects (Gainotti, 2020).

It is also worth noting that the phenomenological experience of participants, as reflected in their remember/know judgments to true and false recognition responses, was not affected by tDCS, echoing prior null findings with stimulation of the parietal cortex (Pergolizzi and Chua, 2015). This finding suggests that the right ATL is not involved in the evaluative processes accompanying recognition decisions for the studied materials. While further systematic analyses are obviously needed, it is interesting that the left ATL has been reported to be involved in the familiarity judgments for verbal materials (e.g., Köhler and Martin, 2020). Given that in the present experiment (as in most other published reports with the DRM procedure) familiarity is as frequently involved as recollection in the production of false memories, additional evidence for the lateralization of familiarity, with a variety of procedures and materials, has the potential to contribute to a better understanding of recognition mechanisms in general and false recognition in particular.

Although the absence of stimulation-related modulation of false memories in the present study can be interpreted in terms of interhemispheric functional differences in the production of semantic memory illusions, other interpretations are also possible. The effects of tDCS on declarative memory are still poorly understood and difficult to replicate at times, most likely due to the multiple factors contributing to them (i.e., stimulation parameters, electrode montages, basal state-dependent neuromodulation, materials to be memorized). A recent meta-analysis of the effects of tDCS on episodic memory

revealed that some moderator variables should be considered (Galli et al., 2019). Thus, for example, recall tasks seem to be more sensitive to anodal tDCS than recognition tasks (with cathodal tDCS the tendency seems to be the opposite), especially when associative memory is involved (i.e., Fiori et al., 2011; Flöel et al., 2012; Matzen et al., 2015). In addition, stimulating frontal regions tends to produce larger effects than the stimulation of temporal areas (Galli et al., 2019). Hence, even when the present study embraced a tDCS protocol that proved to be effective at modulating false recognition with the left ATL as the target area (Díez et al., 2017), it could be entirely possible that this protocol (i.e., the intensity of the electric current that is necessary to change the response threshold of the stimulated neurons) is not suitable to change neural activity in the right ATL. The same asymmetries (in anatomy and connectivity; Barrick et al., 2007) potentially producing differentiable functions between the two temporal lobes (left medial temporal regions involved in processing of verbal/ information vs. right homologous regions specially recruited during visual/ processing; Dalton et al., 2016) could also give rise to differences in neuromodulation effectivity. Moreover, it could even be the case that having both ATLs similar functional properties regarding semantic processing, the ability of tDCS to modify their functionality could be different in both hemispheres. Related to this, a recent HD-tDCS study found that stimulating BA22 in the right hemisphere (a site that is slightly more posterior than the target area in our study: BA38/20) modulated insight problem-solving (thought to require semantic integration) relative to sham and left frontopolar stimulation (Salvi et al., 2020). Hence, we recognize that an electrode montage different to the one used here should be considered in future studies. In addition, because false memory effects with DRM procedures are also robust when performance is assessed using recall tasks, future attempts to conceptually replicate the null effect of tDCS over the right ATL to modulate the production of semantic memory illusions should also consider memory tests of this kind that could be more sensitive to external modulations of neural activity.

Beyond the evidence on the asymmetrical involvement of the temporal cortex in semantic processing tasks, the results of the present experiment make a contribution to the quest for the neural correlates of activation and/or gist-formation processes (Roediger et al., 2001; Brainerd and Reyna, 2005) that result in false memory formation. And, when taken together with the findings that reveal a role for the left ATL in false recognition (e.g., Díez et al., 2017), constitute relevant evidence for the assumption that verbal false memories are, to a large extent, a consequence of higher-order semantic processing in the left lateral cortex. They also offer support to explanations of memory distortions by neuroscience-based semantic approaches that, like the hub-and-spoke model (Patterson and Lambon Ralph, 2016; Lambon Ralph et al., 2017), open the door to assume that such errors are critically related to integrative, conceptual processes taking place in the ATL. At the same time, these results point to the need for such models to be further expanded and replicated with independent samples, more inclusive in terms of characteristics such as gender or

handedness, and to pay closer attention to the specifics of the particular brain areas involved in the different processes and subprocesses considering the combined use of NIBS and neuroimaging techniques.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by CEIBA - Comité de Ética de la Investigación y Bienestar Animal (CEIBA2017-0265). The patients/participants

provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MA collected the data. ED analyzed the data. All authors contributed to the design of the experiment and contributed to writing the manuscript.

FUNDING

The work reported here was partially supported by the Spanish Ministry of Economy, Industry and Competitiveness (Grant PSI2017-82748-P), and by Junta de Castilla y León (Grant SA052G18).

REFERENCES

- Acharya, J. N., Hani, A. J., Thirumala, P. D., and Tsuchida, T. N. (2016). American clinical neurophysiology society guideline 3: a proposal for standard montages to be used in clinical EEG. *J. Clin. Neurophysiol.* 33, 312–316. doi: 10.1097/WNP.0000000000000317
- Antal, A., Terney, D., Poreisz, C., and Paulus, W. (2007). Towards unravelling task-related modulations of neuroplastic changes induced in the human motor cortex. *Eur. J. Neurosci.* 26, 2687–2691. doi: 10.1111/j.1460-9568.2007.05896.x
- Antonenko, D., Külzow, N., Sousa, A., Prehn, K., Grittner, U., and Flöel, A. (2018). Neuronal and behavioral effects of multi-day brain stimulation and memory training. *Neurobiol. Aging* 61, 245–254. doi: 10.1016/j.neurobiolaging.2017.09.017
- Barrick, T. R., Lawes, I. N., Mackay, C. E., and Clark, C. A. (2007). White matter pathway asymmetry underlies functional lateralization. *Cerebral Cortex* 17, 591–598. doi: 10.1093/cercor/bhk004
- Berryhill, M. E., Peterson, D. J., Jones, K. T., and Stephens, J. A. (2014). Hits and misses: leveraging tDCS to advance cognitive research. *Front. Psychol.* 5:800. doi: 10.3389/fpsyg.2014.00800
- Bestmann, S., de Berker, A. O., and Bonaiuto, J. (2015). Understanding the behavioural consequences of non-invasive brain stimulation. *Trends Cogn. Sci.* 19, 13–20. doi: 10.1016/j.tics.2014.10.003
- Bindman, L. J., Lippold, O. C., and Redfearn, J. W. (1964). The action of brief polarizing currents on the cerebral cortex of the rat (1) during current flow and (2) in the production of long-lasting after-effects. *J. Physiol.* 172, 369–382. doi: 10.1113/jphysiol.1964.sp007425
- Binney, R. J., and Lambon Ralph, M. A. (2015). Using a combination of fMRI and anterior temporal lobe rTMS to measure intrinsic and induced activation changes across the semantic cognition network. *Neuropsychologia* 76, 170–181. doi: 10.1016/j.neuropsychologia.2014.11.009
- Bjekić, J., Colić, M. V., Živanović, M., Milanović, S. D., and Filipović, S. R. (2019). Transcranial direct current stimulation (tDCS) over parietal cortex improves associative memory. *Neurobiol. Learn. Mem.* 157, 114–120. doi: 10.1016/j.nlm.2018.12.007
- Boggio, P. S., Fregni, F., Valasek, C., Ellwood, S., Chi, R., Gallate, J., et al. (2009). Temporal lobe cortical electrical stimulation during the encoding and retrieval phase reduces false memories. *PLoS ONE* 4:e4959. doi: 10.1371/journal.pone.0004959
- Bonner, M. F., and Price, A. R. (2013). Where is the anterior temporal lobe and what does it do? *J. Neurosci.* 33, 4213–4215. doi: 10.1523/JNEUROSCI.0041-13.2013
- Brainerd, C. J., and Reyna, V. F. (2005). *The Science of False Memory*. New York, NY: Oxford University Press.
- Brückner, S., and Kammer, T. (2017). Both anodal and cathodal transcranial direct current stimulation improves semantic processing. *Neuroscience* 343, 269–275. doi: 10.1016/j.neuroscience.2016.12.015
- Brunoni, A. R., Amadera, J., Berbel, B., Volz, M. S., Rizziero, B. G., and Fregni, F. (2011). A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. *Int. J. Neuropsychopharmacol.* 14, 1133–1145. doi: 10.1017/S1461145710001690
- Cambiaghi, M., Velikova, S., Gonzalez-Rosa, J. J., Cursi, M., Comi, G., and Leocani, L. (2010). Brain transcranial direct current stimulation modulates motor excitability in mice. *Eur. J. Neurosci.* 31, 704–709. doi: 10.1111/j.1460-9568.2010.07092.x
- Cerruti, C., and Schlaug, G. (2009). Anodal transcranial direct current stimulation of the prefrontal cortex enhances complex verbal associative thought. *J. Cogn. Neurosci.* 21, 1980–1987. doi: 10.1162/jocn.2008.21143
- Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., and Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proc. Nat. Acad. Sci.* 113, 10180–10185. doi: 10.1073/pnas.1610686113
- Chiou, R., Humphreys, G. F., Jung, J., and Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive ‘semantic control’ and hub-and-spoke ‘semantic representation’ systems. *Cortex* 103, 100–116. doi: 10.1016/j.cortex.2018.02.018
- Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., and Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: a review of how structural and conceptual similarity affect false memory. *Front. Psychol.* 12:668550. doi: 10.3389/fpsyg.2021.668550
- Coane, J. H., McBride, D. M., Termonen, M. L., and Cutting, J. C. (2016). Categorical and associative relations increase false memory relative to purely associative relations. *Mem. Cognit.* 44, 37–49. doi: 10.3758/s13421-015-0543-1
- Dalton, M. A., Hornberger, M., and Piguet, O. (2016). Material specific lateralization of medial temporal lobe function: an fMRI investigation. *Hum. Brain Mapp.* 37, 933–941. doi: 10.1002/hbm.23077
- de Boysson, C., Belleville, S., Phillips, N. A., Johns, E. K., Goupil, D., Souchay, C., et al. (2011). False recognition in Lewy-body disease and frontotemporal dementia. *Brain Cogn.* 75, 111–118. doi: 10.1016/j.bandc.2010.10.011
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Díez, E., Gómez-Ariza, C. J., Díez-Álamo, A. M., Alonso, M. A., and Fernandez, A. (2017). The processing of semantic relatedness in the brain: evidence from associative and categorical false recognition effects following transcranial direct current stimulation of the left anterior temporal lobe. *Cortex* 93, 133–145. doi: 10.1016/j.cortex.2017.05.004
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Mem. Cognit.* 24, 523–533. doi: 10.3758/BF03200940
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Fernandez, A., Díez, E., and Alonso, M. A. (2014). *Normas de Asociación libre en castellano de la Universidad de Salamanca*. Available online at: <http://www.usal.es/gimc/nalc> (accessed January 20, 2020).

- Fernandez, A., Díez, E., Alonso, M. A., and Beato, M. S. (2004). Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behav. Res. Methods Instr. Comput.* 36, 577–583. doi: 10.3758/BF03195604
- Fertonani, A., and Miniussi, C. (2017). Transcranial electrical stimulation: what we know and do not know about mechanisms. *Neuroscientist* 23, 109–123. doi: 10.1177/1073858416631966
- Filmer, H. L., Dux, P. E., and Mattingley, J. B. (2014). Application of transcranial direct current stimulation for understanding brain function. *Trends Neurosci.* 37, 742–753. doi: 10.1016/j.tins.2014.08.003
- Fiori, V., Coccia, M., Marinelli, C. V., Vecchi, V., Bonifazi, S., Ceravolo, M. G., et al. (2011). Transcranial direct current stimulation improves word retrieval in healthy and nonfluent aphasic subjects. *J. Cogn. Neurosci.* 23, 2309–2323. doi: 10.1162/jocn.2010.21579
- Fiori, V., Nitsche, M., Iasevoli, L., Cucuzza, G., Caltagirone, C., and Marangolo, P. (2017). Differential effects of bihemispheric and unihemispheric transcranial direct current stimulation in young and elderly adults in verbal learning. *Behav. Brain Res.* 321, 170–175. doi: 10.1016/j.bbr.2016.12.044
- Flöel, A., Suttrop, W., Kohl, O., Kürten, J., Lohmann, H., Breitenstein, C., et al. (2012). Non-invasive brain stimulation improves object-location learning in the elderly. *Neurobiol. Aging* 33, 1682–9. doi: 10.1016/j.neurobiolaging.2011.05.007
- Friedrich, J., and Beste, C. (2018). Paradoxical, causal effects of sensory gain modulation on motor inhibitory control - a tDCS, EEG-source localization study. *Sci. Rep.* 30:17486. doi: 10.1038/s41598-018-35879-2
- Gainotti, G. (2011). What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia* 49, 2273–2282. doi: 10.1016/j.neuropsychologia.2011.04.027
- Gainotti, G. (2012). The format of conceptual representations disrupted in semantic dementia: a position paper. *Cortex* 48, 521–529. doi: 10.1016/j.cortex.2011.06.019
- Gainotti, G. (2020). Representational and connectivity-based accounts of the cognitive consequences of atrophy of the right and left anterior temporal lobes. *Cogn. Neuropsychol.* 37, 466–481. doi: 10.1080/02643294.2020.1739011
- Gallate, J., Chi, R., Ellwood, S., and Snyder, A. (2009). Reducing false memories by magnetic pulse stimulation. *Neurosci. Lett.* 449, 151–154. doi: 10.1016/j.neulet.2008.11.021
- Galli, G., Vadiello, M. A., Sirota, M., Feurra, M., and Medvedeva, A. (2019). A systematic review and meta-analysis of the effects of transcranial direct current stimulation (tDCS) on episodic memory. *Brain Stimul.* 12, 231–241. doi: 10.1016/j.brs.2018.11.008
- Gallo, D. A. (2006). *Associative Illusions of Memory*. New York, NY: Psychology Press.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Mem. Cognit.* 38, 833–848. doi: 10.3758/MC.38.7.833
- Gómez-Ariza, C. J., Martín, M. C., and Morales, J. (2017). Tempering proactive cognitive control by transcranial direct current stimulation of the right (but not the left) lateral prefrontal cortex. *Front. Neurosci.* 11:282. doi: 10.3389/fnins.2017.00282
- Guillory, J. J., and Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences. *Psychon. Bull. Rev.* 17, 73–81. doi: 10.3758/PBR.17.1.73
- Hoffman, P., McClelland, J. L., and Lambon Ralph, M. A. (2018). Concepts, control, and context: a connectionist account of normal and disordered semantic cognition. *Psychol. Rev.* 125, 293–328. doi: 10.1037/rev0000094
- King, B. R., Rumpf, J. J., Heise, K. F., Veldman, M. P., Peeters, R., Doyon, J., et al. (2020). Lateralized effects of post-learning transcranial direct current stimulation on motor memory consolidation in older adults: an fMRI investigation. *Neuroimage* 223:117323. doi: 10.1016/j.neuroimage.2020.117323
- Köhler, S., and Martin, C. B. (2020). Familiarity impairments after anterior temporal-lobe resection with hippocampal sparing: lessons learned from case NB. *Neuropsychologia* 138:107339. doi: 10.1016/j.neuropsychologia.2020.107339
- Kurkela, K. A., and Dennis, N. A. (2016). Event-related fMRI studies of false memory: an activation likelihood estimation meta-analysis. *Neuropsychologia* 81, 149–167. doi: 10.1016/j.neuropsychologia.2015.12.006
- Lambon Ralph, M., Cipolletti, L., Manes, F., and Patterson, K. (2010). Taking both sides: do unilateral anterior temporal lobe lesions disrupt semantic memory? *Brain* 133, 3243–3255. doi: 10.1093/brain/awq264
- Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philos. Trans. R. Soc. B* 369:20120392. doi: 10.1098/rstb.2012.0392
- Lambon Ralph, M. A., Jefferies, E., Patterson, K., and Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55. doi: 10.1038/nrn.2016.150
- Lambon Ralph, M. A., and Patterson, K. (2008). Generalization and differentiation in semantic memory: Insights from semantic dementia. *Ann. N. Y. Acad. Sci.* 1124, 61–76. doi: 10.1196/annals.1440.006
- Leach, R. Y., McCurdy, M. A., Trumbo, M. C., Matzen, L. E., and Leshikar, E. D. (2019). Differential age effects of transcranial direct current stimulation on associative memory. *J. Gerontol. Series B* 74, 1163–1173. doi: 10.1093/geronb/gby003
- Marful, A., Díez, E., and Fernandez, A. (2015). Normative data for the 56 categories of Battig & Montague (1969) in Spanish. *Behav. Res. Methods* 47, 902–910. doi: 10.3758/s13428-014-0513-8
- Matzen, L., Trumbo, M. C., Leach, R. C., and Leshikar, E. D. (2015). Effects of non-invasive brain stimulation on associative memory. *Brain Res.* 1624, 286–296. doi: 10.1016/j.brainres.2015.07.036
- McCabe, D. P., Presmanes, A. G., Robertson, C. L., and Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychon. Bull. Rev.* 11, 1074–1079. doi: 10.3758/BF03196739
- Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* 527, 633–639. doi: 10.1111/j.1469-7793.2000.t01-1-00633.x
- Nitsche, M. A., and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* 57, 1899–1901. doi: 10.1212/WNL.57.10.1899
- 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
- Patterson, K., and Lambon Ralph, M. A. (2016). “The hub-and-spoke hypothesis of semantic memory,” in *Neurobiology of Language*, eds G. Hickok and S. L. Small (Burlington: Academic Press), 765–775.
- 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
- Peretz, Y., and Lavidor, M. (2013). Enhancing lexical ambiguity resolution by brain polarization of the right posterior superior temporal sulcus. *Cortex* 49:1056. doi: 10.1016/j.cortex.2012.03.015
- Pergolizzi, D., and Chua, E. F. (2015). Transcranial direct current stimulation (tDCS) of the parietal cortex leads to increased false recognition. *Neuropsychologia* 66, 88–98. doi: 10.1016/j.neuropsychologia.2014.11.012
- Pergolizzi, D., and Chua, E. F. (2016). Transcranial direct current stimulation over the parietal cortex alters bias in item and source memory tasks. *Brain Cogn.* 108, 56–65. doi: 10.1016/j.bandc.2016.06.009
- Pisoni, A., Turi, Z., Raithel, A., Ambrus, G. G., Alekseichuk, I., Schacht, A., et al. (2015). Separating recognition processes of declarative memory via anodal tDCS: boosting old item recognition by temporal and new item detection by parietal stimulation. *PLoS ONE* 10:e0123085. doi: 10.1371/journal.pone.0123085
- Price, A. R., Peelle, J. E., Bonner, M. F., Grossman, M., and Hamilton, R. H. (2016). Causal evidence for a mechanism of semantic integration in the angular gyrus as revealed by high-definition transcranial direct current stimulation. *J. Neurosci.* 36, 3829–3838. doi: 10.1523/JNEUROSCI.3120-15.2016
- Psychology Software Tools (2016). *E-prime 3.0*. Available online at: <https://support.psnet.com/> (accessed January 20, 2020).
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/> (accessed June 15, 2020).
- Rajaram, S. (1993). Remembering and knowing: two means of access to the personal past. *Mem. Cognit.* 21, 89–102. doi: 10.3758/BF03211168
- Rice, G. E., Caswell, H., Moore, P., Hoffman, P., and Lambon Ralph, M. A. (2018). The roles of left versus right anterior temporal lobes in semantic memory: a neuropsychological comparison of postsurgical temporal lobe epilepsy patients. *Cerebral Cortex* 28, 1487–1501. doi: 10.1093/cercor/bhx362
- Rice, G. E., Hoffman, P., and Lambon Ralph, M. A. (2015a). Graded specialization within and between the anterior temporal lobes. *Ann. N. Y. Acad. Sci.* 1359, 84–97. doi: 10.1111/nyas.12951

- Rice, G. E., Lambon Ralph, M. A., and Hoffman, P. (2015b). The roles of left versus right anterior temporal lobes in conceptual knowledge: an ALE meta-analysis of 97 functional neuroimaging studies. *Cerebral Cortex* 25, 4374–4391. doi: 10.1093/cercor/bhv024
- Roediger, H. L., and Gallo, D. A. (2016). “Associative memory illusions,” in *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgment and Memory*, eds R. F. Pohl. 2nd Edn. (New York, NY: Routledge), 390–405.
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, I. I. I., H. L., Balota, D. A., and Watson, J. M. (2001). “Spreading activation and arousal of false memories” in *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, eds H. L. Roediger, J. S. Nairne, I. Neath, and A. M. Surprenant (Washington, DC: American Psychological Association Press), 95–115.
- Salvi, C., Beeman, M., Bikson, M., McKinley, M., and Grafman, J. (2020). TDCS to the right anterior temporal lobe facilitates insight problem-solving. *Sci. Rep.* 10:946. doi: 10.1038/s41598-020-57724-1
- Simons, J. S., Lee, A. C. H., Graham, K. S., Verfaellie, M., Koutstaal, W., Hodges, J. R., et al. (2005). Failing to get the gist: reduced false recognition of semantic associates in semantic dementia. *Neuropsychology* 19, 353–361. doi: 10.1037/0894-4105.19.3.353
- Smirni, D., Turriziani, P., Mangano, G. R., Cipolotti, L., and Oliveri, M. (2015). Modulating memory performance in healthy subjects with transcranial direct current stimulation over the right dorsolateral prefrontal cortex. *PLoS ONE* 10:e0144838. doi: 10.1371/journal.pone.0144838
- 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
- Stagg, C. J., Jayaram, G., Pastor, D., Kincses, Z. T., Matthews, P. M., and Johansen-Berg, H. (2011). Polarity and timing-dependent effects of transcranial direct current stimulation in explicit motor learning. *Neuropsychologia* 49, 800–804. doi: 10.1016/j.neuropsychologia.2011.02.009
- Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instr. Comput.* 31, 137–149. doi: 10.3758/BF03207704
- Tanaka, S., Hanakawa, T., Honda, M., and Watanabe, K. (2009). Enhancement of pinch force in the lower leg by anodal transcranial direct current stimulation. *Exp. Brain Res.* 196, 459–465. doi: 10.1007/s00221-009-1863-9
- Tanaka, T., Isomura, Y., Kobayashi, K., Hanakawa, T., Tanaka, S., and Honda, M. (2020). Electrophysiological effects of transcranial direct current stimulation on neural activity in the rat motor cortex. *Front. Neurosci.* 14:495. doi: 10.3389/fnins.2020.00495
- Thapar, A., and McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: effects of retention interval and level of processing. *Mem. Cognit.* 29, 424–432. doi: 10.3758/BF03196393
- The jamovi project (2021). *jamovi (Version 1.6)*. Available online at: <https://www.jamovi.org> (accessed June 15, 2020).
- Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12. doi: 10.1037/h0080017
- Tussing, A. A., and Greene, R. L. (1997). False recognition of associates: how robust is the effect? *Psychon. Bull. Rev.* 4, 572–576. doi: 10.3758/BF03214351
- Wong, C., and Gallate, J. (2012). The function of the anterior temporal lobe: a review of the empirical evidence. *Brain Res.* 1449, 94–116. doi: 10.1016/j.brainres.2012.02.017
- Woollams, A. M., and Patterson, K. (2018). Cognitive consequences of the left-right asymmetry of atrophy in semantic dementia. *Cortex* 107, 64–77. doi: 10.1016/j.cortex.2017.11.014
- World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *J. Am. Med. Assoc.* 310, 2191–2194. doi: 10.1001/jama.2013.281053
- Young, S. J., Bertuccio, M., Sheehan-Stross, R., and Sanger, T. D. (2013). Cathodal transcranial direct current stimulation in children with dystonia: a pilot open-label trial. *J. Child Neurol.* 28, 1238–1244. doi: 10.1177/0883073812460092
- Zhang, J., and Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika* 70, 203–212. doi: 10.1007/s11336-003-1119-8

Conflict of Interest: 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.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Alonso, Díez-Álamo, Gómez-Ariza, Díez and Fernandez. 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) and the copyright owner(s) 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.



Lie for Me: Developmental Trends in Acquiescing to a Blatantly False Statement

Amelia Courtney Hritz^{*†} and Stephen J. Ceci

Department of Psychology, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

Sara Cadavid,
Rosario University, Colombia

Reviewed by:

Olivier Dodier,
Université de Nantes, France
Erik Mac Giolla,
University of Gothenburg, Sweden

*Correspondence:

Amelia Courtney Hritz
ach254@cornell.edu

†Present address:

Amelia Courtney Hritz,
Shapiro Arato Bach LLP, New York,
NY, United States

Specialty section:

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

Received: 06 April 2021

Accepted: 18 August 2021

Published: 23 September 2021

Citation:

Hritz AC and Ceci SJ (2021) Lie for Me: Developmental Trends in Acquiescing to a Blatantly False Statement.
Front. Psychol. 12:691276.
doi: 10.3389/fpsyg.2021.691276

A pair of studies demonstrates that simply asking children to make a blatantly false accusation in the guise of helping others can result in both immediate and long-term false claims. In the pilot study, the initial willingness to make a blatantly false statement was associated with some children making false statements a week later despite being told that the first interviewer had made mistakes during the initial interview. On a positive note, the majority of participants accurately stated that they did not have first-hand knowledge of their accusation's accuracy. Across both studies, the rate of false accusation rates was high. The main experiment demonstrated that children who were young, possessed the lowest verbal intelligence or who were from the lowest SES homes made the most accusations. These findings illustrate not only the dangers of encouraging children to make false statements, but the ease and durability of making such false statements.

Keywords: interview, false memory, witness, age, compliance, lie

INTRODUCTION

A large corpus of research has documented the deleterious effects of various interviewing behaviors. This research has demonstrated that children's report accuracy can be reduced as a result of providing either pre- or post-event misinformation. Pre-event misinformation that can damage children's report accuracy includes the provision of false stereotypes, rumors, inaccurate co-witness information, and unfulfilled expectations (e.g., Pynoos and Nader, 1989; Leichtman and Ceci, 1995; Garven et al., 2000; Principe et al., 2006). Similarly, post-event misinformation that can damage children's report accuracy includes misleading suggestions provided by an interviewer (e.g., Poole and Lindsay, 1995; Cassel et al., 1996), visualization inductions (Ceci et al., 1994), reinforcement (Garven et al., 1998, 2000), imagination-inflation techniques, and leading questions (Ceci, 1994; Ceci and Bruck, 1998; Bruck et al., 2006; Howe et al., 2009; Otgaar et al., 2016, 2017).

In addition to these suggestive techniques associated with report inaccuracy, a growing body of research has documented developmental trends in children's willingness to mislead interviewers. Talwar, Lee, and their associates have produced a large body of research on the developmental course and cognitive and social correlates of lie-telling (e.g., Talwar and Lee, 2002a,b, 2008; Talwar et al., 2007a,b; Talwar and Crossman, 2011; Evans and Lee, 2013). Lies told in the service of achieving selfish ends, such as gaining material rewards or escaping punishment, begin early during the pre-school years and are reduced in magnitude by middle childhood, whereas lies told in the service of socially desirable ends, such as pretending to appreciate an undesirable gift, tend to begin later. Both types of lies

are associated with cognitive and social factors such as theory of mind, social skills, and parenting (e.g., Clarke-Stewart et al., 2004; Lavoie et al., 2016).

The current study builds on the developmental research on children's report inaccuracy and deception in several ways: first, we examine a context that differs from spontaneous lies generated to serve selfish or prosocial aims, namely, willingness to acquiesce to a blatantly false persuasion request, using a paradigm that has not been used for this purpose previously but which may have practical relevance to legal cases. Second, we are interested in the intersection between lies and false beliefs, asking whether the former can influence the latter as some have opined (e.g., Leichtman and Ceci, 1995; Zaragoza et al., 2001; Otgaar and Baker, 2018). Finally, we include sociodemographic variables that have only rarely been included in the memory development literature, but which theoretical research suggests could be important sources of systematic developmental variability (Bronfenbrenner and Ceci, 1994; Ceci et al., 2010). Very few studies have examined children's acquiescence to a misleading persuasion as a function of parent social class, which Talwar, Lee and their colleagues suggest may moderate children's acquiescence. We also include the role of verbal intelligence as a potential moderator; although it has received some attention (e.g., Roebbers and Schneider, 2001; Chae and Ceci, 2005), it has heretofore not been studied in the context of acquiescence to blatantly false persuasion requests. The closest it has come to this context was a study by Clarke-Stewart et al. (2004) that found that 5-year-olds with the highest verbal intelligence were the most resistant to adult attempts at false persuasion.

Children may make statements about events they initially know to be false if the statements are suggested by adults who hold *a priori* beliefs about their authenticity. Adults may signal their beliefs through methods such as repeating specific misinformation during questioning (e.g., Warren et al., 1991; Leichtman and Ceci, 1995; Poole and Lindsay, 2001; Moore et al., 2018), offering praise, bribes or threats of punishment (Garven et al., 1998; Schreiber et al., 2006), rejecting or ignoring children's denials (White et al., 1997; Garven et al., 1998), and selectively reinforcing their incorrect statements (Zaragoza et al., 2001). Such social pressure can cause a child to make statements that, while consistent with the belief of the interviewer, are inconsistent with the child's actual perceptual experience (for a review see Ceci and Bruck, 1998; Bruck et al., 2006). Even mild forms of suggestion can increase inaccurate reports by children, such as descriptions of events by parents (e.g., Thompson et al., 1997; Poole and Lindsay, 2001), visualization inductions (Ceci et al., 1994) being informed that co-witnesses have made a disclosure (Principe and Ceci, 2002), stereotypes (Leichtman and Ceci, 1995; Moore et al., 2018), or even naturally-occurring conversations with parents and peers (Bruck et al., 1999; Principe and Schindewolf, 2012; Kim et al., 2017).

Once encoded, false memories can lead children to maintain inaccurate reports in later neutral interviews. Efforts to retrieve accurate memories following the creation of false memories—for instance by instructing children to say when they are unsure, or

correcting the interviewer when she makes a false suggestion—often will not offset the impact of the false memory (e.g., Poole and Lindsay, 2001; Zaragoza et al., 2001), although there is some evidence that warning children that questions may be tricky does result in a small but significant reduction in errors (~5%) to suggestive questions (Warren et al., 1991). Even efforts to talk children out of their false beliefs can be unsuccessful (e.g., Ceci et al., 1994; Ceci, 1995; Leichtman and Ceci, 1995; Kim et al., 2017). And the metacognitive strategies that are useful in rejecting false information—recollection rejection, retrieval editing, and monitoring—are less effective for young children (Brainerd et al., 2008; Moore et al., 2018). These past findings set the stage for the current study by moving the issue from subtle misleading suggestions by an interviewer to blatantly false persuasive statements.

In contrast to the voluminous literature on the effects of suggestive interviews and misinformation [e.g., see integrative reviews by Poole and Lamb (1998), Quas et al. (2000), Otgaar et al. (2014), Schneider (2015), Goodman et al. (2017)] as well as on factors associated with developmental trends in lie-telling (e.g., the work of Talwar, Crossman, Lee and their colleagues), we were interested in what would happen if instead of exposing participants to techniques designed to bias their report accuracy—such as suggestive interviews, stereotypes, visual inductions, post-event misinformation, incentives to lie—an adult blatantly makes a false accusation in the guise of helping others. Hence, in the current research, we were interested in (a) what happens when children are exposed to an adult's blatantly false assertion and they are asked to repeat it and sign their agreement with it, an extreme form of forced confabulation. Related questions include: (b) will children acquiesce and sign an adult's blatantly erroneous statement, and if so, (c) will they subsequently incorporate it into their long-term reports to a neutral interviewer who instructs them to ignore the blatantly incorrect prior interviewer, and (d) what, if any, developmental trends or social or cognitive correlates will influence their performance?

Thus, in the following experiments we sought to determine whether will children make a blatantly false accusation merely because someone asks them to do so to help unnamed others? And, if so, will children later maintain the false accusation when interviewed outside the presence of the previously biased adult by an unbiased person who urges them to report only what they actually experienced, not what the blatantly biased adult told them? We hypothesized that (a) children will be more likely to make a false accusation in an initial interview if they are blatantly asked to do so in service of helping unnamed others (children who attended a different school, they were told), and (b) they will subsequently maintain this false assertion with a neutral interviewer, especially if they are younger and more suggestible and/or have lower verbal intelligence. These were all *a priori* expectations that we made based on previous research showing that the mechanisms that could drive this effect are unfolding rapidly over early and middle childhood as we briefly describe below. We undertook a mini test of the planned procedure to make sure it would work with the intended age groups in a larger-scale experiment with controls; this mini test gave us confidence

that the youngest children understood the procedure (e.g., the wording to request them to endorse a blatantly false statement) although it also led to changes in the procedure.

There are empirical and theoretical reasons for positing that the period between early and middle childhood is one of rapid development of the factors of interest in the current study. A confluence of cognitive, social, and neurobiological developments unfold between early and middle childhood that are relevant to understanding the effect of exposure to blatantly false statements on later report accuracy. Specifically, social developments are occurring between the ages of 4 and 10 that could be relevant in the context of children's compliance with false persuasions by an adult. As noted, younger children are more influenced by a powerful adult authority figure than are older children (e.g., Ceci et al., 1987); preschoolers are more inclined to conform to false persuasion than are adults (Clarke-Stewart et al., 2004) but older children are actually less likely to conform than adults, thus a U-shaped function (Kim et al., 2017). There are also improvements in source monitoring over early and middle childhood (e.g., Poole and Lindsay, 1995). Quas et al. (2000) reported that younger children had significantly more difficulty sourcing their memories than did older children, a skill that is relevant in the present study in which children are asked to describe an event to a neutral interviewer after receiving a false persuasion request from a prior interviewer, thus creating a potential source misattribution. On the other hand, in their extensive review, Bruck and Melnyk (2004) did not find correlations between source monitoring and suggestibility-proneness even though both displayed normative developmental trends (Quas et al., 2000); the majority of the studies Bruck and Melnyk reviewed showed no relationship between individual differences in source monitoring and suggestibility. Young children also lack metamemorial insights that limit their recollective accuracy, as shown in Wellman's (1978) classic work, when they were more impeded by misinformation (for review see Ceci and Bruck, 1993). Preschool-aged children also lack strategy knowledge, such as elaborative rehearsal (Ornstein, 1978), and they have relatively undeveloped retrieval-time editing (Brainerd and Reyna, 2001), and second-order theory of mind in which participants must identify a third person's beliefs based on what the second person believes, which does not asymptote before at least age 8 (e.g., Arslan et al., 2012; Hiatt and Trafton, 2015). Although the right dorsolateral region of the pre-frontal cortex continues to develop into young adulthood, it is disproportionately undeveloped among preschoolers (Giedd et al., 1999; Gogtay et al., 2004), limiting their ability to inhibit, track, and monitor the contents of their memory (Ceci et al., 2010). In short, this confluence of developments in cognitive (memory, strategy use, theory of mind), social (conformity, deferral of memory to those of perceived as more authoritative), and neuromaturational (tracking, monitoring) mechanisms converge to anticipate that the period between early and middle childhood is one of heightened relevance for the present hypotheses. It was for this reason that we focused on this transition.

The departure point for the present set of experiments is that unlike studies that employed techniques known to bias report

accuracy, in the present study none of the documented explicit or implicit suggestive techniques that damage report accuracy were employed—i.e., there were no post-event suggestions, misleading questions, visually-guided inductions, stereotypes, reinforcements, or automatic semantic associations that have previously been shown to cause children's reporting errors (Bruck and Ceci, 1999; Ceci and Friedman, 2000; Ceci and Bruck, 2006). Instead, the present approach seems straightforward and could have implications for disclosures that come about in the context of one adult urging a child to acquiesce to a false statement about another. For example, in an acrimonious custody dispute a child, upon the request of one parent, may initially repeat a blatantly false statement made by one parent about another parent, and over time elaborate the statement, as has been suggested by some to occur in actual cases (see Bruck and Ceci, 2013). What effect might repeating a blatantly false assertion have on a children's later reports to a neutral interviewer who encourages report accuracy?

PILOT STUDY

Based on insights from a demonstration study of 16 children who watched clowns perform magic tricks, we designed a somewhat larger pilot study in which children watched a chemist perform "magic tricks" with chemistry. An interviewer later asked children to tell them that the chemist had broken a test tube during the demonstration. At no time did the chemist break any of the test tubes in the magic show and children were not shown any evidence of a broken test tube during either the event itself or during either of the two interviews that followed it. Thus, the children in this experiment were not provided with any evidence that the chemist broke a test tube during the show they watched; the only basis for claiming he did so was an assertion by the initial interviewer that others who attended a different camp session had told her he broke a test tube at their session.

Methods

Participants

Data for this pilot study were collected on 54 children attending a summer camp hosted by a children's science museum. These participants ranged in age from 6 to 11 years old ($M = 8.52$, $SD = 1.42$); 41% were male. Ideally, we planned to extend this age range downward to capture the entire early childhood-middle childhood span that has been implicated in the developmental research reviewed earlier.

Procedure

The measures and procedure employed in this study were approved by the university's institutional review board (1303003677 "Children's False Accusations without Suggestive Questioning"). Participants watched a chemist perform 12 different magic tricks including a trick where she mixed acids, bases, and indicators in a test tube to create a rainbow tube. Several days later ($M = 2.02$ days, $SD = 0.94$), participants were interviewed individually by an unfamiliar female adult. This interviewer solicited a free narrative about what transpired during the magic show, followed by two directive questions

about the magic show: “Do you remember seeing a test tube?”; “Did the chemist do any tricks with the test tube?” (Answers to these questions and the child’s free narrative served as a baseline to judge later responses that followed this interviewer’s request to endorse her blatant lie). Next, the interviewer informed the child that she had heard from children who attended another chemistry show that the chemist broke his test tube during their show. Then she blatantly asked the child to say the chemist broke a test tube during the magic show they saw (even though he had not). Specifically, she said: “I need your help so he doesn’t accidentally do this again, or other children will not get to see the magic show. Can you help me? Can you tell me the chemist broke the test tube so I can make sure he doesn’t do this again?” The direct request for affirmation was done to resemble a situation in which an adult asks a child to affirm a non-event, one they could not have witnessed but which they might assume the adult had witnessed and needed their assent for prosocial reasons (allow future children to see the show). After the child answered, the interviewer thanked the child and offered stickers.

Approximately 1 week later ($M = 7.56$ days, $SD = 6.38$), a neutral interview took place during which an unfamiliar female adult explained that the previous interviewer had made mistakes when talking to children about the magic show, and she wanted the child to report only what they actually saw in the magic show rather than what the previous interviewer told them. Following these instructions from this neutral interviewer, she asked the child if the chemist accidentally broke his test tube. If the child said yes, the interviewer asked if they had seen it happen with their own eyes. Following this neutral interview, children completed intelligence and suggestibility measures.

In light of the literature on verbal ability and suggestibility, we administered the vocabulary subset of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) to examine if a *g*-loaded measure of children’s verbal intelligence predicted responses during the first and second interview. The vocabulary scaled scores from the Wechsler series of intelligence tests were collected as a rough measure of general intelligence, given its very high saturation on the general intelligence factor, *g* (Flynn, 2007; Nisbett, 2009). Raw scores on the vocabulary subset of the Wechsler Abbreviated Scale of Intelligence were calculated based on the general scoring principles (Wechsler, 1999). These raw scores were converted to *T*-scores, which are age-corrected (thus removing any correlation with chronological age) and have a wide range (Wechsler, 1999). Results indicated that intelligence was normally distributed ($M = 63.53$, $SD = 11.44$, $n = 53$) with negative skew.

We administered the Video Suggestibility Scale for Children (VSSC) to discover whether suggestibility predicted responses during the first and second interview. The VSSC produces two parameters of suggestibility-proneness: yield (succumbing to erroneous statements) and shift (changing originally correct answers to false answers in response to negative feedback) (Scullin and Ceci, 2001; Scullin et al., 2002). Results did not differ when the “yield” and “shift” measures from the VSSC were examined separately, and therefore the combined parameter scores were used for the analysis. Scores on the VSSC were

normally distributed with a slight positive skew ($M = 9.6$, $SD = 4.35$, $n = 52$).

We found no correlation among our individual difference measures: verbal intelligence and suggestibility were not correlated, and suggestibility was also not correlated with age.

Results

Most participants were willing to accuse the chemist of breaking the test tube, even though they had not witnessed it and were not presented with any physical evidence of a broken test tube. When blatantly asked to make a false statement, 34 out of 54 children (63%), asserted that the chemist broke the test tube. Importantly, no child made such false allegations spontaneously during their prior free narrative; it was only done in response to the subsequent blatant request from the interviewer. One week later after being told the initial interviewer made mistakes and got children to make mistakes about what happened during the magic show, most children recanted. Only 13 of the 34 children who had previously asserted the chemist broke the test tube maintained their false accusation. All but two of the children who asserted the accusation during the second interview had also made the assertion in the first interview. These children who made false accusations in both interviews were 32% of the 34 people who made the accusation in the first interview and 20% of the total original sample of 54. In addition, five children asserted that they saw the chemist break the test tube with their own eyes. This represented 16% of the 34 who made the accusation and 9% of the original 54.

In sum, a fraction of the participants who watched the chemist do a magic show went along with the initial interviewer’s blatant request to affirm her false allegation and many subsequently maintained this false allegation with some claiming to have witnessed it with their own eyes despite no suggestive techniques being employed during the second interview (i.e., there was no provision of erroneous post-event information, misleading questions, imagery inductions, requests to speculate, clumsy stereotypes, forced confabulation questions, etc.). In contrast to the children who went along with the first interviewer’s blatant request for a false affirmation, none of the 54 children had made such false allegations spontaneously in their free narratives of what occurred before the interviewer requested the child make a false affirmation. Thus, the damage to children’s report accuracy was the result of children assenting to the initial interviewer’s assertion, a small fraction of whom subsequently claimed when speaking with a neutral interviewer a week later not only that it occurred but to have witnessed it with their own eyes.

We next compared the children who made a false accusation by age, verbal intelligence, and suggestibility-proneness. Results are displayed in **Table 1**. As we expected, the youngest children appeared to make most of the false allegations. Because it is possible that with more participants, age would emerge as a significant predictor, we tested this in the main experiment, using a predetermined sample size that possessed ample power to detect differences of the observed magnitude. Similarly, as noted, ultimately a total of five children in this Pilot Study who watched the chemistry show said they saw with their own eyes the chemist

TABLE 1 | Pilot study: average age, intelligence, and suggestibility by whether they made an accusation during the first interview, second interview, or both.

	Two accusations	T 1 accusation only	No accusation
Age	7.81 (1.09)	8.76 (1.59)	8.66 (1.13)
Intelligence	62.20 (6.84)	64.05 (11.80)	63.6 (14.1) ^a
Suggestibility	9.82 (5.17)	10.00 (4.21) ^a	9.06 (4.30)
N	11	22	18

Values displayed are means (standard deviations in parentheses). Two participants made an accusation in the second interview but not the first, due to the small number they are excluded from the table. ^a1–2 participants missing scale data.

break the test tube, despite being told by the neutral interviewer that the initial interviewer had caused children to make mistakes and that they should only report what they actually witnessed rather than what the prior interviewer may have told them. All five of these children were 8 years old and younger. Even with the small sample size in this pilot, this age difference was reliable, $t(51) = 2.02$, $p < 0.05$, Hedges' $g = 0.95$ (large effect). Further comparisons of these five participants did not lead to additional significant results, which could be due to the small sample size.

We did not notice trends when we compared children who made a false accusation by verbal intelligence or suggestibility. Follow-up work with larger and more diverse (in terms of verbal intelligence) samples was conducted in the following main experiment to provide a more robust test of these findings.

MAIN EXPERIMENT

The data and the findings from pilot study revealed a number of interesting results despite the small sample size. In view of these results, the main experiment was a modification designed to broaden the context by substituting a more active role for the child than was the case in the pilot study where the child passively observed a magic show, and also to include another dependent variable: parent socio-economic status (Talwar et al., 2017). In this experiment children were once again asked by an interviewer to agree with a blatantly false assertion but in a less passive context from the one used in the pilot study. Because of the narrow age differences in the pilot study, this experiment was designed with finer age gradations to shed light on developmental vs. reverse developmental effects (e.g., Kim et al., 2017). Once again, vocabulary scaled scores from the Wechsler series of intelligence tests were collected as a rough measure of general intelligence, g (Flynn, 2007; Nisbett, 2009). Finally, we endeavored to recruit a broad range of SES given that compliance with adult authority figures may be related to parent educational attainment as well as suggestibility (Chae et al., 2016).

Method

Participants

Children were recruited through schools, preschools, and a university-run summer camp. One hundred and seventy-one children and adolescents participated, 43 4-year-olds (23 females; $M = 50.40$ months, $SD = 2.53$), and 44 6-year-olds (25 females;

$M = 73.70$ months, $SD = 3.86$), 44 8-year-olds (22 females; $M = 99.09$ months, $SD = 5.02$), and 44 12-year-olds (19 females; $M = 151.27$ months, $SD = 6.27$).

Procedure

This experiment involved one male and one female research assistant. Children were brought into a testing room and greeted by an opposite-sex research assistant. The youngest children were escorted by a parent or guardian and the older children were escorted by a teacher's aide or a camp counselor. The opposite-sex assistant provided crayons and a coloring book for the child to play with for ~ 10 min. Toys were displayed prominently on the table, as well as several items of clothing, including a straw hat. After ~ 10 min, a same-sex research assistant entered the room and was introduced to the child as Jenn or John by the opposite-sex assistant who then departed. To avoid cross-sex confounds, all final interviews were conducted by same-sex research assistants. After entering the room, the same-sex assistant engaged the child in actively playing a couple rounds of Simon Says. During this game Jenn or John instructed the child "Simon Says [John/Jenn] put on a straw hat," which the same-sex assistant donned; they went through five such actions. After playing this game, the original opposite-sex assistant returned to the room and the same-sex assistant departed. After several minutes of amiable interaction with the child, the opposite-sex research assistant asked the child two memory questions related to real and suggested actions in the Simon Says game ("Do you remember the name of the person who played Simon Says with you?", "Do you remember what Simon Says told him (or her) to do with the straw hat?"). Next, this opposite-sex assistant made two blatantly false assertions about events: the child was told that John (or Jenn) had broken a non-present cell phone and asked the child to say that they saw this happen. Following their response, they were presented with a typewritten document and asked to make a mark to indicate if they saw John (or Jenn) break the cell phone. Then they were told that John (or Jenn) ripped a colored drawing and they were asked to make a mark to indicate that they observed John (or Jenn) rip it. In reality, John (or Jenn) did neither thing (Older participants were asked to sign their name on a line indicating they saw John (or Jenn) rip the drawing and break the cell phone; pre-school-aged children were shown two blank spaces on the sheet where they were asked to make crayon marks to indicate that they saw John (or Jenn) damage each of these items, if they agreed that this happened).

Individual Differences

To measure verbal intelligence, raw scores on the vocabulary subtest of the Wechsler Preschool and Primary Scale of Intelligence (for the youngest age group) and the Wechsler Intelligence scale for Children-IV (for the three older groups) were converted to scaled scores with a mean of 10 and a standard deviation of 3 based on national norms by age. As was true in the pilot study, the sample was skewed slightly above the national average, with a mean of 11.11, and standard deviation of 2.52.

Parent educational attainment is reported in Table 2. Parent SES may moderate the effect of parenting practices on children's willingness to lie (Talwar et al., 2017) and has been tied to report

TABLE 2 | Parent education of participants.

	Parent 1	Parent 2
Some high school or less	1	0
High school diploma/GED	22	10
Some college	35	23
College degree	46	51
Some graduate school	48	58
Graduate degree	22	28
Unreported	1	5

TABLE 3 | Descriptive statistics by age group.

	4 year olds	6 year olds	8 year olds	12 year olds
Intelligence	10.98 (3.01)	11.18 (2.58)	11.23 (2.22)	11.05 (2.26)
Parent Education	4.19 (1.33)	4.32 (1.08)	4.09 (1.04)	4.19 (0.96)
Accuracy				
One correct	22 (51%)	16 (36%)	12 (27%)	18 (41%)
Two correct	17 (40%)	25 (57%)	29 (66%)	23 (52%)
N	43	44	44	44

For intelligence, parent education and accuracy, values are mean (standard deviation), all other values are sample size. Parent education is coded on a six-point scale and averaged across parents: 1, did not complete high school; 2, high school graduate; 3, some college attendance; 4, college completion; 5, some post-baccalaureate coursework; 6, PhD or professional degree.

accuracy and suggestibility (e.g., Chae et al., 2016). After coding parent education on a six-point scale and averaging across both parents, parent education was strongly correlated with children's verbal intelligence scaled scores, $r(173) = 0.37$, $p < 0.001$.

Participants were asked two questions about salient details in the Simon Says game to gauge how well they attended to it. The questions probed the name of the assistant with whom they played Simon Says (Jenn/John) and what was placed on their partner's head (a straw hat). Based on the accuracy of their answers, they were assigned values of 0, 1, or 2. Thirteen participants answered both questions incorrectly (7%), 68 answered one correctly (39%), and 94 answered both correctly (54%).

For each age group descriptive information on the individual difference variables is displayed in Table 3.

Results

One-hundred-forty-five participants claimed to have witnessed at least one false event, and 66 participants claimed to have witnessed both false events. A logistic regression model was estimated to predict which children made a false assertion as a function of age, verbal intelligence, and parent education. We mean-centered age, standardized intelligence, and mean-centered the average parent education. Table 4 shows the results of the regression. As can be seen, age and parent education were significant predictors for making at least one false assertion. Younger children and children whose parents had less education were more likely to make at least one false assertion. We

estimated a 41% increase in the odds of making a false accusation for a 1-year decrease in age.

Post-hoc analysis further confirmed the strong relationship between age and making false assertions. When we substituted the continuous age variable with the age group variable in the logistic regression, keeping the other predictor variables the same, and ran pairwise comparisons, the 12-year-olds were significantly less likely to make a false assertion compared to all other age groups. More specifically, compared to 12-year-olds, the odds of making a false assertion were 14.35 times higher for 4-year-olds, 12.29 times higher for 6-year-olds, and 5.13 times higher for 8-year-olds (p 's = 0.002, 0.003, 0.05, respectively). Thus, like the pilot study, this study also documented age differences.

There was a similarly strong relationship between parent education and false assertions. As displayed in Table 4, we estimated the odds of signing a false statement were 4.89 times higher for each unit decrease in parent education. In fact, all of the children who refrained from signing either of the false statements had parents with at least a baccalaureate degree. Of the children whose mothers did not have a college degree, 66% signed both false statements (compared to 33% in the group whose mothers had a college degree and 17% in the group with mothers with post baccalaureate education). A similar trend was found with father's education.

Intelligence was not a significant predictor, however, as noted, it was significantly correlated with parent education. When parent education was omitted from the regression, intelligence was significant, $p = 0.05$.

We noticed that there may be a relationship between accuracy and making a false assertion: of the 13 participants who answered both salient questions incorrectly, all claimed to have witnessed at least one false event and 8 of 13 claimed to have witnessed both (62%). In case the 13 participants who answered both questions about salient details incorrectly were not sufficiently paying attention, we reran the regression after omitting their responses and results were not significantly different: age and parent education remained significant predictors of making false assertions (both p s < 0.001).

In addition, we ran a logistic regression comparing individuals who made both false assertions to individuals who made one or zero false assertions. Again, in this regression age and parent education were significant predictors (both p s < 0.001).

DISCUSSION

In the pilot study a sizable portion of children complied with an adult's request to make a blatantly false accusation even though they lacked first-hand knowledge of the alleged infraction. On the other hand, most children who made these false accusations did accurately disclose the truth in a subsequent neutral interview after the interviewer gave them releasing instructions ("The person who talked to you before made a lot of mistakes and got children to make mistakes Please tell me only what you actually saw, not what someone told you."). And even when they complied with the blatant request to make a false statement

TABLE 4 | Logistic regression predicting signing at least one false statement.

	<i>B</i>	<i>SE</i>	95% CI	Odds	Odds 95% CI
Intercept	−2.59***	0.40	−3.48, 1.89	0.75	0.03, 0.15
Age	0.34***	0.08	0.19, 0.51	1.41	1.21, 1.67
Intelligence	0.22	0.29	−0.33, 0.80	1.25	0.72, 2.23
Parent's education	1.59***	0.37	0.92, 2.39	4.89	2.51, 10.90

*** $p < 0.001$.

and maintained this falsehood when subsequently interviewed by a neutral interviewer, most of these children did not claim to have seen it with their own eyes. In sum, these findings demonstrate that under conditions in which the child is only subjected to a single blatant request for false information, in later interviews a small number of the young children may purport to remember the accusation and misattribute its source to personal experience observing the infraction rather than to the blatantly false statement by the initial interviewer who asked them to claim they actually observed it.

After the neutral interviewer informed children that the prior interviewer had made mistakes and misled children, not all children accurately disclosed that they had not observed the infraction. It is possible that children's initial compliance may have created false memories, retrieval competitions, or source misattributions in some of the youngest children who claimed to have seen the infraction with their own eyes despite being given "release" instructions by the neutral interviewer. Such release instructions should, if anything, have motivated them to retract their former false assertion if they were aware of its falsity. Of course, this is speculative as we have no direct test of the hypothesis that the initial compliance request actually distorted memory as opposed to other possibilities such as children's loyalty to, or even fear of, the adult in the first session, could plausibly lead to them to continuing lying in the follow-up interview. Future research will be needed to test this.

In contrast to reversed developmental trends in which younger children are more resistant to spontaneous false memory due to their less developed semantic associative networks (e.g., Brainerd et al., 2010; Brainerd and Reyna, 2012; Otgaar et al., 2016; Kim et al., 2017), in both the pilot and main study, false assertions were more likely among the youngest children. This developmental finding is consistent with well-documented age trends in source misattributions which routinely document that younger children have greater difficulty separating various sources or inputs into their memories (e.g., Ceci et al., 1994; Quas et al., 2000; Goodman et al., 2017). This literature suggests that children who make source misattributions genuinely come to believe in the veracity of their misattributions and "generally were not able to report that they had been asked about these events in prior interviews" (Quas et al., 2000, p. 218).

In the main study, false assertions were found to occur disproportionately more often among children who had the least educated parents. Relatively little research has examined children's acquiescence to misleading suggestions as a function of socioeconomic and intellectual factors. In the present study, a

proxy for parent social class proved to exert a powerful influence. In light of this, one wonders whether prior findings in the developmental literature might be qualified if future researchers were to replicate former designs but include a socioeconomic measure. The role of verbal intelligence has received some attention (e.g., Roebbers and Schneider, 2001; Chae and Ceci, 2005). This literature suggests that children with higher verbal intelligence provide more accurate recall and children with very low intelligence can be more suggestible in response to misleading questions. In our study, intelligence, which was significantly correlated with parent education, did not uniquely predict making a false assertion in the context of conformity to blatant lies.

Taken together, the findings from these experiments have implications for cases in which adults articulate biases during conversations with children. The influence of the interviewer can lead some children to make accusations that they initially know are false. In itself, this is hardly a new finding, as decades of deception research have documented that children will lie in response to various incentives (e.g., Talwar et al., 2011; Wyman et al., 2016). However, the present findings demonstrate that this is more strongly observed among the youngest participants from the lowest educated families.

Caveats and Limitations

The results of the pilot study and main experiment are limited in their forensic implications because we refrained from creating the stress associated with an actual forensic interview in which children: (a) usually know the individual they are accusing, (b) understand that their answers may influence others' opinion of this individual, and (c) are cognizant that this individual could face adverse consequences as a result of their statement. Ethical considerations preclude us from making children feel seriously uncomfortable or protective of loved ones. Thus, the children in these experiments were told that it was probably an accident that the item was broken, and that their help was needed to make sure this did not happen again so that children in another school could enjoy the use of the item. This was done to minimize stress but at the same time it deviates from legal contexts where stress is inherent, thus limiting its practical import.

Furthermore, there were no negative consequences associated with making a false accusation. There was also far less pressure on the children to make false accusations than may inhere in child abuse cases in which multiple suggestive methods might be used in interviews, such as introduction of new suggestive information, positive reinforcement, interviewer's expression of

disbelief when a child fails to disclose, conformity pressure, and invitations to pretend or speculate (e.g., Garven et al., 1998; Schreiber et al., 2006; Bruck and Ceci, 2013). If in real cases interviewers with biases exerted more pressure and used more suggestions than was done in the current study, the deleterious consequences could be even greater than what were observed under these less intensive circumstances. Thus, even though a substantial fraction of the children in these experiments affirmed the false statement and maintained this affirmation over time, this might have been elevated by increasing the intensity and number of requests.

In addition, the present experiments, although careful to include variables that have often been missing from past studies (socioeconomic status, verbal intelligence), nevertheless did not examine other potentially important individual differences that could be instrumental. Recently, a number of researchers have begun to examine such factors as child and parental attachment status as it relates to suggestibility (Chae et al., 2014, 2018), children's frontal neurological status (Poole et al., 2014), children's social skills (Lavoie et al., 2016), and parental rearing styles as they relate to children's compliance with a false report (Kim et al., 2017). It would be interesting for future work to add such variables to the study of blatantly false statements.

Future research will have to chart the boundary conditions of this effect, although some experimental evidence indicates that children's suggestibility is exhibited even under conditions of stressful physical experience such as during painful medical procedures (e.g., Bruck et al., 1995), and case studies of contested custody are rife with analogs of the present procedure. For example, Bruck and Ceci (2013) describe a custody case that progressed into a series of accusations of sexual abuse by the father of two preschool-aged daughters. The children likely overheard claims made by their mother and repeated statements to a counselor such as "Mommy says Daddy is mean." The

present findings suggest that initially agreeing with an adult's request to affirm such assertions may result in some children later repeating it to a neutral interviewer even if the child initially was aware they had not witnessed them (e.g., in the above case the allegation that the father had harshly snatched a credit card from the mother and cut it half).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB, Cornell University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AH and SC designed the study together. SC drafted the introduction and discussion. AH drafted the methods and results. The manuscript has been read and approved by both authors. Once the full manuscript was assembled, both authors contributed to editing and finalizing the manuscript. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Caisa Royer, Grace Monks, Victoria Dahl, Logan Kenney, and James Dunlea for collecting and coding data and we are grateful to all the families who participated.

REFERENCES

- Arslan, B., Hohenberger, A., and Verbrugge, R. (2012). "The development of second-order social cognition and its relation with complex language understanding and working memory," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (Austin, TX), 1290–1295.
- Brainerd, C., Reyna, V., Wright, A. H., and Mojardin, R. (2008). Recollection rejection: false memory editing in children and adults. *Psychol. Rev.* 110, 762–784. doi: 10.1037/0033-295X.110.4.762
- Brainerd, C. J., Holliday, R. E., Reyna, V. F., Yang, Y., and Toglia, M. P. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *J. Exp. Child Psychol.* 107, 137–154. doi: 10.1016/j.jecp.2010.04.013
- Brainerd, C. J., and Reyna, V. F. (2001). Fuzzy-trace theory: dual processes in memory, reasoning, and cognitive neuroscience. *Adv. Child Dev. Behav.* 28, 41–100. doi: 10.1016/S0065-2407(02)80062-3
- Brainerd, C. J., and Reyna, V. F. (2012). Reliability of children's testimony in the era of developmental reversals. *Dev. Rev.* 32, 224–267. doi: 10.1016/j.dr.2012.06.008
- Bronfenbrenner, U., and Ceci, S. J. (1994). Nature-nurture in developmental perspective: a bioecological theory. *Psychol. Rev.* 101, 568–586. doi: 10.1037/0033-295X.101.4.568
- Bruck, M., and Ceci, S. J. (1999). The suggestibility of children's memory. *Ann. Rev. Psychol.* 50, 419–439. doi: 10.1146/annurev.psych.50.1.419
- Bruck, M., and Ceci, S. J. (2013). Expert testimony in a child sex abuse case: translating memory development research. *Memory* 21, 556–565. doi: 10.1080/09658211.2013.769606
- Bruck, M., Ceci, S. J., and Francoeur, E. (1999). The accuracy of mothers' memories of conversations with their preschool children. *J. Exp. Psychol.* 5, 89–106. doi: 10.1037/1076-898X.5.1.89
- Bruck, M., Ceci, S. J., Francoeur, E., and Barr, R. (1995). "I hardly cried when I got my shot!": influencing children's reports about a visit to their pediatrician. *Child Dev.* 66, 193–208. doi: 10.2307/1131200
- Bruck, M., Ceci, S. J., and Principe, G. (2006). "The child and the law," in *Handbook of Child Psychology: Psychology in Practice, 6th Edn, Vol. 4*, eds I. Sigel and K. A. Renninger (NY: Wiley), 776–816.
- Bruck, M., and Melnyk, L. (2004). Individual differences in children's suggestibility: a review and synthesis. *Appl. Cogn. Psychol.* 18, 947–996. doi: 10.1002/acp.1070
- Cassel, W. S., Roebbers, C. E. M., and Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *J. Exp. Child Psychol.* 61, 116–133. doi: 10.1006/jecp.1996.0008
- Ceci, S. J. (1994). "Cognitive and social factors in children's testimony," in *Psychology in Litigation and Legislation*, eds B. D. Sales and G. R. Vandenbos (Washington, DC: American Psychological Association), 14–54.
- Ceci, S. J. (1995). "False beliefs: some developmental and clinical considerations," in *Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past*, eds D. L. Schacter, J. T. Coyle, G. D. Fischbach, M.-M. Mesulam, and L. E. Sullivan (Cambridge, MA: Harvard University Press), 91–125.

- Ceci, S. J., and Bruck, M. (1993). The suggestibility of the child witness: a historical review and synthesis. *Psychol. Bull.* 113, 403–439. doi: 10.1037/0033-2909.113.3.403
- Ceci, S. J., and Bruck, M. (1998). “Children’s testimony: applied and basic issues,” in *The Handbook of Child Psychology, 5th Edn.*, Vol. 4, eds W. Damon, I. Sigel, and K. A. Renninger (Hoboken NJ: Wiley and Sons Inc), 713–774.
- Ceci, S. J., and Bruck, M. (2006). Children’s suggestibility: characteristics and mechanisms. *Adv. Child Dev. Behav.* 34, 247–281. doi: 10.1016/S0065-2407(06)80009-1
- Ceci, S. J., Fitneva, S. A., and Williams, W. M. (2010). Representational constraints on the development of memory and metamemory: a developmental-representational-theory. *Psychol. Rev.* 117, 464–495. doi: 10.1037/a0019067
- Ceci, S. J., and Friedman, R. D. (2000). The suggestibility of children: scientific research and legal implications. *Cornell Law Rev.* 86, 34–108.
- Ceci, S. J., Loftus, E. F., Leichtman, M., and Bruck, M. (1994). The possible role of source misattributions in the creation of false beliefs among preschoolers. *Int. J. Clin. Exp. Hypnosis* 42, 304–320. doi: 10.1080/00207149408409361
- Ceci, S. J., Ross, D. F., and Toglia, M. P. (1987). Suggestibility of children’s memory: psycholegal implications. *J. Exp. Psychol.* 116, 38–49. doi: 10.1037/0096-3445.116.1.38
- Chae, Y., Goodman, G. S., Larsen, R. P., Augusti, E. M., Alley, D., VanMeenen, K. M., et al. (2014). Children’s memory for distressing events: the role of children’s and parents’ attachment. *J. Exp. Child Psychol.* 123, 90–111. doi: 10.1016/j.jecp.2014.01.005
- Chae, Y., Goodman, M., Goodman, G. S., Troxel, N., McWilliams, K., Thompson, R. A., et al. (2018). How children remember the strange situation: the role of attachment. *J. Exp. Child Psychol.* 166, 360–379. doi: 10.1016/j.jecp.2017.09.001
- Chae, Y., Kulkofsky, S., Debaran, F. B., Wang, Q., and Hart, S. L. (2016). Low-SES preschool children’s eyewitness memory: the role of narrative skill. *Behav. Sci. Law* 34, 55–73. doi: 10.1002/bsl.2242
- Chae, Y. J., and Ceci, S. J. (2005). Individual differences in children’s recall and suggestibility: the effect of intelligence, temperament, and self-perceptions. *Appl. Cogn. Psychol.* 19, 383–407. doi: 10.1002/acp.1094
- Clarke-Stewart, K. A., Malloy, L. C., and Allhusen, V. D. (2004). Verbal ability, self-control, and close relationships with parents protect children against misleading suggestions. *Appl. Cogn. Psychol.* 18, 1037–1058. doi: 10.1002/acp.1076
- Evans, A. D., and Lee, K. (2013). Emergence of lying in very young children. *Dev. Psychol.* 49, 1958–1963. doi: 10.1037/a0031409
- Flynn, J. R. (2007). *What is intelligence?* NY: Cambridge University Press. doi: 10.1017/CBO9780511605253
- Garven, S., Wood, J. M., Malpas, R. S., and Shaw, J. S. (1998). More than suggestion: The effect of interviewing techniques from the McMartin preschool case. *J. Appl. Psychol.* 83, 347–359. doi: 10.1037/0021-9010.83.3.347
- Garven, S., Wood, J. M., and Malpass, R. (2000). Allegations of wrongdoing: the effects of reinforcement on children’s mundane and fantastic claims. *J. Appl. Psychol.* 85, 38–49. doi: 10.1037/0021-9010.85.1.38
- Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, E. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* 2, 861–863. doi: 10.1038/13158
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8174–8179. doi: 10.1073/pnas.0402680101
- Goodman, G. S., Jones, O., and McLeod, C. (2017). Is there a consensus about children’s memory and suggestibility? *J. Interpers. Violence* 32, 926–939. doi: 10.1177/0886260516657358
- Hiatt, L. M., and Trafton, J. G. (2015). “Understanding second-order theory of mind,” in *HRI 2015 Extended Abstracts*, ACM 978-1-4503-3318-4/15/03. doi: 10.1145/2701973.2702030
- Howe, M. L., Wimmer, M. C., Gagnon, N., and Plumpton, S. (2009). An associative activation theory of children’s and adults’ memory illusions. *J. Mem. Lang.* 60, 229–251. doi: 10.1016/j.jml.2008.10.002
- Kim, I. K., Kwon, E., and Ceci, S. J. (2017). Developmental Reversals in report conformity: psycho-legal implications. *Appl. Cogn. Psychol.* 31, 128–138. doi: 10.1002/acp.3309
- Lavoie, J., Yachison, S., Crossman, A. M., and Talwar, V. (2016). Polite, instrumental, and dual liars: relation to children’s developing social skills and cognitive ability. *Int. J. Behav. Dev.* 41, 257–264. doi: 10.1177/0165025415626518
- Leichtman, M. D., and Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers’ reports. *Dev. Psychol.* 31, 568–578. doi: 10.1037/0012-1649.31.4.568
- Moore, K. N., Lampinen, J. M., Gallo, D. A., Adams, E. J., and Bridges, A. J. (2018). Children’s use of memory editing strategies to reject source misinformation. *Child Dev.* 89, 219–234. doi: 10.1111/cdev.12716
- Nisbett, R. (2009). *Intelligence and How to Get It*. NY: Norton.
- Ornstein, P. A. (ed.). (1978). *Memory Development in Children*. Hillsdale, NJ: Erlbaum.
- Otgaar, H., and Baker, A. (2018). When lying changes memory for the truth. *Memory* 26, 2–14. doi: 10.1080/09658211.2017.1340286
- Otgaar, H., Howe, M. L., Brackmann, N., and van Helvoort, D. H. J. (2017). Eliminating age differences in children’s and adults’ suggestibility and memory conformity effects. *Dev. Psychol.* 53, 962–970. doi: 10.1037/dev0000298
- Otgaar, H., Howe, M. L., Peters, M., Smeets, T., and Moritz, S. (2014). The production of spontaneous false memories across childhood. *J. Exp. Child Psychol.* 121, 28–41. doi: 10.1016/j.jecp.2013.11.019
- Otgaar, H., Scoboria, A., Howe, M. L., Moldoveanu, G., and Smeets, T. (2016). Challenging memories in children and adults using an imagination inflation procedure. *Psychol. Conscious.* 3, 270–283. doi: 10.1037/cns0000087
- Poole, D. A., Dickinson, J. J., Brubacher, S. P., Liberty, A. E., and Kaake, A. M. (2014). Deficient cognitive control fuels children’s exuberant false allegations. *J. Exp. Child Psychol.* 118, 101–109. doi: 10.1016/j.jecp.2013.08.013
- Poole, D. A., and Lamb, M. E. (1998). *Investigative Interviews of Children: A Guide for Helping Professionals*. Washington, DC: American Psychological Association. doi: 10.1037/10301-000
- Poole, D. A., and Lindsay, D. S. (1995). Interviewing preschoolers: effects of nonsuggestive techniques, parental coaching and leading questions on reports of nonexperienced events. *J. Exp. Child Psychol.* 60, 129–154. doi: 10.1006/jecp.1995.1035
- Poole, D. A., and Lindsay, D. S. (2001). Children’s eyewitness reports after exposure to misinformation from parents. *J. Exp. Psychol.* 7, 27–50. doi: 10.1037/1076-898X.7.1.27
- Principe, G., Kanaya, T., Ceci, S. J., and Singh, M. (2006). Believing is seeing: how rumors can engender false memories in Preschoolers. *Psychol. Sci.* 17, 243–248. doi: 10.1111/j.1467-9280.2006.01692.x
- Principe, G., and Schindewolf, E. (2012). Natural conversations as a source of false memories in children: implications for the testimony of young witnesses. *Dev. Rev.* 32, 205–223. doi: 10.1016/j.dr.2012.06.003
- Principe, G. F., and Ceci, S. J. (2002). “I saw it with my own ears”: the effects of peer conversations on preschoolers’ reports of nonexperienced events. *J. Exp. Child Psychol.* 83, 1–25. doi: 10.1016/S0022-0965(02)00120-0
- Pynoos, R. S., and Nader, K. (1989). Children’s memory and proximity of violence. *J. Am. Acad. Child Adolesc. Psychiatry* 28, 236–241. doi: 10.1097/00004583-198903000-00015
- Quas, J. A., Schaaf, J. M., Alexander, K. W., and Goodman, G. S. (2000). “Do you really remember it happening or do you only remember being asked about it happening?: children’s source monitoring in forensic contexts,” in *Children’s Source Monitoring*, eds K. Roberts and M. Blades (Mahwah, NJ: Erlbaum Associates), 197–226.
- Roebbers, C. M., and Schneider, W. (2001). Individual differences in children’s eyewitness recall: the influence of intelligence and shyness. *Appl. Dev. Sci.* 5, 9–20. doi: 10.1207/S1532480XADS0501_2
- Schneider, W. (2015). *Memory Development From Early Childhood Through Emerging Adulthood*. NY: Springer. doi: 10.1007/978-3-319-09611-7
- Schreiber, N., Bellah, L. D., Martinez, Y., McLaurin, K. A., Strok, R., Garven, S., et al. (2006). Suggestive interviewing in the McMartin Preschool and Kelly Michaels daycare abuse cases: a case study. *Soc. Influence* 1, 16–47. doi: 10.1080/15534510500361739
- Scullin, M., Kanaya, T., and Ceci, S. J. (2002). Measurement of individual differences in children’s suggestibility across situations. *J. Exp. Psychol.* 8, 233–246. doi: 10.1037/1076-898X.8.4.233

- Scullin, M. H., and Ceci, S. J. (2001). A suggestibility scale for children. *Pers. Individ. Dif.* 30, 843–856. doi: 10.1016/S0191-8869(00)00077-5
- Talwar, V., and Crossman, A. (2011). From little white lies to filthy liars: the evolution of honesty and deception in young children. *Adv. Child Dev. Behav.* 40, 139–141. doi: 10.1016/B978-0-12-386491-8.00004-9
- Talwar, V., Crossman, A., Williams, S., and Muir, S. (2011). Adult detection of children's selfish and polite lies: experience matters. *J. Appl. Soc. Psychol.* 41, 2837–2857. doi: 10.1111/j.1559-1816.2011.00861.x
- Talwar, V., Gordon, H. M., and Lee, K. (2007a). Lying in the elementary school years: verbal deception and its relation to second-order belief understanding. *Dev. Psychol.* 43, 804–810. doi: 10.1037/0012-1649.43.3.804
- Talwar, V., Lavoie, J., Gomez-Garibello, C., and Crossman, A. M. (2017). The influence of social factors on the relation between lie-telling and children's cognitive abilities. *J. Exp. Child Psychol.* 159, 185–198. doi: 10.1016/j.jecp.2017.02.009
- Talwar, V., and Lee, K. (2002a). Emergence of white-lie-telling in children between 3 and 7 years of age. *Merrill Palmer Q.* 48, 160–181. doi: 10.1353/mpq.2002.0009
- Talwar, V., and Lee, K. (2002b). Development of lying to conceal a transgression: children's control of expressive behavior during verbal deception. *Int. J. Behav. Dev.* 26, 436–444. doi: 10.1080/01650250143000373
- Talwar, V., and Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Dev.* 79, 866–881. doi: 10.1111/j.1467-8624.2008.01164.x
- Talwar, V., Murphy, S. M., and Lee, K. (2007b). White lie-telling for children in politeness purposes. *Int. J. Behav. Dev.* 31, 1–11. doi: 10.1177/0165025406073530
- Thompson, W. C., Clarke-Stewart, K. A., and Lepore, S. J. (1997). What did the janitor do? Suggestive interviewing and the accuracy of children's accounts. *Law Human Behav.* 21, 405–426. doi: 10.1023/A:1024859219764
- Warren, A. R., Hulse-Trotter, K., and Tubbs, E. (1991). Inducing resistance to suggestibility in children. *Law Hum. Behav.* 15, 273–285. doi: 10.1007/BF01061713
- Wechsler, D. (1999). *Wechsler Intelligence Scale for Children-IV*. San Antonio, TX: Psychological Corp.
- Wellman, H. M. (1978). Knowledge of the interaction of memory variables: a developmental study of metamemory. *Dev. Psychol.* 14, 24–29. doi: 10.1037/0012-1649.14.1.24
- White, T. L., Leichtman, M. D., and Ceci, S. J. (1997). The good, the bad, and the ugly: accuracy, inaccuracy, and elaboration in preschoolers' reports about a past event. *Appl. Cogn. Psychol.* 11, S37–S54. doi: 10.1002/(SICI)1099-0720(199712)11:7<S37::AID-ACP546>3.0.CO;2-4
- Wyman, J., Foster, I., and Talwar, V. (2016). An experimental analysis of children's ability to provide a false report about a crime. *J. Visual. Expe.* 111:e53773. doi: 10.3791/53773
- Zaragoza, M. S., Payment, K. E., Ackil, J. K., Drivdahl, S. B., and Beck, M. (2001). Interviewing witnesses: forced confabulation and confirmatory feedback increase false memories. *Psychol. Sci.* 12, 473–477. doi: 10.1111/1467-9280.00388

Conflict of Interest: After the writing of this article but before publication, AH became employed by the law firm Shapiro Arato Bach LLP.

The remaining 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Hritz and Ceci. 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) and the copyright owner(s) 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.



False Memories in Native and Foreign Languages

Aleksandra Dolgoarshinnaia* and Beatriz Martin-Luengo*

Centre for Cognition and Decision Making, Institute for Cognitive Neuroscience, HSE University, Moscow, Russia

OPEN ACCESS

Edited by:

Hui Chen,
Zhejiang University, China

Reviewed by:

Nikolay Novitskiy,
The Chinese University of Hong
Kong, China
Julia Morales,
Loyola Andalusia University, Spain

*Correspondence:

Aleksandra Dolgoarshinnaia
adolgoarshinnaya@hse.ru
Beatriz Martin-Luengo
bmartinluengo@hse.ru

Specialty section:

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

Received: 28 May 2021

Accepted: 26 August 2021

Published: 28 September 2021

Citation:

Dolgoarshinnaia A and
Martin-Luengo B (2021) False
Memories in Native and Foreign
Languages.
Front. Psychol. 12:716336.
doi: 10.3389/fpsyg.2021.716336

Human memory is prone to memory errors and distortion. Evidence from studies on cognitive functions in bilinguals indicates that they might be prone to different types of memory errors compared to monolinguals; however, the effect of language in false memories is still understudied. Source monitoring processes required for proper memory functioning, presumably, rely on inhibitory control, which is also heavily utilized by bilinguals. Moreover, it is suggested that thinking in a second language leads to more systematic and deliberate reasoning. All these results lead to expect that bilinguals are more analytical when processing information in their second language overcoming some memory errors depending on the language of information. To test this hypothesis, we run a classical misinformation experiment with an explicit source monitoring task with a sample of Russian–English bilinguals. The language of the misinformation presentation did not affect the degree of the misinformation effect between the Russian and English languages. Source monitoring demonstrated an overall higher accuracy for attributions to the English source over the Russian source. Furthermore, analysis on incorrect source attributions showed that when participants misattributed the sources of false information (English or Russian narrative), they favored the Russian source over the not presented condition. Taken together, these results imply that high proficiency in the second language does not affect misinformation and that information processing and memory monitoring in bilinguals can differ depending on the language of the information, which seems to lead to some memory errors and not others.

Keywords: eyewitness memory, false memory, bilingualism, misinformation effect, source monitoring

INTRODUCTION

Globalization and worldwide migration lead to an increase in the percentage of the population that simultaneously speaks two and more languages. Evidence from studies on cognitive functions in bilinguals indicates that they might be prone to different types of memory errors compared to monolinguals (Marian and Neisser, 2000; Boroditsky et al., 2003, 2009, 2019; Fausey and Boroditsky, 2011). These individuals may face a situation where they have to testify in a foreign country using their second language to communicate. Unlike monolingual witnesses, they can be exposed to misleading information from sources in several languages. Results of the existing studies are not enough to lead to any sort of conclusion on how such situations might affect their memories and recollections. Moreover, the misinformation effect is rarely checked through explicit source monitoring tasks, as it is usually implied that if the misinformation effect is present, source confusion has occurred. Furthermore, to our knowledge, only few studies investigated possible implications in bilingual populations (Tosun et al., 2013; Ünal et al., 2016). However, these

studies investigated the effects of grammatically expressed evidentiality of the action agent. While undoubtedly important, this research is specific to particular languages with such expressed features (e.g., Turkish), and the effect in other languages remains unclear. Therefore, the present research aimed to investigate the effect the language of the presentation can have on memory errors, such as the misinformation effect and source misattributions.

The misinformation paradigm is one of the available research designs that enable the research on false memories and misinformation effect (Loftus and Palmer, 1974; Loftus, 1975; Loftus and Guido, 1975; Loftus and Hoffman, 1989). In short, the misinformation effect occurs when postevent false information is wrongly reported as the original. In the laboratory, the misinformation paradigm is operationalized as follows: witnessing the original event (Phase 1), exposure to false information about the event (Phase 2), and memory test to measure the acceptance of false information (Phase 3). As a result, participants report that they remember false information from witnessing the event rather than the later sources. The degree of memory distortion and the resulting false memories can vary due to a variety of factors. Susceptibility to misinformation was found to increase with age (Ceci et al., 1987; Loftus, 2005) and the time passed between witnessing the event, exposure to false information, and recollection of the event (Loftus, 2005). On the other hand, misinformation acceptance can decrease if people are warned that they have received or will receive information that is not entirely correct (Echterhoff et al., 2005; Loftus, 2005). Finally, the nature of the details is also important; schema-consistent false information is easier accepted than schema-inconsistent or schema irrelevant information (Tuckey and Brewer, 2003); furthermore, memory for the central information is generally better remembered than for the peripheral (Luna and Migueles, 2009); however, it can also differ depending on the emotional arousal and valence of the situation and information (Christianson and Loftus, 1991; Porter et al., 2003).

One of the most accepted explanations for the misinformation effect is “source monitoring errors,” that is, errors in the attribution of the information (Lindsay et al., 1991; Johnson et al., 1993; Johnson, 1997; Luna and Martín-Luengo, 2013). There are two key mechanisms underlying the failures in source discrimination. First, depending on the quality of the available information (perceptual cues and schema consistency or inconsistency), the credibility of the memory is either defined through rapid and heuristic processing or requires more deliberate and systematic processing (Johnson et al., 1993). It is considered that source attributions are mainly made heuristically (“System-1” processing), whereas systematic reasoning (“System-2” reasoning) is engaged to a lesser degree, but both can be activated for “more careful” judgments. The decision criteria that are used in both, the System-1 and System-2 processes differ, and the criterion for the latter is stricter. Thus, when both the heuristic and systematic processes are used, the criterion tightens, and better judgments are made. Generally, the criteria are rather flexible, can be influenced by biases, meta-memory assumptions, and vary across different types of events and situations (Johnson et al., 1993). Second,

research on suggestibility and reality monitoring in children revealed that inhibitory control can be another important mechanism underlying the source monitoring processes. Several studies showed that younger children experience difficulties when performing source monitoring tasks, and the performance usually improves with age (Foley et al., 1983). Neuroscientific findings associate source monitoring processes to activations in the prefrontal cortex (Johnson et al., 1993; Ruffman et al., 2001), which is associated with executive functioning including the inhibitory control (Waltz et al., 1999). Developmentally, the prefrontal cortex matures later compared to the other regions of the brain (Diamond, 2002), and its maturation is correlated with the ability of the children to inhibit task-irrelevant or competing for information (Sinopoli and Dennis, 2012). Moreover, several studies investigated the relationship between these processes directly by testing the source memory performance and measuring the level of inhibitory control (Ruffman et al., 2001). The results of these studies indicate that greater inhibitory control is positively correlated with better performance on source monitoring tasks.

The effect of language on memory, specifically, episodic memory was repeatedly observed within and between the languages. Many studies investigated the effect of wording in leading questions similar to those asked witnesses during interrogations (Loftus and Palmer, 1974; Loftus, 1975; Loftus and Guido, 1975). These studies demonstrated that even slight lexical variations in questions, such as the usage of synonyms can remarkably influence the answer of an individual and the recollection of an event. Also, the linguistic influence on the memory of an individual remains persistent across different language groups as shown by further studies. For example, Fausey and Boroditsky, 2011 compared English and Spanish as well as English and Japanese monolinguals in their recall of intentional and accidental events and the agent who acted in these events. Observed cross-linguistic differences in memory performance were described as caused by the different grammar patterns specific to these languages, e.g., more frequent use of an agent in the English language compared to Japanese or Spanish. Similar studies also explored the effect in several other languages, including Turkish (Aydin and Ceci, 2013) and Indonesian (Fausey and Boroditsky, 2011), as well as different linguistic characteristics, such as the usage of definite or indefinite articles (Loftus, 1975), grammar tenses (Boroditsky et al., 2019), or gender (Boroditsky et al., 2003), all showing variations in the memory performance attributed to the specific linguistic characteristics.

These findings and a general rise of interest in the cognitive functions in bi and multilingual populations sparked the interest in memory processing and memory mistakes in individuals who utilize more than one language. Providing a mental representation for the event, language can be labeled as a contextual cue which affects the way information is encoded and retrieved, implying that in bilinguals, the memory of an event can be better when encoded and retrieved in the same language as opposed to situations when languages are inconsistent (Schroeder and Marian, 2014). Furthermore, it has been long argued that linguistic characteristics, being specific to

a particular language, can shape the thoughts and behavior of an individual (Schroeder and Marian, 2014). As per the thinking-for-speaking concept proposed by Slobin, when speaking, a person can direct his attention to particular details through syntactic structures that are established by the language he speaks (Slobin, 1987). This implies that the structure of the language and the cultural representations associated with this language can create a certain perspective through which a person processes information. When bilinguals are concerned, it also means that encoding can be different depending on the language they use at the moment. Yet, while episodic and autobiographical memory in bilinguals has been widely researched and differences in encoding and retrieval processes were observed (Boroditsky et al., 2009; Fausey and Boroditsky, 2011; Aydin and Ceci, 2013), there have been only a handful of studies investigating the suggestibility in bilinguals (Shaw et al., 1997; Smith et al., 2017; Calvillo and Mills, 2020). Although misinformation effect was present in all these cross-linguistic studies, the results regarding the influence of the language on the endorsement of false information show inconsistent findings, reporting no significant differences between the conditions (Shaw et al., 1997) or explaining the results in relation to the different levels of proficiency between the languages (Calvillo and Mills, 2020). At the same time, the potential effects of proficiency levels of the second language were not yet explored. All of this, therefore, demonstrates the need for further research of the phenomenon.

Research on source monitoring has also advanced in the last decade. Nevertheless, it is still rarely explicitly tested in the misinformation paradigm, as the presence of the effect implies that source confusion has occurred. Moreover, it is still unclear whether source monitoring processes can be affected by language and bilingualism. Several studies (Tosun et al., 2013; Ünal et al., 2016) investigated the effects of grammatically expressed agency; however, as this grammar feature is present in some languages and not others, the results cannot be generalized. Thus, language influence on source monitoring processes requires further examination.

In sum, research on the effects of bilingualism on cognitive functions, such as decision-making, attention, and memory processing suggests that bilinguals may have certain advantages in performing non-linguistic tasks due to their more trained inhibitory control because of the selection of languages compared to monolinguals (Bialystok et al., 2005; Weissberger et al., 2015). Furthermore, bilinguals could rely on more analytical and deliberate System-2 processing as opposed to heuristic System-1 (Caldwell-Harris, 2014; Costa et al., 2014; Hayakawa et al., 2016, 2017; Bialek et al., 2019). Therefore, we suggested that bilinguals would be less susceptible to misleading details in their second language, as their monitoring mechanisms would be more engaged in information processing. Thus, the misinformation effect would be reduced. Following this hypothesis, we expected a lower level of accuracy on the recognition task in the native language compared to the foreign language. Alternatively, we also hypothesized that bilinguals could be more susceptible to misleading information in their second language, as their cognitive resources would be more engaged with the linguistic processing of the information,

therefore leaving fewer resources for monitoring the source of this information. In this case, we could expect a lower level of accuracy on the recognition task in foreign, rather than in the native language. Our second expectation was that due to a higher degree of misinformation effect in the native language, a higher confidence rate for incorrect answers on the recognition task for the native language would be observed. More expectations were made regarding the source-monitoring judgments. Based on the previous expectations of observable misinformation effect, we expected to see lower accuracy for misleading items in the native language compared to foreign for the source monitoring task, as the presence of the misinformation effect implies that memory for misleading information is treated as a memory for the original event. In relation to that, we expected higher confidence ratings for these items on the source monitoring task. In the case of the alternative hypothesis, the pattern should be reversed with lower accuracy for misleading items in a foreign language. To test these hypotheses, we run an experiment using the misinformation paradigm with the classical three phases: presentation of the original information (no sound video), exposure to misleading information (narratives describing the event of the video in different languages), and memory test in this case with an explicit source monitoring task.

MATERIALS AND METHODS

Participants

A total of 70 bilingual volunteers (46 women, mean age = 24.2, $SD = 4.63$) recruited through social media took part in this online study, created, and hosted in the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants received a small monetary reward (250 rubles) for the successful completion of the experiment. *A priori* analysis conducted in GPower (version 3.1) determined the sample size as 48 participants (effect size = 0.40, power = 0.80, and alpha = 0.05). The selection requirements for the participants included having Russian as the native language, speaking intermediate or higher level of English proficiency (according to the Common European Framework of Reference, CEFR), having no prior experience of immersion in an entirely English-speaking environment for a prolonged period of time (more than 5 months) (Costa et al., 2014). All the participants completed an online Cambridge Assessment test (www.cambridgeenglish.org), and nine participants who scored below 16 points (Pre-Intermediate and lower, CEFR) on the Cambridge test were excluded, leaving mainly high-proficient participants (mean proficiency = 20 out of 25 points, $SD = 2.9$). Five more were excluded due to the violation of the English-speaking environment requirement. The final sample reported in this study consisted of 56 participants (40 women, mean age = 24.1, $SD = 4.66$).

Normative Study and Materials

As the main stimulus, we used 61-s real footage of a car robbery (as shown in **Figure 1** for the video outline). First, true items were extracted directly from the contents of the video (11). Then, false items (20) were created that either distorted the information of the video (for example, white T-shirt instead of black) or



FIGURE 1 | The temporal sequence of footage of a car robbery. The video is recorded from a camera in a car parked at the side of the road. It captures a gray car parked in front of the car with a camera. A young man wearing a white T-shirt leaves the gray car and enters the bushes on the side of the road. At the same time, a second man leaves the car but stays beside it looking around. The young man then reappears from the bushes hiding something behind his back and heads toward one of the cars parked in front. After a while, he returns carrying several plastic bags. Both men then enter the gray car and drive away.

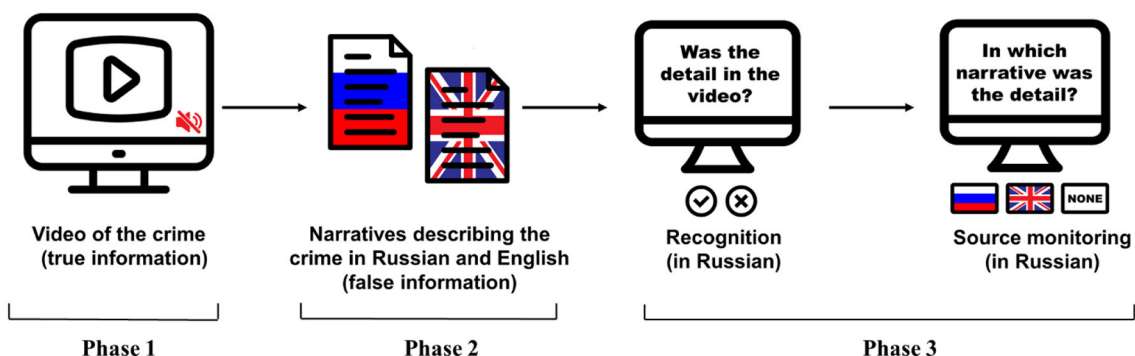


FIGURE 2 | Experimental design. Phase 1: encoding of true information from a video. Phase 2: introduction of misinformation from two narratives (Russian and English). Phase 3: true/false recognition + source monitoring (Russian narrative, English narrative, none).

did not appear in the video but were plausible to happen in that environment.

Before the main experiment, we ran a normative study to guarantee equal memorability to both true and false items. The normative study was conducted in Russian, with 20 volunteers (15 women, mean age = 25, and $SD = 3.34$), who did not take part in the main experiment. Participants of the normative study were not required to watch the video; instead, they read two narratives describing the event in the video. Each narrative

contained 10 critical details, which were considered true for the normative study and were going to be used as misleading in the main experiment. To check the memory of these details, 11 false control items were used on the recognition task. These false items were originally identified as critical true elements in the video for the main experiment. The narratives were presented separately for the two groups of participants, one narrative for one group (each group consisting of 10 participants). This separation in two groups is based on the counterbalance of the elements in

the design of the main experiment as they should not have referred to the same object, person, or modality, and should not have interfered with each other in one set of the narratives. Procedurally, the participants were required to read a narrative, and after a distractor task to perform a recognition test, they had to indicate whether a particular piece of information was present in the text or not (True or False) and rate their confidence in their answer on a 1 (not confident at all) to 10 (totally confident) scale. The normative study confirmed a similar memorability of both true and false items ($p > 0.05$), which allowed us to use all 32 previously identified items in the further experiment.

For the main experiment, a total of 31 details were used (**Supplementary Appendix A**). The true items (11) were the same for all the conditions. Counterbalancing false items (20), some of them were used as misleading information introduced during the misinformation stage and some were used as false control items on the recognition task. For introducing misleading information, we created four pairs of narratives in Russian and English to counterbalance the misleading and control details as well as the language in which they were presented. Each narrative contained five misleading details (making it 10 for a set).

Procedure

Participants accessed the experiment *via* a link shared by the experimenter. They read and explicitly stated their agreement with an electronic consent form before starting the experiment. Then they filled in a social and linguistic questionnaire adopted from Marian et al. (2007). After that, the participants were randomly assigned to one of the four counterbalanced conditions and completed a misinformation paradigm consisting of the following stages: witnessing the event, encountering misleading information, recognition test, and explicit source monitoring test (**Figure 2**).

First, the participants watched the 1-min long video of a car robbery. Immediately after it, they were asked to perform a series of distracting numeric tasks that took ~4 min to complete. Upon completing these numeric tasks, the participants read a pair of narrative accounts of what took place in the video. The narratives were described to the participants as reports made by two witnesses of the crime, a Russian native speaker and his friend who happened to be an English native, reported in the corresponding languages. Each of the narratives consisted of five misleading details about the event. Then the participants performed another series of distracting numeric tasks analogous to the first one and of the same duration. Next, the participants completed a recognition test along with the source-monitoring test. For the recognition test, they indicated whether the detail or item described in a statement was present in the original video (true or false) and indicated the level of confidence in their answer on a scale from 1 (not confident at all) to 10 (totally confident). Immediately after the recognition question, the participants performed a source monitoring task for which they had to indicate whether the detail or item from the previous question was mentioned in the Russian or English narrative, or it did not appear in any of the narratives (for true and false control items). Participants rated their confidence for their answers on this question again from 1 (not confident at

all) to 10 (totally confident). Finally, the participants had to complete a general English proficiency test adapted from an online Cambridge Assessment.

Statistical Analysis

Misinformation

Typically, the misinformation effect is checked and calculated as a difference in the proportion of false alarm rates of misleading and control items reflecting the acceptance of misinformation. Due to the specificity of design and materials, in particular, the absence of control items that would fully match the conditions (i.e., there were no control English items), a 2×2 ANOVA was not feasible. Instead, we made direct comparisons on the proportions of false alarms for misleading and control items by *t*-tests.

Source Monitoring

To analyze the performances on source monitoring task and corresponding confidence ratings, we ran a 3×3 repeated measures ANOVA with correct and incorrect source attributions of the participants (Russian narrative, English narrative, and None, where information was not presented in the narratives). We then investigated the differences within these factors by running a repeated measures ANOVA for correct source attributions (Russian-to-Russian, English-to-English, and None-to-None) and Student's *t*-tests for incorrect source attributions (Russian-to-English, Russian-to-None, English-to-Russian, English-to-None, None-to-Russian, and None-to-English).

RESULTS

Below are reported analyses of ANOVA on the recognition and source monitoring tasks and their corresponding confidence ratings. When appropriate two-tailed pairwise comparisons with the Student's *t*-test were used. For pairwise comparisons, we applied Bonferroni correction to avoid Type I error for conducting multiple statistical tests. Partial-eta squared (η^2_p) and Cohen's *d* (*d*) are also reported as the measures of effect size.

Misinformation

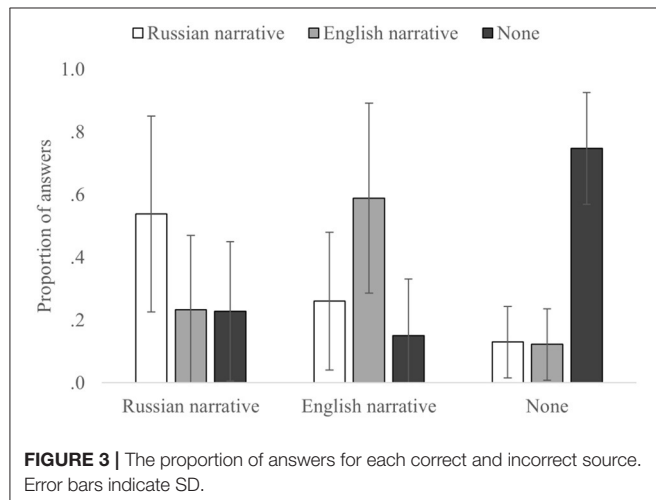
First, the misinformation effect was confirmed since the proportion of false alarms for misleading information ($M = 0.32$, $SD = 0.19$) were significantly higher than for the control items ($M = 0.21$, $SD = 0.12$), $t_{(55)} = 5.134$, $p < 0.001$, and $d = 0.686$. Moreover, mean confidence ratings were higher for misleading ($M = 24.34$, $SD = 16.70$), than for the control items ($M = 14.77$, $SD = 10.37$), $t_{(55)} = 4.392$, $p < 0.001$, and $d = 0.587$. However, misinformation effect did not differ between the Russian ($M = 0.13$, $SD = 0.11$) and English languages ($M = 0.13$, $SD = 0.11$), $t_{(55)} = 1.169$, $p = 0.866$, and $d = 0.023$.

Source Monitoring

First, a 3×3 ANOVA on the correct and incorrect source attributions of the participants (**Table 1**, **Figure 3**) showed no main effect for the correct source, $F_{(2,110)} = 0.003$, $p = 0.999$, and $\eta^2_p = 0.00006$; however, it showed main effect for the incorrect source, $F_{(2,110)} = 3.910$, $p = 0.023$, and $\eta^2_p = 0.066$, and their interaction, $F_{(4,220)} = 73.580$, $p < 0.0001$, and $\eta^2_p = 0.572$.

TABLE 1 | Means (SD) of correct and incorrect source attributions.

Actual source	Participants' attributions		
	Russian narrative	English narrative	None
Russian narrative	0.54 (0.31)	0.23 (0.23)	0.23 (0.22)
English narrative	0.26 (0.21)	0.59 (0.30)	0.15 (0.18)
None	0.13 (0.11)	0.12 (0.10)	0.75 (0.18)

**FIGURE 3** | The proportion of answers for each correct and incorrect source. Error bars indicate SD.

For the correct source attributions (diagonal in **Table 1**), we found significant differences, $F_{(2,110)} = 12.250$, $p < 0.0001$, and $\eta^2_p = 0.183$, None-to-None attributions were significantly higher than Russian-to-Russian [$t_{(55)} = -5.733$, $p < 0.001$, and $d = -0.766$] and English-to-English [$t_{(55)} = -3.890$, $p < 0.001$, and $d = -0.520$]. No significant differences were found in correct attributions between Russian-to-Russian and English-to-English [$t_{(55)} = -0.958$, $p = 0.342$, and $d = -0.128$].

Analysis on incorrect attributions for the Russian source (when Russian narrative was the correct source), showed that the differences between Russian-to-English [$t_{(55)} = 4.493$, $p < 0.001$, and $d = 0.600$] and Russian-to-None [$t_{(55)} = 4.766$, $p < 0.001$, and $d = 0.637$] source misattributions were not significant [$t_{(55)} = 0.131$, $p = 0.896$, and $d = 0.018$]. Similarly, for the None source (when information was not presented in the narratives), differences between None-to-Russian [$t_{(55)} = 16.403$, $p < 0.001$, and $d = 2.192$] and None-to-English [$t_{(55)} = 17.522$, $p < 0.001$, and $d = 2.341$] source misattributions were also insignificant [$t_{(55)} = 0.442$, $p = 0.639$, and $d = 0.063$]. On the other hand, for the English source (when the English narrative was the correct source), the proportion of English-to-Russian [$t_{(55)} = 4.939$, $p < 0.001$, and $d = 0.660$] misattributions was higher than that of the English-to-None [$t_{(55)} = 7.323$, $p < 0.001$, and $d = 0.979$] misattributions [$t_{(55)} = 3.129$, $p = 0.003$, and $d = 0.418$].

To summarize, these results revealed several patterns for source attributions in the misinformation paradigm. For correct source attributions, they were significantly higher than incorrect source attributions in all the conditions; however, there was no

significant difference between the Russian and English sources. Further analyses on incorrect attributions showed no significant results for the Russian and None sources, but not for the English source misattributions to the Russian source, which were significantly higher than the English source misattributions to the None source.

Source Monitoring Confidence

First, a 3×3 ANOVA for confidence ratings of correctly and incorrectly attributed sources (**Table 2**, **Figure 4**) showed significant main effect on confidence when the source was identified correctly, $F_{(2,110)} = 18.501$, $p < 0.001$, and $\eta^2_p = 0.252$, or incorrectly, $F_{(2,110)} = 4.357$, $p = 0.015$, and $\eta^2_p = 0.073$, as well as their interaction, $F_{(4,220)} = 100.599$, $p < 0.00001$, and $\eta^2_p = 0.647$.

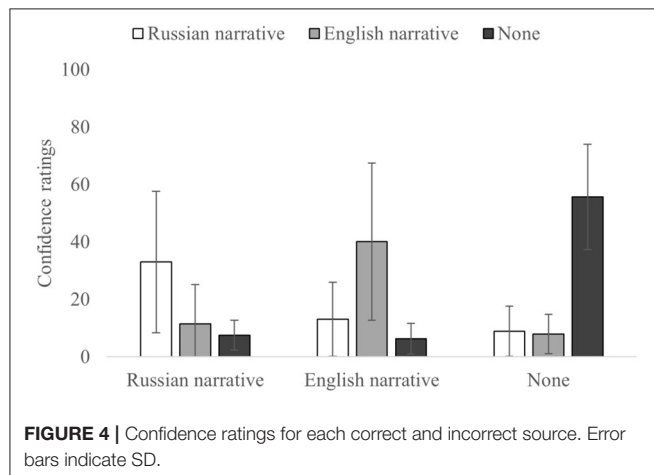
For the correct source attributions (diagonal in **Table 2**), there were differences, $F_{(2,110)} = 17.569$, $p < 0.0001$, and $\eta^2_p = 0.242$. In particular, the confidence for None-to-None attributions was higher than for Russian-to-Russian [$t_{(55)} = -7.109$, $p < 0.001$, and $d = -0.950$] and English-to-English [$t_{(55)} = -4.013$, $p < 0.001$, and $d = -0.536$]. At the same time, there was no difference in confidence ratings for correct source attributions between Russian-to-Russian and English-to-English [$t_{(55)} = -1.557$, $p = 0.125$, and $d = -0.208$].

Analysis on confidence ratings for the incorrect attributions for the Russian source (when the Russian narrative was the correct source) demonstrated that differences in confidence between Russian-to-English [$t_{(55)} = 5.058$, $p < 0.001$, and $d = 0.676$] and Russian-to-None [$t_{(55)} = 6.716$, $p < 0.001$, and $d = 0.897$] source misattributions were not significant [$t_{(55)} = 1.867$, $p = 0.067$, $d = 0.250$]. Similarly, for the None source (when information was not presented in the narratives), differences in confidence between None-to-Russian [$t_{(55)} = 14.494$, $p < 0.001$, and $d = 1.937$] and None-to-English [$t_{(55)} = 15.932$, $p < 0.001$, and $d = 2.129$] source misattributions were also not significant [$t_{(55)} = 0.760$, $p = 0.451$, and $d = 0.102$]. Finally, for the English source (when English narrative was the correct source), confidence ratings were higher for the English-to-Russian [$t_{(55)} = 5.974$, $p < 0.001$, and $d = 0.798$] and the English-to-None [$t_{(55)} = 8.088$, $p < 0.001$, and $d = 1.081$] misattributions, [$t_{(55)} = 3.540$, $p < 0.001$, and $d = 0.473$].

To sum up, the results of the analyses on confidence complemented the results of source monitoring analyses and showed similar patterns. First, confidence for correct source attributions was higher than that for incorrect source attributions for all the conditions. Second, there was no significant difference in the confidence between correct attributions for the Russian

TABLE 2 | Means (SD) of confidence ratings for correct and incorrect source attributions.

Actual source	Participants' attributions		
	Russian narrative	English narrative	None
Russian narrative	33.0 (24.61)	11.39 (13.69)	7.54 (5.21)
English narrative	13.04 (12.88)	40.07 (27.38)	6.28 (5.33)
None	8.86 (8.74)	7.93 (6.81)	55.65 (18.36)



and English sources. Finally, analyses on the confidence for misattributions showed no significant differences for the Russian and None sources; however, for the English source, the confidence was higher when incorrect attributions were made in favor of the Russian source rather than the None source, which fully mirrors the results of the source monitoring analyses.

Additional Analysis by the Level of Proficiency

One of the main criteria for the participants in this study was intermediate or higher levels of English language proficiency, which was necessary to ensure that the participants would not experience any difficulties in understanding the materials. Nevertheless, in an attempt to further explore how different levels of language proficiency can affect memory performance for misinformation and source monitoring, we ran an additional analysis between the groups of participants with intermediate (scoring from 16 to 19) and higher levels of proficiency (scoring from 20 to 25). Each group consisted of 28 participants. Mean proficiency for the Intermediate group was 17.39 ($SD = 1.1$) out of 25 points, and for the High proficient group, it was 22.6 ($SD = 1.5$) points out of 25.

Analysis on the false alarm rates (as shown in **Table 3**), confirmed the acceptance of misinformation in both the groups [Int: $t_{(27)} = 4.740$, $p < 0.001$, and $d = 0.896$; High: $t_{(27)} = 2.748$, $p = 0.011$, and $d = 0.519$]. The degree of misinformation accepted from the Russian and English languages did not differ within the groups [Int: $t_{(27)} = 1.622$, $p = 0.116$, and $d =$

0.307; High: $t_{(27)} = -1.451$, $p = 0.158$, and $d = -0.274$]. However, comparison between the level of groups revealed that the degree of misinformation acceptance from the English language was significantly higher in the High proficient group compared to the Intermediate group, $t_{(27)} = 2.460$, $p = 0.021$, and $d = 0.465$.

Further analysis on the source monitoring within the proficiency groups demonstrated patterns similar to the results of the main analysis with the full sample (**Table 4**, **Figures 5, 6**). In both the groups, 3×3 ANOVA with the correct and incorrect source attributions of the participants showed no main effect for the correct source [Int: $F_{(2,54)} = 2.681$, $p = 0.078$, and $\eta^2_p = 0.090$; High: $F_{(2,54)} = 0.458$, $p = 0.635$, and $\eta^2_p = 0.017$] and incorrect source [Int: $F_{(2,54)} = 1.940$, $p = 0.154$, and $\eta^2_p = 0.067$; High: $F_{(2,54)} = 1.999$, $p = 0.145$, and $\eta^2_p = 0.069$]; at the same time, their interaction was significant [Int: $F_{(2,54)} = 28.700$, $p < 0.0001$, and $\eta^2_p = 0.515$; High: $F_{(2,54)} = 46.088$, $p < 0.0001$, and $\eta^2_p = 0.631$].

For the correct source attributions (diagonal in **Table 4**), for both the groups, None-to-None attributions were significantly higher than the Russian-to-Russian [Int: $t_{(27)} = -3.682$, $p = 0.001$, and $d = -0.501$; High: $t_{(27)} = -4.342$, $p < 0.001$, and $d = -0.821$] and English-to-English attributions [Int: $t_{(27)} = -2.649$, $p = 0.013$, and $d = -0.501$; High: $t_{(27)} = -2.827$, $p = 0.009$, and $d = -0.534$]. No significant differences were found in correct attributions between the Russian-to-Russian and English-to-English [Int: $t_{(27)} = -0.284$, $p = 0.779$, and $d = -0.054$; High: $t_{(27)} = -1.176$, $p = 0.250$, and $d = -0.222$].

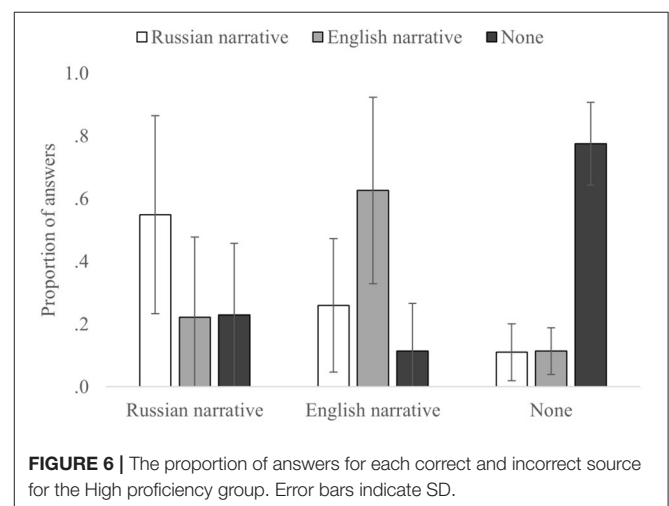
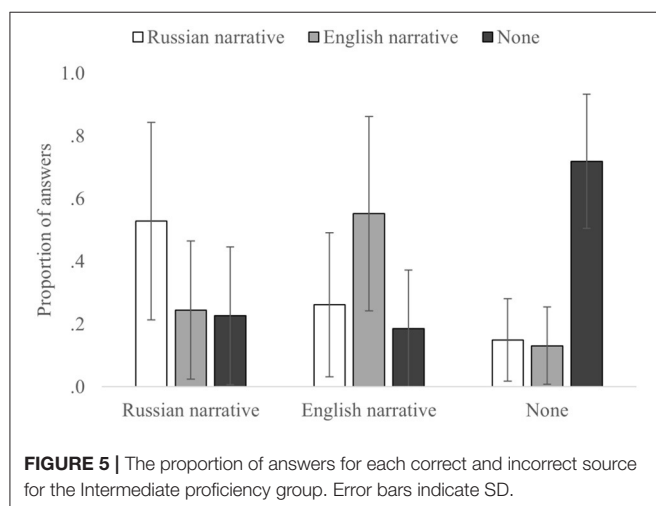
As for the incorrect source attributions, both the groups mostly followed the same patterns revealed by the main analyses; however, some minor differences were also found. For misattributions for the Russian source (when Russian narrative was the correct source), there was no significant difference between the Russian-to-English [Int: $t_{(27)} = 3.021$, $p = 0.005$, and $d = 0.571$; High: $t_{(27)} = 3.285$, $p = 0.003$, and $d = 0.621$] and Russian-to-None [Int: $t_{(27)} = 3.213$, $p = 0.003$, and $d = 0.607$; High: $t_{(27)} = 3.456$, $p = 0.002$, and $d = 0.653$] source misattributions [Int: $t_{(27)} = 0.306$, $p = 0.761$, and $d = 0.058$; High: $t_{(27)} = -0.113$, $p = 0.911$, and $d = -0.021$]. Similarly, for the None source (when information was not presented in the narratives), the difference between None-to-Russian [Int: $t_{(27)} = 9.050$, $p < 0.001$, and $d = 1.710$; High: $t_{(27)} = 16.526$, $p < 0.001$, and $d = 3.123$] and None-to-English [Int: $t_{(27)} = 9.613$, $p < 0.001$, and $d = 1.817$; High: $t_{(27)} = 18.023$, $p < 0.001$, and $d = 3.406$] source misattributions was also not significant [Int: $t_{(27)}$

TABLE 3 | Means (SD) of false alarm rates for control and misleading items for Intermediate and High proficient groups.

Level of proficiency	FAs control	FAs misleading		
		Overall	Russian origin	English origin
Intermediate	0.18 (0.12)	0.32 (0.21)	0.14 (0.12)	0.09 (0.11)
High proficient	0.23 (0.12)	0.33 (0.17)	0.13 (0.10)	0.17 (0.11)

TABLE 4 | Means (SD) of correct and incorrect source attributions for Intermediate (Int) and High proficiency groups.

Actual source	Participants' attributions					
	Russian narrative		English narrative		None	
	Int	High	Int	High	Int	High
Russian narrative	0.53 (0.31)	0.55 (0.32)	0.25 (0.22)	0.22 (0.25)	0.23 (0.22)	0.23 (0.23)
English narrative	0.26 (0.23)	0.26 (0.21)	0.55 (0.31)	0.63 (0.30)	0.19 (0.19)	0.11 (0.15)
None	0.15 (0.13)	0.11 (0.09)	0.13 (0.12)	0.11 (0.08)	0.72 (0.21)	0.78 (0.13)



$= 0.683$, $p = 0.500$, and $d = 0.129$; High: $t_{(27)} = -0.208$, $p = 0.837$, and $d = -0.039$]. However, for the English source (when English narrative was the correct source), English-to-Russian misattributions [Int: $t_{(27)} = 3.026$, $p = 0.005$, and $d = 0.572$; High: $t_{(27)} = 4.521$, $p < 0.001$, and $d = 0.854$] were significantly higher than English-to-None misattributions [Int: $t_{(27)} = 5.128$, $p < 0.001$, and $d = 0.780$; High: $t_{(27)} = 6.993$, $p < 0.001$, and $d = 1.322$] only in High proficient group [Int: $t_{(27)} = 1.313$, $p = 0.200$, and $d = 0.248$; High: $t_{(27)} = 3.993$, $p < 0.001$, and $d = 0.755$].

Analysis on the confidence ratings of correctly and incorrectly attributed sources (Table 5; Figures 7, 8) showed the main effect for the correct source [Int: $F_{(2,54)} = 13.655$, $p < 0.001$, and $\eta^2_p = 0.336$; High: $F_{(2,54)} = 7.280$, $p = 0.002$, and $\eta^2_p = 0.0212$]. However, only the Intermediate group demonstrated the main effect for the incorrect source [Int: $F_{(2,54)} = 3.964$, $p = 0.025$, and $\eta^2_p = 0.128$; High: $F_{(2,54)} = 1.673$, $p = 0.197$, and $\eta^2_p = 0.058$]. Meanwhile, the interaction was shown to be significant for both

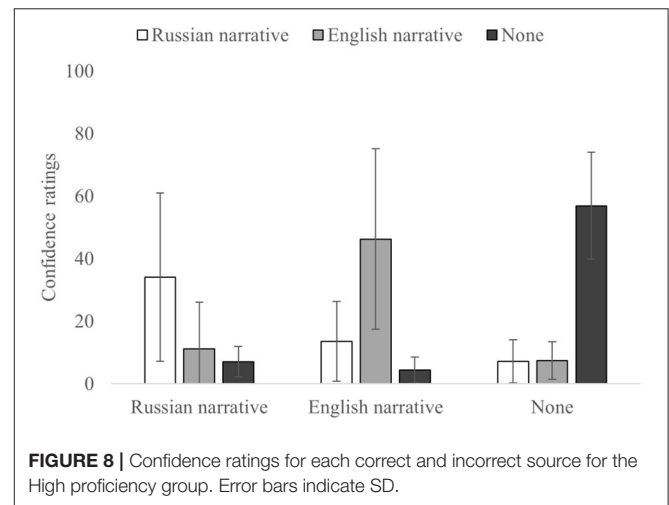
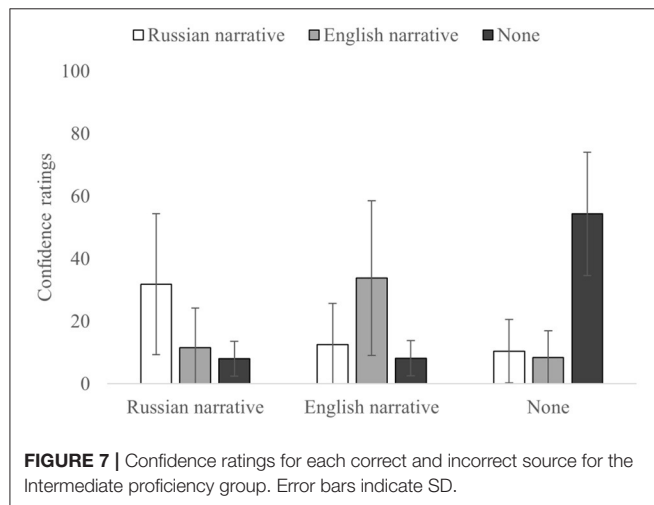
the groups [Int: $F_{(2,54)} = 42.146$, $p < 0.0001$, and $\eta^2_p = 0.610$; High: $F_{(2,54)} = 59.860$, $p < 0.0001$, and $\eta^2_p = 0.689$].

For the correct source attributions (diagonal in Table 5), for both the groups, confidence in None-to-None attributions was significantly higher than the Russian-to-Russian [Int: $t_{(27)} = -6.097$, $p < 0.001$, and $d = -1.152$; High: $t_{(27)} = -4.329$, $p < 0.001$, and $d = -0.818$]; however, confidence in None-to-None attributions was higher than for the English-to-English only in the Intermediate group [Int: $t_{(27)} = -3.891$, $p < 0.001$, and $d = -0.735$; High: $t_{(27)} = -1.889$, $p = 0.070$, and $d = -0.357$]. No significant differences were found for confidence in correct attributions between the Russian-to-Russian and English-to-English [Int: $t_{(27)} = -0.325$, $p = 0.748$, and $d = -0.061$; High: $t_{(27)} = -1.826$, $p = 0.079$, and $d = -0.345$].

Confidence for misattributions complemented previous results within the proficiency groups and mostly followed the patterns of the main analysis on confidence. For the Russian source (when Russian narrative was the correct source), in both

TABLE 5 | Means (SD) of confidence ratings for correct and incorrect source attributions for Intermediate (Int) and High proficiency groups.

Actual source	Participants' attributions					
	Russian narrative		English narrative		None	
	Int	High	Int	High	Int	High
Russian narrative	31.86 (22.50)	34.14 (26.90)	11.57 (12.64)	11.21 (14.90)	8.04 (5.55)	7.04 (4.91)
English narrative	12.50 (13.25)	13.57 (12.72)	33.86 (24.73)	46.29 (28.89)	8.23 (5.66)	4.38 (4.24)
None	10.47 (10.17)	7.19 (6.83)	8.47 (7.61)	7.4 (6.0)	54.35 (19.75)	56.94 (17.11)



the groups differences in confidence between Russian-to-English [Int: $t_{(27)} = 3.857$, $p < 0.001$, and $d = 0.732$; High: $t_{(27)} = 3.351$, $p = 0.002$, and $d = 0.633$] and Russian-to-None [Int: $t_{(27)} = 4.741$, $p < 0.001$, and $d = 0.896$; High: $t_{(27)} = 4.710$, $p < 0.001$, and $d = 0.890$] source misattributions were not significant [Int: $t_{(27)} = 1.238$, $p = 0.226$, and $d = 0.234$; High: $t_{(27)} = 1.376$, $p = 0.180$, and $d = 0.260$]. Similarly, for the None source (when information was not presented in the narratives), difference in the confidence ratings between None-to-Russian [Int: $t_{(27)} = 8.728$, $p < 0.001$, and $d = 1.649$; High: $t_{(27)} = 12.216$, $p < 0.001$, and $d = 2.309$] and None-to-English [Int: $t_{(27)} = 9.527$, $p < 0.001$, and $d = 1.800$; High: $t_{(27)} = 13.694$, $p < 0.001$, and $d = 2.588$] source misattributions were also not significant for both the groups [Int: $t_{(27)} = 1.089$, $p = 0.286$, and $d = 0.206$; High: $t_{(27)} = -0.136$, $p = 0.893$, and $d = -0.026$]. However, for the English source (when the English narrative was the correct source), confidence for English-to-Russian [Int: $t_{(27)} = 3.488$, $p = 0.002$, and $d = 0.659$; High: $t_{(27)} = 4.956$, $p < 0.001$, and $d = 0.937$] misattributions was significantly higher than for the English-to-None [Int: $t_{(27)} = 4.761$, $p < 0.001$, and $d = 0.900$; High: $t_{(27)} = 6.878$, $p < 0.001$, and $d = 1.300$] misattributions in the High proficient group [Int: $t_{(27)} = 1.570$, $p = 0.128$, and $d = 0.297$; High: $t_{(27)} = 3.498$, $p = 0.002$, and $d = 0.661$].

To sum up, exploratory analyses on misinformation and source monitoring within the groups of participants with Intermediate and High levels of proficiency, overall,

followed the patterns revealed by the main analyses. Regarding misinformation, results, on the one hand, confirmed the misinformation effect in both the groups and no difference in the degree of the effect depending on the language of misinformation presentation within the groups. On the other hand, between groups analysis revealed higher misinformation acceptance coming from the English narrative for more proficient participants. As for source monitoring, correct source attributions were significantly higher than the incorrect source attributions in all the conditions; however, there was no significant difference between the Russian and English sources. Analyses for incorrect attributions showed no significant differences between the Russian and None sources; however, the English-to-Russian source misattributions were significantly higher than the English-to-None misattributions only in the High proficient group. Confidence ratings complemented these patterns and for the most part, showed similar results as in the main analysis on confidence.

DISCUSSION

This experiment aimed to examine the possible effects of language on memory errors. To investigate this relationship, we used the misinformation paradigm and explicit source monitoring task. Our main expectations were that the participants would be better, or, worse at rejecting misleading

information that was presented in their second language due to increased, or, decreased cognitive processing of this information. Analyses showed that the misinformation effect was present; however, there was no difference in the degree of misinformation acceptance between the Russian and English narratives. Other studies on the misinformation effect in bilinguals did not either report significant differences between the conditions (Shaw et al., 1997) or attributed it to language proficiency (Calvillo and Mills, 2020). Specifically, Calvillo and Mills (2020) reported an increased misinformation effect in less proficient language. This study had some limitations; however, as the sample was not properly balanced between the English-Spanish and Spanish-English bilinguals, and what is more important, the measure of language proficiency was performed in the form of object-naming task and self-assessment. Although self-assessment of the abilities of the participants can be useful, it can be argued that object-naming tasks could not be an entirely reliable measure of proficiency when participants have to process information in more complex structures, such as texts. Therefore, a more reliable measure is necessary to ensure that the observed effect is not caused by a lack of linguistic command. In the current experiment, we used objective measures as the main measure of second language proficiency. All the participants completed a Cambridge test, and analyses were based on mainly high-proficient (intermediate and higher levels of proficiency) participants. Thus, the absence of the expected interaction between misinformation endorsement and the language can be explained in that as people get more proficient in a second language, their interaction with information does not differ in both languages.

This conclusion was further supported by the exploratory analysis of misinformation effect and source monitoring between the participants with the Intermediate level of English language proficiency and participants with the higher levels of proficiency. Comparison of the degree of misinformation acceptance in English showed significant differences suggesting that the misinformation effect was greater for the High-proficient group than for the Intermediate group. A recent review of false memories in the bilingual Deese-Roediger-McDermott (DRM)-paradigm by Suarez and Beato (2021), also found that false memories are greater in participants with higher proficiency compared to the ones with lower levels and do not differ when the command of the participants in both the languages is similar. Importantly, the authors propose that it is not proficiency *per se* but rather dominance and environmental and interactional context that plays a major role in such differences (Beatty-Martínez et al., 2020; Suarez and Beato, 2021).

Indeed, both the results by Calvillo and Mills (2020), as well as the results of the current experiment could also be explained by the expectations, the participants have in a particular experimental setting and linguistic environment (Marian and Fausey, 2006), as well as their general higher confidence in their more proficient language. Representative to this notion is a study by Marian and Neisser (2000) examining a cued recall for autobiographical memories in Russian-English bilinguals. The results indicated that regardless of the language in which the cue was presented (Russian or English), the recall of the participants

was mainly influenced by the linguistic environment of the experiment, i.e., when interactions between the experimenter and participant were overtly in the Russian language, recalled were the memories that had been mainly encoded in the Russian language context. Indeed, almost all of the materials (with the exception of the English narrative) and instructions and recognition tasks were presented in Russian, and therefore, the retrieval of information was conducted in Russian. The effect of the linguistic environment and the language of information retrieval could be further tested in the future using a mirroring experiment, in which English is the main language of the experiment, with some of the materials being in Russian, as well as the cross-linguistic paradigm introduced by Shaw et al. (1997).

In relation to source monitoring performance, our expectations of the differences in misinformation effect among the languages were not confirmed. However, the ANOVA revealed that participants favored Russian sources when making incorrect source attribution for the English source (i.e., when the original source of information was the English narrative). Analysis of confidence ratings complemented these results, showing higher confidence ratings in such misattributions. This implies that participants, indeed, treated this information as coming from a Russian source. There was no similar pattern for other observed sources (Russian and None). These results may be an indication that the participants might have invested more resources in processing the information in the English language, which was one of the theories underlying the hypotheses in the current research. As mentioned in the previous sections, the source monitoring framework argues that memory judgments are thought to be based on phenomenological cues and meaningful details that are assigned to particular sources (Raye and Johnson, 1980). In the case that information processing in English was more effortful, more cognitive information would be assigned to this source, making it more distinguishable in that dimension. Therefore, participants were able to recognize the information as coming from the narrative, but for some reason attributed it to the wrong source.

Alternatively, it could also be argued that observed differences in source misattributions were caused by the structural differences between the languages themselves. Although Russian and English do not have such specific grammar features like Japanese or Turkish, they still demonstrate many differences (absence or presence of case system, personal ending in verbs, declination of adjectives by gender, etc.) that can influence the source monitoring processes. So, in the future, the effect of these linguistic features on source monitoring could be investigated with more precision.

Finally, there might be one more possible explanation for this effect. As discussed, source confusion is considered to underlie the misinformation effect, resulting in individuals confusing the source of original information and the source of postevent information. On the one hand, the presence of the misinformation effect in the current experiment implies that participants confused the sources of original (video) and postevent (narratives) information. On the other hand, the results of source monitoring showing a preference for Russian narrative in the English source suggest that participants

recognized that information came from the narrative but could not indicate the exact narrative. It was argued that it is more difficult to determine the origin of memories derived from external sources than, for example, to discriminate between the external and internal sources (Raye and Johnson, 1980). Memories from external sources are to be distinguished based on specific sensory content. Thus, it is more difficult to distinguish between the external sources when they are of the same or similar modality as sensory information related to these sources is similar.

In the case of our experiment, all the sources were external, but they differed in modality and could have provided different cues to justify source judgments. It might have been easier to distinguish between the information that came from the video and narratives (different modalities) than to discriminate between the information coming from two narratives (same modality). Therefore, participants were better at recognizing the modality of information, but not the source (English or Russian). In the future, the influence of modality can be further tested in other source monitoring paradigm, such as reality monitoring.

To our knowledge, this study is the first bilingual misinformation paradigm that did not manipulate the language of encoding and retrieval, but rather presented misleading information in two languages and measured its acceptance directly using explicit source monitoring tasks. The degree of misinformation acceptance did not seem to be affected by the language of misleading information; however, there might be still an influence on general information processing as shown by the results from the source monitoring task. Specifically, findings of the source monitoring task raise the basis for further examination of how exactly the monitoring processes work in bilinguals which can have important implications, both theoretically and practically.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Anwyl-Irvine, A. L., Massonni  , J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behav. Res. Methods* 52, 388–407. doi: 10.3758/s13428-019-01237-x
- Aydin, C., and Ceci, S. (2013). The role of culture and language in avoiding misinformation: pilot findings. *Behav. Sci. Law* 31, 559–573. doi: 10.1002/bsl.2077
- Beatty-Martinez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., and Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *J. Exp. Psychol. Learn. Memory Cogn.* 46, 1022–1047. doi: 10.1037/xlm0000770
- Bialek, M., Paruzel-Czachura, M., and Gawronski, B. (2019). Foreign language effect on moral dilemma judgments: an analysis using the CNI model. *J. Exp. Soc. Psychol.* 85, 1–8. doi: 10.1016/j.jesp.2019.103855
- Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., et al. (2005). Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. *NeuroImage* 24, 40–49. doi: 10.1016/j.neuroimage.2004.09.044

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the HSE University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The conceptualization of the research study was conducted by AD and BM-L. Data collection was conducted by AD, and analyses by AD and BM-L. The initial writing was completed by AD. Both the authors revised and contributed to the final version.

FUNDING

This article is an output of a research project implemented as a part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

ACKNOWLEDGMENTS

This research is a part of the master thesis entitled, Second Language Effect on Memory Errors, conducted by AD under the supervision of BM-L at the Educational Programme: Cognitive Sciences and Technologies: From Neuron to Cognition (<https://www.hse.ru/en/edu/vkr/470536235>). Portions of this research were presented at the 34th CUNY Conference on Human Sentence Processing, CUNY- virtual, 34th APPE annual Portuguese conference on Experimental Psychology, and 5th Human Brain Project—Student Conference, HBP-virtual. Thanks to Dr. Karlos Luna for helpful comments on an earlier version of this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.716336/full#supplementary-material>

- Boroditsky, L., Fausey, C., and Long, B. (2009). “The role of language in eye-witness memory: remembering who did it in English and Japanese,” in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 31, 2426–2431. Available online at: <https://escholarship.org/uc/item/4k31500t>
- Boroditsky, L., Ham, W., and Ramscar, M. (2019). “What is universal in event perception? Comparing English and Indonesian speakers,” in *Proceedings of the twenty-fourth annual conference of the Cognitive Science Society* (London: Routledge), 136–141. doi: 10.4324/9781315782379-61
- Boroditsky, L., Schmidt, L. A., and Phillips, W. (2003). “Sex, syntax, and semantics,” in *Language in Mind: Advances in the Study of Language and Thought*, eds. D. Gentner and S. Goldin-Meadow (New York, NY: MIT Press), 61–79.
- Caldwell-Harris, C.L. (2014). Emotionality differences between a native and foreign language: theoretical implications. *Front. Psychol.* 5:1055. doi: 10.3389/fpsyg.2014.01055
- Calvillo, D. P., and Mills, N. V. (2020). Bilingual witnesses are more susceptible to misinformation in their less proficient language. *Curr. Psychol.* 39, 673–680. doi: 10.1007/s12144-018-9787-9

- Ceci, S. J., Ross, D. F., and Toglia, M. P. (1987). Suggestibility of children's memory: psycholegal implications. *J. Exp. Psychol.* 116, 38–49. doi: 10.1037/0096-3445.116.1.38
- Christianson, S. -Å., and Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cogn. & Emo.*, 5, 81–108. doi: 10.1080/02699939108411027
- Costa, A., Foucart, A., Arnon, I., Aparici, M., and Apesteguia, J. (2014). “Piensa” twice: on the foreign language effect in decision making. *Cognition* 130, 236–254. doi: 10.1016/j.cognition.2013.11.010
- Diamond, A. (2002). “Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy, and biochemistry,” in *Principals of Frontal Lobe Function*, eds. D. T. Stuss and R.T. Knight (London, UK: Oxford University Press), 466–503. doi: 10.1093/acprof:oso/9780195134971.003.0029
- Echterhoff, G., Hirst, W., and Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cogn.* 33, 770–782. doi: 10.3758/BF03193073
- Fausey, C. M., and Boroditsky, L. (2011). Who dunnit? Cross linguistic differences in eye-witness memory. *Psychol. Bull. Rev.* 18, 150–157. doi: 10.3758/s13423-010-0021-5
- Foley, M. A., Johnson, M. K., and Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Dev.* 54, 51–60. doi: 10.2307/1129860
- Hayakawa, S., Costa, A., Foucart, A., and Keysar, B. (2016). Using foreign language changes our choices. *Trends Cogn. Sci.* 20, 791–793. doi: 10.1016/j.tics.2016.08.004
- Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., and Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign language effect on moral judgment. *Psychol. Sci.* 28, 1387–1397. doi: 10.1177/0956797617720944
- Johnson, M. K. (1997). Source monitoring and memory distortion. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 352, 1733–1745. doi: 10.1098/rstb.1997.0156
- Johnson, M. K., Hashtroudi, S., and Lindsay, D. S. (1993). Source monitoring. *Psychol. Bull.* 114, 3–28. doi: 10.1037/0033-2909.114.1.3
- Lindsay, D. S., Johnson, M. K., and Kwon, P. (1991). Developmental changes in memory source monitoring. *J. Exp. Child Psychol.* 52, 297–318. doi: 10.1016/0022-0965(91)90065-Z
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cogn. Psychol.* 7, 560–572. doi: 10.1016/0010-0285(75)90023-7
- Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & Memory*, 12, 361–366. doi: 10.1101/lm.94705
- Loftus, E. F., and Guido, Z. (1975). Eyewitness testimony: the influence of the wording of a question. *Bull. Psychon. Soc.* 5, 86–88. doi: 10.3758/BF03336715
- Loftus, E. F., and Hoffman, H.G. (1989). Misinformation and memory: the creation of new memories. *J. Exp. Psychol.* 118, 100–104. doi: 10.1037/0096-3445.118.1.100
- Loftus, E. F., and Palmer, J. C. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory. *J. Verbal Learn. Verbal Behav.* 13, 585–589. doi: 10.1016/S0022-5371(74)80011-3
- Luna, K., and Martín-Luengo, B. (2013). Monitoring the source monitoring. *Cogn. Process.* 14, 347–356. doi: 10.1007/s10339-013-0558-0
- Luna, K., and Migueles, M. (2009). Acceptance and confidence of central and peripheral misinformation. *Spanish J. Psychol.* 12, 405–413. doi: 10.1017/S1138741600001797
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marian, V., and Fausey, C. M. (2006). Language-dependent memory in bilingual learning. *Appl. Cogn. Psychol.* 20, 1025–1047. doi: 10.1002/acp.1242
- Marian, V., and Neisser, U. (2000). Language-dependent recall of autobiographical memories. *J. Exp. Psychol.* 129, 361–368. doi: 10.1037/0096-3445.129.3.361
- Porter, S., Spencer, L., and Birt, A. R. (2003). Blinded by emotion? Effect on the emotionality of a scene on susceptibility of false memories. *Canadian J. Behav. Sci.*, 35, 165–175. doi: 10.1037/h0087198
- Raye, C., and Johnson, M. K. (1980). Reality monitoring vs. discriminating between external sources of memories. *Bull. Psychon. Soc.* 15, 405–408. doi: 10.3758/BF03334572
- Ruffman, R., Rustin, C., Gamham, W., and Parkin, A. J. (2001). Source monitoring and false memories in children: Relation to certainty and executive functioning. *J. Exp. Child Psychol.* 80, 95–111. doi: 10.1006/jecp.2001.2632
- Schroeder, S. R., and Marian, V. (2014). “Bilingual episodic memory: how speaking two languages influences remembering,” in *Foundations of Bilingual Memory* (New York, NY: Springer), 111–132. doi: 10.1007/978-1-4614-9218-4_6
- Shaw, J. S., Garcia, L. A., and Robles, B. E. (1997). Cross-language postevent misinformation effects in Spanish-English bilingual witnesses. *J. Appl. Psychol.* 82, 889–899. doi: 10.1037/0021-9010.82.6.889
- Sinopoli, K., and Dennis, M. (2012). Inhibitory control after traumatic brain injury in children. *Int. J. Dev. Neurosci.* 30, 207–215. doi: 10.1016/j.ijdevneu.2011.08.006
- Slobin, D. (1987). “Thinking for speaking,” in *Proceedings of Annual Meeting of the Berkeley Linguistics Society*, vol 13, 435–445. doi: 10.3765/bls.v13i0.1826
- Smith, K. C., Multhaup, K. S., and Ihejirika, R. C. (2017). From eyewitnesses to academic contexts: examining the effect of misinformation in first and second languages. *Appl. Cogn. Psychol.* 31, 546–557. doi: 10.1002/acp.3352
- Suarez, M., and Beato, M. S. (2021). The role of language proficiency in false memory: a mini review. *Front. Psychol.* 12:659434. doi: 10.3389/fpsyg.2021.659434
- Tosun, S., Vaid, J., and Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *J. Memory Lang.* 69, 121–134. doi: 10.1016/j.jml.2013.03.004
- Tuckey, M., and Brewer, N. (2003). How schemas affect eyewitness memory over repeated retrieval attempts. *Appl. Cogn. Psychol.* 17, 785–800. doi: 10.1002/acp.906
- Ünal, E., Pinto, A., Bunger, A., and Papafragou, A. (2016). Monitoring sources of event memories: a cross-linguistic investigation. *J. Memory Lang.* 87, 157–176. doi: 10.1016/j.jml.2015.10.009
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Santons, M., et al. (1999). A system for relation reasoning in the human prefrontal cortex. *Psychol. Sci.* 10, 119–125. doi: 10.1111/1467-9280.00118
- Weissberger, G., Gollan, T., Bondi, M., Clark, L., and Wierenga, C. (2015). Language and task switching in the bilingual brain: bilinguals are staying, not switching, experts. *Neuropsychologia* 66, 193–203. doi: 10.1016/j.neuropsychologia.2014.10.037

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Dolgoarshinnaia and Martin-Luengo. 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) and the copyright owner(s) 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.



Identification Performance During Quarantine by COVID-19 Pandemic: Influence of Emotional Variables and Sleep Quality

Facundo A. Urreta Benítez^{1,2,3*}, Candela S. Leon^{1,3}, Matías Bonilla^{1,2,3}, Pablo Ezequiel Flores-Kanter^{1,2} and Cecilia Forcato^{1,2*}

¹ Laboratorio de Sueño y Memoria, Departamento de Ciencias de la Vida, Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina, ² Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina, ³ Innocence Project Argentina, Buenos Aires, Argentina

OPEN ACCESS

Edited by:

Beatriz Martín-Luengo,
National Research University Higher
School of Economics, Russia

Reviewed by:

Catriona Havard,
The Open University, United Kingdom
Dawn Rachelle Weatherford,
Texas A&M University-San Antonio,
United States
Sergii Yaremenko,
Maastricht University, Netherlands

*Correspondence:

Facundo A. Urreta Benítez
furreta@itba.edu.ar
Cecilia Forcato
cforcato@itba.edu.ar

Specialty section:

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

Received: 06 April 2021

Accepted: 24 September 2021

Published: 15 October 2021

Citation:

Urreta Benítez FA, Leon CS, Bonilla M, Flores-Kanter PE and Forcato C (2021) Identification Performance During Quarantine by COVID-19 Pandemic: Influence of Emotional Variables and Sleep Quality. *Front. Psychol.* 12:691583. doi: 10.3389/fpsyg.2021.691583

The COVID-19 pandemic has caused major disruptions in people's lives around the globe. Sleep habits and emotional balance have been disturbed in a way that could be comparable to the havoc caused by a deep personal crisis or a traumatic experience. This unfortunate situation provides a unique context in which to study the impact of these imbalances on cognitive processes. In particular, the field of eyewitness science could benefit from these conditions, since they are also often present in crime victims, but can only be generated in the laboratory up to a certain ethical and practical limit. For several decades, eyewitness studies have tried to discover what variables affect people's ability to properly recognize faces. However, the disparity of experimental designs and the limitations of laboratory work could be contributing to the lack of consensus around several factors, such as sleep, anxiety, and depression. Therefore, the possibility of observing the influence of these agents in natural contexts could shed light on this discussion. Here, we perform simple and repeated lineups with witnesses of mock-crime, considering the conditions related to the COVID-19 pandemic, which to some extent allow emulating the deterioration in general well-being that often afflicts crime victims. For this, 72 participants completed symptomatology scales, and watched a video portraying a staged violent episode. Subsequently, they gave testimony and participated in two lineups, in which we manipulated the presence/absence of the perpetrator, to recreate critical scenarios for the appearance of false recognitions. We found an increase in recognition errors in those individuals who did not have access to the perpetrator during the Initial lineup. Additionally, the conditions of the pandemic appear to have adversely affected the ability to witness and accurately perform lineups. These results reaffirm the need to move toward the standardization of research practices and methods for assessing testimonial evidence, especially in relation to the results of the lineups. Considering the degree of fallibility of these processes can lead to a reduction of wrongful convictions.

Keywords: recognition, lineup, episodic memory, anxiety, depression, sleep, COVID-19

INTRODUCTION

Since the beginning of the COVID-19 pandemic, mental health surveys have been conducted consistently around the world. The concern arises from the fact that social isolation, confinement, and sedentary lifestyle are directly associated with a broad range of mental disorders (Shah et al., 2021). The results of these studies indicate a significant prevalence of negative feelings, derived from the fear of being infected, economic instability, frustration, and boredom (among others) (Brooks et al., 2020). As a result of this unfavorable situation, intense symptoms of anxiety (a psychophysiologic sign of stress, Robinson, 1990) and depression have been observed in large segments of the population (Ozamiz-Etxebarria et al., 2020; Shah et al., 2021; Sheridan et al., 2021), as well as significant disturbances in sleep habits (Barros et al., 2020; Jahrami et al., 2021). This unfortunate situation provides a unique opportunity to study how these disturbances affect various human activities, especially in those areas where the lack of consensus could be a direct consequence of ethical and methodological limitations. Such is the case of eyewitness studies.

A difficulty that constantly appears throughout most of the background in eyewitness science is the limitation (fundamentally ethical) to produce ecological conditions in the laboratory (Valentine and Mesout, 2009), that is, levels of anxiety, depression, or sleep loss like those that could be observed in crime victims. The need to find a way to compensate for these methodological flaws becomes evident when contemplating the consequences of the scarcity of applicable knowledge.

According to information from the Innocence Project organization¹, 72% of wrongful convictions in the United States are strongly determined by incorrect identifications in lineups. Even acting in good faith, people may misrecognize an individual in a lineup, due to various environmental and personal factors. In order to understand this, we must understand memory as highly malleable and subject to distortion and not as a video camera that faithfully reproduces the past (Clifasefi et al., 2007). These distortions can lead to the formation of false memories, that is, memories of events that never occurred, or memories with added or altered details (Loftus, 2003; Newman and Lindsay, 2009).

The formation of false memories constantly occurs in everyday life. Without realizing it, people incorporate details and characters that come from dreams, external suggestions, or confusion to their anecdotes (Reyna et al., 2016). In terms of the ability of making decisions in lineups and describing events, memory errors can have drastic consequences (mainly in the legal field). Therefore, it is imperative to discover what processes favor their appearance and how they can be reduced.

The emotional conditions of individuals are some of the most studied factors for their influence on the ability to recognize faces (Edelstein et al., 2004). In particular, the impact of a witness's level of stress has been extensively addressed, considering the high stress that criminal acts, and subsequent police processes can generate (Dobson and Markham, 1992; Deffenbacher et al., 2004).

The models that relate stress and general cognitive performance have become more complex over the years, to capture more and more nuances, and account for the vastly disparate results in experimentation. This complexity is transferred directly to eyewitness studies, in general, the results found in the existing literature are mixed (Valentine and Mesout, 2009). Evidence of a facilitating effect has been found, in which subjects under higher stress have more accurate memories of a witnessed event (Lindberg et al., 2001). Yuille et al. (1994) presented a group of police recruits with a simulation task in which a situation of greater or lesser stress was set. In the following weeks, the participants were interviewed, and it was observed that those who had participated in the high-stress set-up, reported more precise details. Opposite findings have also been made (Pezdek et al., 2020), in which witness performance worsens as stress arises. Morgan et al. (2004) studied the performance in lineups of a group of soldiers enrolled in a military survival academy. After a 12-h confinement in a simulated prison camp, the participants experienced a low- or high-stress interrogation (the interrogators were more or less aggressive). A day later, the soldiers went through a lineup trying to identify their interrogators. A dramatic decrease in accuracy was observed in the high-stress group, compared to the low-stress group. According to Tyng et al. (2017) stress can affect performance on memory tasks, both positively and negatively. This depends mainly on the intensity of the stimulus, type of memory involved, and the phase of the memory process where the excitation is applied (Deffenbacher et al., 2004). A very popular current theoretical model proposes that the release of hormones during stress (particularly catecholamines and glucocorticoids) turns the stressed organism in a “memory formation mode” that prioritizes the encoding and storage, to the detriment of retrieval. This mechanism has an adaptive value, since it prioritizes the acquisition of information in a potentially dangerous environment and can explain the performance of our participants, who learned better, but retrieved worse at higher levels of arousal (Schwabe et al., 2012).

Depressive mood is another emotional disturbance that usually appears in witnesses and crime victims (Norris and Kaniasty, 1994). This affliction can persist for long periods (it can encompass the entire judicial process) and can result into a major depressive disorder, interpersonal problems, and even lead to suicide (Rounding et al., 2014). There is evidence suggesting that chronically negative mood states increase the possibility of selecting the target in a lineup. Rounding et al. (2014) showed a group of healthy subjects a series of images of faces, while assessing the possible presence and intensity of their depressive symptoms. A week later, the same individuals had to make an identification on a series of six-person lineups, attempting to recognize the faces previously observed. The results showed that those with mild sustained dysphoria generally had greater accuracy than those without symptoms and those severely depressed. This effect is usually stable and is not affected by acute mood changes, whether it is positive feelings or new depressants (unless the latter are very intense). In general, it is considered that individuals with high depressive symptoms elaborate the information stored in memory in an active and biased way, that

¹<https://innocenceproject.org/>

is, they tend to select the details that make up negative events with violent or unpleasant elements, which in a certain way reaffirm their biased perception of reality (Watkins et al., 1996). As a result, one of the most robust findings in the literature on depression is that depressed subjects had a stronger and more persistent memory of negative events, while they more easily forget the positive or pleasant (Gotlib and Joormann, 2010).

Outside emotional factors, there are variables related to the individual that must be considered when studying their performance in lineups. Among them, sleep is one of fundamental importance, given its role in the acquisition, consolidation, and integration of new information (Rasch and Born, 2013). In general aspects, it is quite clear that sleep is beneficial for memory (Walker, 2009). However, when observing this relationship in greater depth, it becomes evident not only that this facilitating effect differentially affects the different memory phases, but that it is also capable of producing undesired results. Sleep favors subsequent memory acquisition while sleep disturbances can lead to an encoding decline (Van Der Werf et al., 2009) and this effect is usually explained by the synaptic homeostasis hypothesis (Tononi and Cirelli, 2003). According to this hypothesis, sleep favors the decay of weak synaptic connections formed during wakefulness, performing a downscaling of the synapses, which benefits strong connections (increasing the signal-noise ratio), and highlighting the information that is most valuable. Because of this downscaling, adequate sleep translates into increased encoding ability during later wakefulness. However, the beneficial effect of sleep does not stop there. Sleep also improves memory consolidation (Born and Wilhelm, 2012) through active consolidation processes. This theory proposes that during sleep, specifically during Slow Wave Sleep (SWS) recently acquired memories are spontaneously reactivated, promoting the gradual redistribution of hippocampus dependent memories from the hippocampus to neocortical areas where they will be stored in long term networks. Furthermore, Rapid Eye Movement (REM) sleep favors integration of memories (Payne, 2014). An integral part of this process is the joint reactivation of new and old memories, to find overlaps and extract central ideas that link the information units to each other (Diekelmann and Born, 2010). These central ideas tend to be more durable in time than specific details (these tend to disappear more easily) and have been identified as possible causes of false memories (Payne et al., 2009). Thus, when we remember a certain event, it is much easier to access its meaning than its details. For this reason, when producing a detailed account, we may find blank spaces, that is, elements that are missing. Furthermore, the stronger a central idea (gist) is, and the weaker the recall of details, the more we will tend to produce false memories (Brainerd and Reyna, 2002).

These claims about the impact of sleep on general memory processes have considerable consensus and are well documented. However, when attempting to translate these results into the specific field of eyewitness science, drawbacks arise. Studies are scarce, and the results are mixed. For example, Stepan et al. (2017) showed the participants a mock-criminal video and performed a photographic lineup 12 h after in the presence or absence of the

target. The participants that slept between the training and the testing sessions had a better performance rejecting the innocent when the perpetrator was absent in the lineup. Nevertheless, in a similar procedure, Morgan et al. (2019) did not observe differences between groups of participants that either slept after the training or remained awake.

When considering the ability to recall details, it has been observed that the quality of sleep prior to the mock-crime influences performance, in fact, as the quality of sleep decreases, the ability to recall details also decreases. Thorley (2013) showed a video of a simulated crime to a group of subjects who reported their level of sleepiness. They also completed sleep quality scales, referring to the night prior to the experiment. In a later recall of what was watched in the video, it was observed that as sleep quality decreases and sleepiness increases, individuals tend to report less detail. This result goes in line with several studies showing that sleep deprivation impairs memory acquisition during subsequent wakefulness (Kaida et al., 2015). Furthermore, the relationship between confidence in choice and accuracy decreases under conditions of sleep deprivation (Blagrove and Akehurst, 2000).

As has been observed there are few clear trends and consensus still needs to be reached on multiple issues. For this reason, locating the experiment in a natural environment, which spontaneously presents one or more of these characteristics (Morgan et al., 2013) could be beneficial, and bring some clarity about misidentifications.

A particular problem very present in police practices in some regions of the world is that of repeated lineups. Particularly in the case of Latin America, these procedures are applied despite being strongly contraindicated. Repeated lineups have been extensively studied and are generally considered unreliable. During the second lineup, people are often retroactively influenced by the first (Stebly and Dysart, 2016). This may be due to the occurrence of a "compromise effect," that is, a person tends to repeat their choices, to show consistency to themselves and others (Valentine et al., 2012; Lin et al., 2019). It can also be due to the "transference effect" (Loftus, 1976), the inability of an eyewitness to distinguish between a familiar but innocent person, from the actual criminal that was observed at the scene of crime (Ross et al., 1994). Finally, it can also be caused by the process of memory reconsolidation (especially in the absence of the perpetrator during the first lineup). That is, during the initial lineup some faces of the foils could have similar features as the target, triggering a prediction error, i.e., the mismatch between what is predicted according to previous experiences and what is encountered during re-exposition allowing memory labilization (Forcato et al., 2020). In this case, the memory would be updated during reconsolidation, incorporating erroneous information from the faces present in the lineup into the original memory.

However, studies on multiple identifications often do not consider the impact of sleep and emotional variables on the process (since observation tends to focus on repetition itself). Given the influence that these variables seem to have on simple identifications, and the fact that repeated lineups are still practiced in some countries, it is of interest to contemplate this case from an exploratory perspective.

In this exploratory study, we will analyze the impact on performance in simple and repeated lineups, of emotional states and sleep habits during the lockdown related to the COVID-19 pandemic.

Finally, the influence of the same variables in tasks of free recall and chronological order of images will be studied, seeking to determine if the relationships are repeated through different memory modalities.

To this end, two groups of subjects completed psychometric scales, watched a video of an incident at a conference, and gave their oral testimony (day 1). 24 h later, they tried to identify the perpetrator in a photographic lineup, in the presence (With perpetrator group) or absence (Without perpetrator group) of the perpetrator (day 2). On day 8 both groups carried out a definitive lineup in present condition. Additionally, they gave a final testimony, and were tested for the episodic temporal order of the event.

MATERIALS AND METHODS

The participants were 78 Argentines recruited online through the official social networks of the Sleep and Memory Lab. Applicants underwent a prior online interview with the experimenter, to ensure that they met the inclusion criteria. They also had to demonstrate that they had the appropriate technical resources (PC and fast enough Internet etc.), and basic knowledge of how to use them. The sample size was decided according to previous studies sharing similar designs (Wells, 1984; Steblay et al., 2013). These studies with comparable sample size have demonstrated significant effects of behavioral intervention, suggesting the reproducibility of these effects on memory with similar sample size.

For the collection of sociodemographic data and symptomatological scales, the Google Forms platform was used. This form of data collection has been shown to be equivalent to traditional forms of collection (Weigold et al., 2013). Then, the experiment was carried out through the Google Meet video calling platform with the experimenter guiding the entire process. It was controlled that all people have access to a computer screen (not cell phone) and that they have quality internet connection. The experiments were approved by the Alberto C. Taquini Biomedical Research Ethics Committee. 6 participants were excluded from the data analyzes because they only completed the first session of the experiment (day 1) and did not show up to the following meetings. The final sample consisted of 72 subjects (Age $M = 29 \pm 6$, years, **Table 1**). Inclusion criteria: The participants stated that they were not ill during the experiment, did not suffer from mental disorders, took psychiatric medication, or had sleep disorders. The experiments were carried out between the months of April and September 2020, within the period of preventive and mandatory social isolation in Argentina.

Procedure

The entire study was conducted online, at the beginning of each session, the participants entered a video call with the

TABLE 1 | Sociodemographic data.

		With perpetrator group	Without perpetrator group
	N	39	33
	Age	29.33 $\pm$ 5.61	28.45 $\pm$ 7.16
Gender	Females	79.48%	78.78%
	Males	17.94%	21.21%
	Non-binary	2.56%	
Education	High school graduates	10.25%	9.09%
	College students	17.94%	33.33%
	College graduates	71.79%	51.51%

Number of participants in each group, mean age $\pm$ SD, percentage of different genders, and education level.

experimenter, who provided instructions, showed the stimuli through streaming, and supervised the tasks.

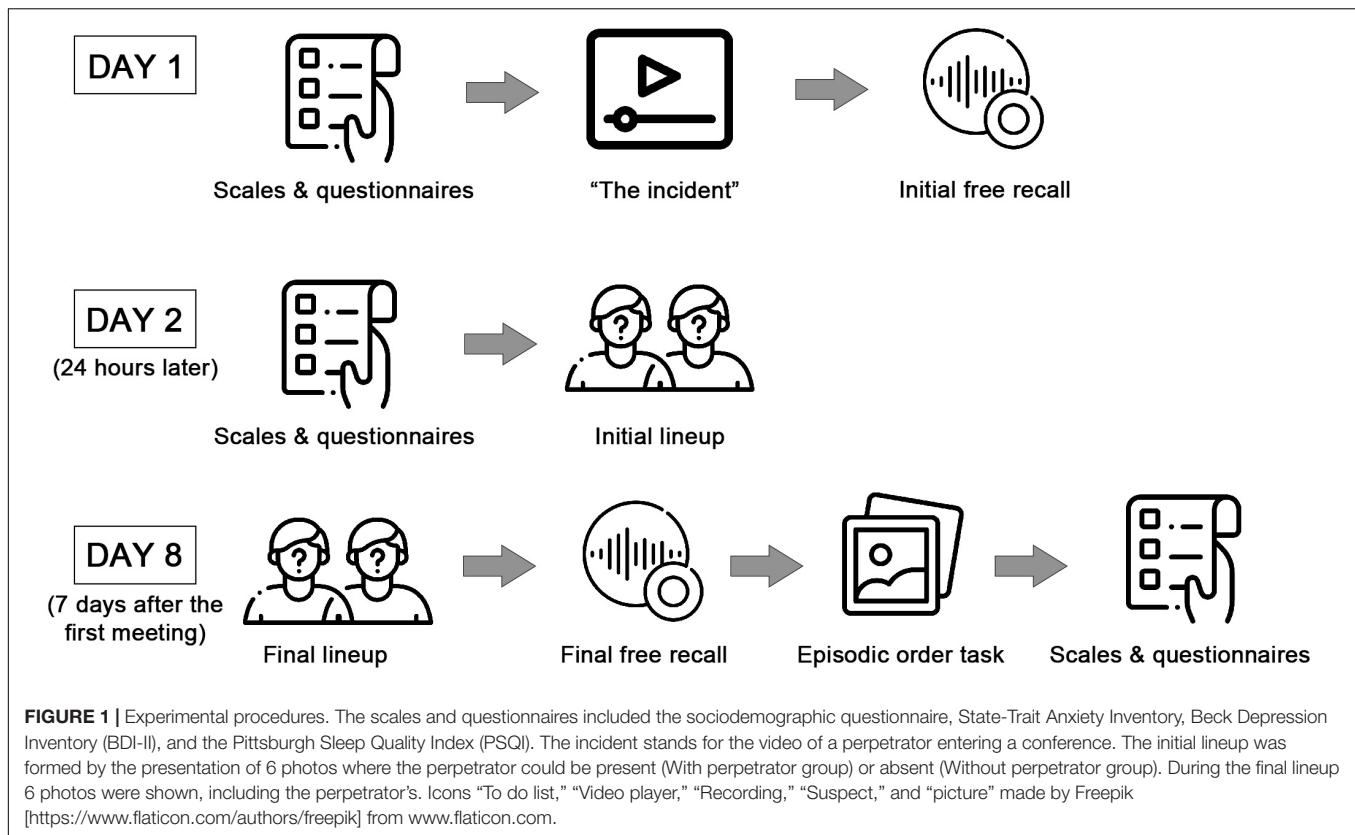
On day 1, after signing the informed consent, they completed the first part of the sociodemographic questionnaire, State-Trait Anxiety Inventory (STAI-Y), and Beck Depression Inventory (BDI-II). Immediately after, they watched a video called “The Incident,” featuring an individual acting aggressively in front of a crowd, and after that, the Initial free recall was performed. 24 h later (day 2), the participants completed the Stanford Sleepiness Scale (SSS), the State Anxiety Inventory, and carried out the Initial lineup. On day 8, the participants performed the Final lineup. Immediately after, a Final free recall of the video watched on day 1 was executed. Finally, they completed the Chronological order task, the Pittsburgh Sleep Quality Index (PSQI), the STAI-Y, and the second part of the sociodemographic questionnaire (**Figure 1**).

The Incident (Video)

The video showed a conference in a room with numerous people. After about 30 s, a young man broke into the meeting to deliver a message. At that moment he argued with the main speaker of the talk, and became violent, yelling, and throwing objects to the ground. After this, he withdraws muttering. The speaker tried to retake the talk, and the video ended. It was filmed in high definition, by three cameras that alternated presenting a general panorama of the front of the room. It took 90 s.

Initial Lineup

Subjects were presented with an image, simultaneously showing the photographs of 6 bearded males of similar ages and builds, randomly ordered, in black and white, numbered from left to right from one to six, in the form of a typical six-person lineup (lineup consisted of 5 foils and the perpetrator). For the Without perpetrator group, the perpetrator was extracted and another man with similar characteristics was added (lineup consisted of 6 foils). The participants were asked to observe the image for as long as necessary, noting the number that accompanies each photo. It was instructed that they had to identify the person who had broken into the talk or reject the lineup. The subjects received an unbiased instruction, which indicates the possibility of the absence of the perpetrator (“Now you are going to see a lineup



with six photos, among which the person who broke into the video you saw yesterday may or may not be found. Take your time to see them. If you identify the suspect, I will ask you to tell me the number that accompanies his photo. If you consider that he is not present, tell me”). In response, they provided a number, or rejected the lineup. Immediately afterward, they were asked for an estimate of the degree of confidence in their own decision, with a number between zero and one hundred, representing with zero the absolute lack of confidence.

A total of 40 participants were recruited to assess an online fairness test of the with perpetrator six-persons lineup (lineup consisted of 5 foils and the perpetrator), and 64 participants for the six-person lineup that had no perpetrator, via a mock witness paradigm (lineup consisted of 6 foils) (Malpass and Lindsay, 1999). A group of simulated witnesses, who have not witnessed the crime video and who did not know the identity of the perpetrator, received a brief description of the perpetrator, and were asked to select the suspect from the list based on this description. For the lineup to be considered fair, the mock witnesses should not be able to identify the suspect at a rate greater than chance (lineup bias), and the distribution of their choices should be spread equally over the lineup members (lineup size, Brigham et al., 1999). In order to measure the lineup size, the Acceptable Lineup Members technique (ALM) was used (Malpass and Devine, 1983). A total of 75% was the minimum percentage of the probability expectation considered acceptable (Brigham et al., 1990). The Functional Size was used to measure the lineup bias (Wells et al., 1979).

In the lineup that included the perpetrator, an ALM of 3.40 and a Functional Size of 5 were obtained. In the lineup without the perpetrator an ALM of 2.50 was obtained.

Initial and Final Free Recall

The participants were instructed to describe in as much detail as possible what they had watched in the video, mentioning that it might include dialogues, characteristics of the people (clothes and physical qualities, etc.) and the place, among other elements. The free recall was recorded, and the total number of details was counted and classified according to their veracity in correct and incorrect details. When counting the number of details, we considered actions, persons, objects, and elements of the environment. Every detail was counted only once, no matter how many times it was repeated in history. The instruction was “Now I’m going to ask you to describe, in as much detail as possible, what you have watched in the video. You can include dialogues, characteristics of the people (clothes and physical characteristics, etc.) and the place. I am going to record everything you say with the recorder.

Final Lineup

During this test, the lineups of both groups included the perpetrator. It was similar to the Initial lineup, but the set of photographs was personalized, so that each subject could repeat their previous choice. Within the group of 6 photos, there were 3 new faces, and 3 previously seen (among which was the suspect chosen in the Initial lineup, if one has been chosen). On this

occasion, both groups had access to the perpetrator and 5 foils, and the order of the photos was semi-randomized, so that none of the previously seen individuals occupied the same place. The fairness control of this lineup was carried out in the same way as in the Initial lineup and 40 mock witnesses were used. It obtained an ALM of 4.13 and a Functional Size of 5.

Socio-Demographic Questionnaire

It included contact information, sex, age, educational level, occupation, cohabitation group, intake of medication, and presence of sleep disorders.

State-Trait Anxiety Inventory

Based on a 4-point Likert scale, it contains 40 questions that are used to estimate two types of anxiety: state anxiety (the level of anxiety experienced at the time of performing the task) and trait anxiety (the personality-integrated anxiety of the individual) (Spielberger et al., 1970). The adaptation of this test for Argentina was used (Leibovich de Figueroa, 1991).

Beck Depression Inventory

It is a multiple-choice inventory used to measure severity of depressive symptoms during the past 2 weeks. It is made up of 21 items that cover emotional, cognitive, and behavioral aspects (BDI-II; Beck et al., 1996).

Chronological Order Task

The participants were provided with 5 images semi-randomly extracted from the video, and they were asked to order them chronologically, starting with the one that was observed first. The five images were presented simultaneously and the time to respond was unlimited. The task was scored considering the performance in terms of two factors: absolute location (that the first observed image was assigned to place 1) and relative location (that the first observed image was assigned to a position prior to the second observed image). The result obtained varies between 0 and 5, a higher score represents a better performance.

Pittsburgh Sleep Quality Index

It is a self-administered questionnaire that assesses the quality of sleep integrating several factors, such as its latency, duration, and efficiency. Each component receives a score between 0 and 3, and is subsequently added to the others, having a result of between 0 and 21 points. Higher scores represent poorer quality of sleep (Buysse et al., 1989).

Statistical Analysis

Statistical analyzes were carried out in the statistical software IBM SPSS Statistics 25, and RStudio Version 1.3.1073. The scores of the three symptomatology scales (STAI-Y, BDI-II, PSQI) were transformed into categories above and below average (low/high) in a similar way to the procedure of Valentine and Mesout (2009). The results of the Initial lineup for the With perpetrator group were considered: “target selection” if they selected the suspect, “foil selection” if they selected a wrong suspect and “incorrect rejection” if they rejected the lineup. In the case of the

Without perpetrator group, the results of the Initial lineup were considered as “correct rejection” if they rejected the lineup and “foil selection” if they selected a wrong suspect. The results of the Final lineup were considered in the same way for both groups: “target selection” if they selected the suspect, “incorrect rejection” if they rejected the lineup and “foil selection” if they selected a foil. Considering the low number of participants who did not select anyone, “foil selection” and “incorrect rejection” were analyzed together. The score obtained from the Chronological order task was used as a direct value. The correct, incorrect, and total details were used as direct values. Additionally, the difference in the number of details (total, correct and incorrect) between day 1 and day 8 was treated as a direct value and as a proportion (memory change). We referred to the set of variables related to recall as: recall variables.

The frequency of target selection for the With/Without perpetrator groups, in the Initial and Final lineup, was analyzed with Pearson's chi-squared test. We further calculated the Odds ratio. Additionally, chi-squared was used to compare the frequency of target selections of the subgroups (within With/Without perpetrator) of high or low anxiety, depression and quality of sleep, both in the Initial lineup and in the Final lineup, between them and against the chance level.

The recall variables were analyzed with two-tailed T-test comparing the high and low level of the symptomatology variables. The score of the chronological order task was related to symptomatology variables in the same way.

A paired *t*-test was used to compare recall variables between Initial and Final free recall. For the non-parametric variables, we used the Wilcoxon Signed-Ranks Test.

All tests were performed with a fixed alpha of 5%.

Additionally, to evaluate the relationship between the accuracy of the elections and the confidence attributed to them, two types of analysis were carried out. Initially, Point Biserial Correlation was applied (Krug, 2007) and then CAC curves were performed. To calculate the value of the correct proportion corresponding to each confidence level (low 0–50%, medium 60–80%, or high 90–100%) the following formula was used: # Correct identifications/# Correct identifications + # Incorrect identifications (Mickes, 2015).

RESULTS

Lineup Recognition

Regarding the repetition of the lineup, we observed that in the With perpetrator group 43% ($N = 17$) of the participants selected the target on Day 8, significantly higher than the 22% ($N = 7$) selected by the Without perpetrator group [Figure 2A, $\chi^2(1) = 4.03$, $p = 0.045$, $\phi_c = 0.24$]. Thus, the target selection was alarmingly low for both groups. However, if we consider that by chance 17% (one sixth) of the subjects would select the target, we observed that the With perpetrator group was significantly higher than the chance level [$\chi^2(2) = 6.02$, $p = 0.01$, $\phi_c = 0.28$]. This difference was not observed for the Without perpetrator group [$\chi^2(2) = 0.10$, $p = 0.76$, $\phi_c = 0.04$]. Furthermore, based on the odds ratio, the odds that a subject would recognize the target on

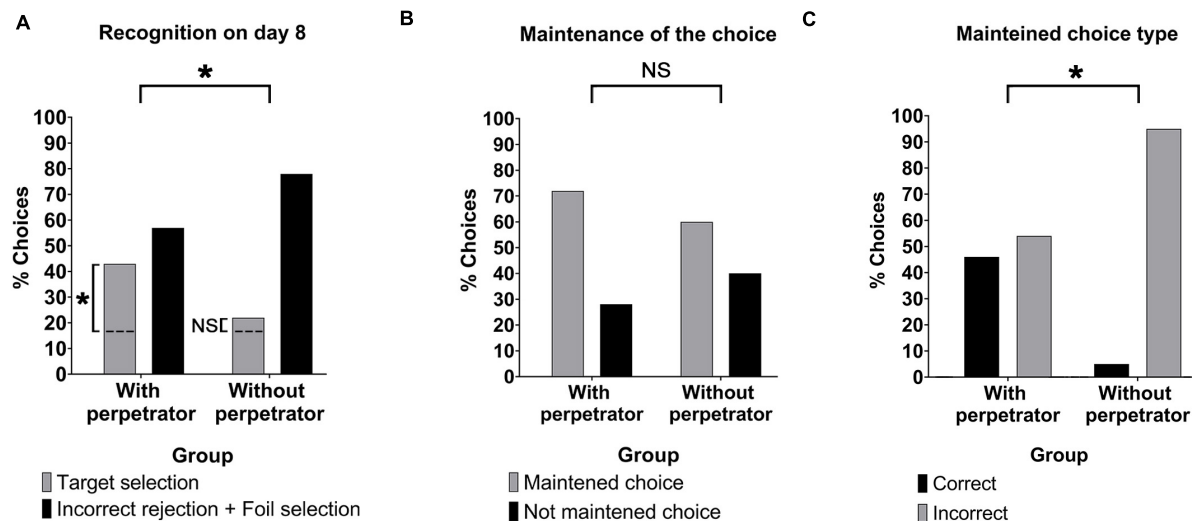


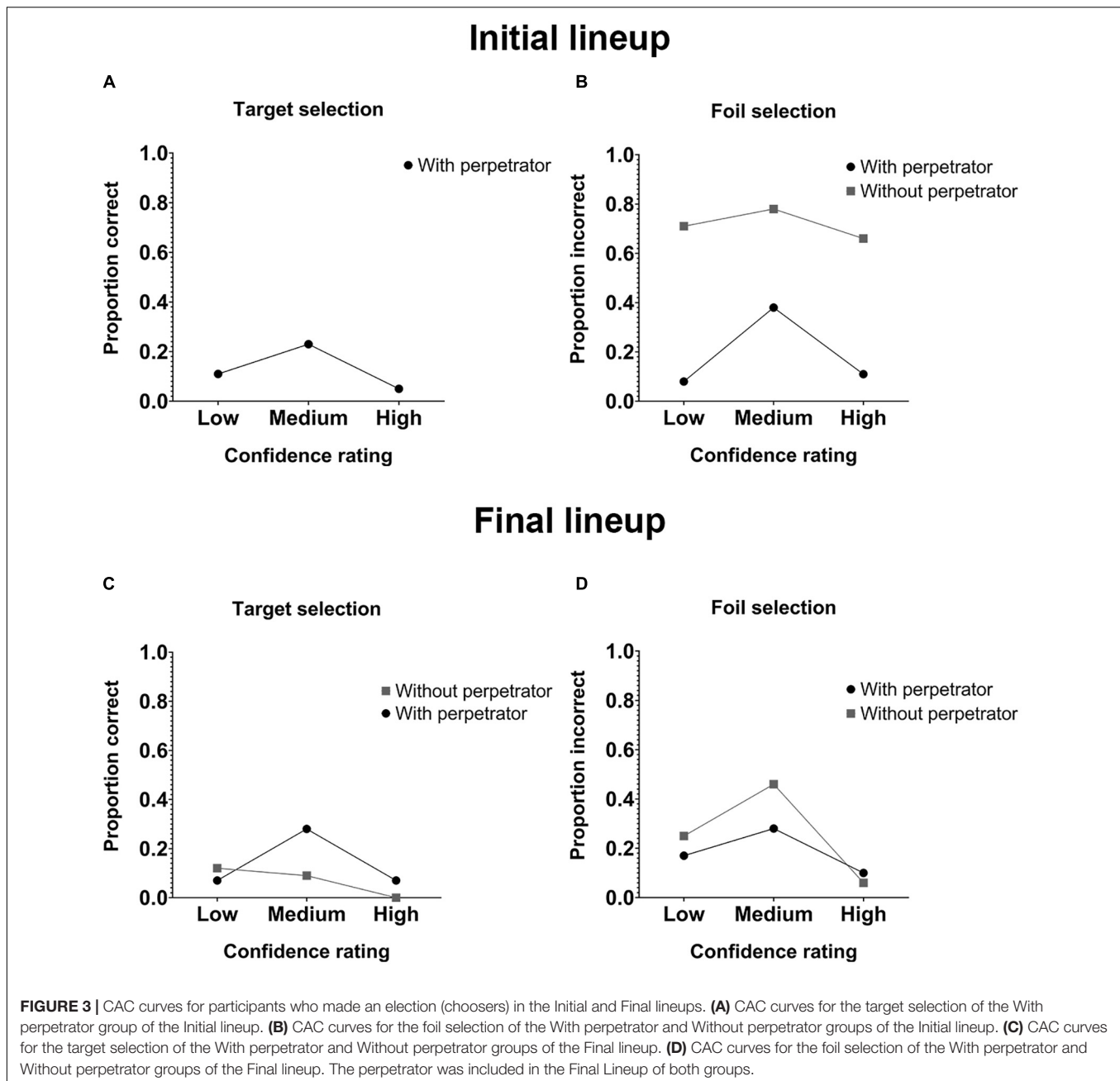
FIGURE 2 | Lineup recognition. **(A)** Percentage of target selection and incorrect rejection + foil selection for the With perpetrator and Without perpetrator groups on day 8. With perpetrator group: target selection ($N = 17$), incorrect rejection ($N = 0$), and foil selection ($N = 22$). Without perpetrator group: target selection ($N = 7$), incorrect rejection ($N = 1$), and foil selection ($N = 25$). **(B)** Percentage of maintained and not maintained choices between day 2 and 8 for the With perpetrator and Without perpetrator groups. With perpetrator group: maintained choices ($N = 28$), not maintained choices ($N = 11$). Without perpetrator group: maintained choices ($N = 20$) and not maintained choices ($N = 13$). **(C)** Percentage of correct and incorrect maintained choices. With perpetrator group: correct (target selection/target selection, $N = 12$); incorrect (incorrect rejection/foil selection, $N = 4$ and foil selection/foil selection, $N = 17$). Without perpetrator group: correct (correct rejection/target selection, $N = 1$); incorrect (foil selection/foil selection, $N = 24$). The following choices were not present in the sample: incorrect rejection/incorrect rejection (for the With perpetrator group) and correct rejection/target selection and foil selection/incorrect rejection (for Without perpetrator group). * $p < 0.05$; NS, $p > 0.05$. The dashed lines stand for the chance level (17% of correct responses).

Day 8 would be 2.96 times higher if the perpetrator had been available during recognition on Day 2.

There were no significant differences between groups for the maintained choices between the recognition at Day 2 and the test at Day 8 (choose the same photo in both lineups) (**Figure 2B**, With perpetrator group: 72% (28), Without perpetrator group: 60% (20) [$\chi^2(2) = 1.01$, $p = 0.32$]). However, within these repeated identifications, the rate of correct choices was different for both groups. While the With perpetrator group showed that 46% ($N = 13$) of their repeated choices were correct (target selection-target selection), the Without perpetrator group showed that only one of their maintained choices was correct (correct rejection-target selection, **Figure 2C**, $\chi^2(2) = 9.69$, $p < 0.01$). It is important to consider that in the case of the Without perpetrator group a maintained correct identification would imply that the participant would have rejected the foils on Day 2 and have chosen the perpetrator on Day 8 while participants in the With perpetrator group have access to the perpetrator in both tests. We further analyzed the incorrectly maintained choices of the Without perpetrator group. We observed that 39% ($N = 12$) of the participants did not repeat the same choice in the Final lineup. However, 61% ($N = 19$) chose the same foil twice, significantly different to the choice level [$\chi^2(1) = 13.33$, $p = 0.0003$, $\phi_c = 0.46$].

Regarding the Initial lineup, 36% (14) of the participants in the With perpetrator group achieved a target selection in the Initial lineup, while only 6% ($N = 2$) of the Without perpetrator group correctly rejected the lineup.

To verify the accuracy-confidence relationship in the Initial and Final lineup, Point Biserial Correlation was applied (Krug, 2007). Regarding the Initial lineup, the calculation was only performed with the With perpetrator group (due to the low frequency of correct rejection of the lineup for the Without perpetrator group) and no significant associations were found ($r_{pb} = 0.240$, $p = 0.140$). In the Final lineup no significant relationships were found, both for the With perpetrator group ($r_{pb} = 0.003$, $p = 0.987$), and for the Without perpetrator group ($r_{pb} = 0.073$, $p = 0.685$). In addition, the CAC curves were performed to comprehensively evaluate the confidence-precision relationship. Confidence was divided into low confidence (0–50%), medium confidence (60–80%) and high confidence (90–100%). To calculate the value of the correct proportion corresponding to each confidence level, the following formula was used: # Correct identifications/# Correct identifications + # Incorrect identifications (Mickes, 2015). In the Initial lineup for the target selection, only the With perpetrator group was analyzed, since the other group had no target in the lineup. Correct proportions of 0.11 ($N = 4$), 0.23 ($N = 8$), and 0.05 ($N = 2$) were obtained for low, medium, and high confidence, respectively (**Figure 3A**). In the Initial lineup for the foil selection, the With perpetrator group presented an incorrect proportion of 0.08 ($N = 3$), 0.38 ($N = 13$), and 0.11 ($N = 4$) in low, medium and high confidence, respectively. The Without perpetrator group presented at low confidence an incorrect proportion of 0.71 ($N = 7$), at medium confidence 0.78 ($N = 23$), and in high



confidence an incorrect proportion of 0.66 ($N = 1$) was found (Figure 3B). In the Final lineup for the target selection, the With perpetrator group presented correct proportions of 0.07 ($N = 3$), 0.28 ($N = 11$), and 0.07 ($N = 3$) in low, medium, and high confidence, respectively. The Without perpetrator group presented correct proportions of 0.12 ($N = 4$), 0.09 ($N = 3$), and 0 ($N = 0$) in low, medium, and high confidence, respectively (Figure 3C). Finally, in the Final lineup for the foil selection, the With perpetrator group presented incorrect proportions of 0.17 ($N = 7$), 0.28 ($N = 11$), and 0.10 ($N = 4$) in low, medium and high confidence, respectively. The Without perpetrator group presented incorrect proportions of 0.25 ($N = 8$), 0.46

($N = 15$), and 0.06 ($N = 2$) in low, medium, and high confidence, respectively (Figure 3D).

Symptomatology Scales and Lineup Recognition

We divided the symptomatology scales into high and low scores (depression score: high ≥ 9.17 , low < 9.17 ; anxiety score: high ≥ 35.14 , low < 35.14 ; and sleep quality score: high ≤ 6.78 , low > 6.78). No significant differences were found between the With and Without perpetrator groups in terms of levels of depression [$\chi^2(1) = 2.36$, $p = 0.12$, $V = 0.18$], anxiety at

TABLE 2 | Symptomatology scales.

	With perpetrator	Without perpetrator	t (70)	p
State anxiety (day 1)	34.92 ± 7.54	35.39 ± 9.62	-0.23	0.81
State anxiety (day 2)	32.43 ± 7.84	32.24 ± 7.94	0.10	0.91
State anxiety (day 8)	34.87 ± 8.85	32.75 ± 8.56	1.02	0.30
Trait anxiety	38.02 ± 8.76	37.30 ± 9.31	0.33	0.73
Depression	8.46 ± 5.89	10.00 ± 6.56	-1.04	0.29
Sleep quality	6.69 ± 3.13	6.84 ± 4.22	-0.21	0.83

Mean state anxiety (day 1, 2, and 8), trait anxiety, depression, and sleep quality ± SD. A two-tailed *t*-test was used to compare symptom scale scores between the With/Without perpetrator groups.

day 1 [$\chi^2(1) = 0.14$, $p = 0.91$, $V = 0.14$], anxiety at day 8 [$\chi^2(1) = 0.56$, $p = 0.45$, $\phi_c = 0.09$], and quality of sleep [$\chi^2(1) = 0.21$, $p = 0.65$, $\phi_c = 0.05$]. A summary of the symptomatology scale scores (divided by groups) can be found in **Table 2**.

No significant associations were found between the performance of the With perpetrator group on the Initial lineup, and levels of anxiety at day 1 [$\chi^2(1) = 0.03$, $p = 0.86$, $V = 0.03$], depression [$\chi^2(1) = 0.25$, $p = 0.62$, $V = 0.08$] and sleep quality [$\chi^2(1) = 1.1$, $p = 0.30$, $V = 0.17$]. None of the high/low subgroups of the three symptom scales achieved target selections above the chance level in the Final lineup.

Once again, due to the low number of correct rejections (2) in the Without perpetrator group, the possibility of making comparisons within this group in the Initial lineup was ruled out.

We further differentiated the identifications made in the Final lineup for the With perpetrator group, dividing the participants in high and low symptomatology scores. We observed no significant differences for the target selection between high and low anxiety score on day 8 [**Figure 4A**, $\chi^2(2) = 1.23$, $p = 0.27$, $V = 0.18$], high and low anxiety score on day 1 [**Figure 4B**, $\chi^2(2) = 1.77$, $p = 0.18$, $V = 0.21$], high and low depression score [**Figure 4C**, $\chi^2(2) = 3.75$, $p = 0.053$, $V = 0.31$], high and low sleep quality score [**Figure 4D**, $\chi^2(2) = 1.95$, $p = 0.16$, $V = 0.22$]. However, we find that the target selection for the low anxiety at day 8 were significantly above the chance level [$\chi^2(2) = 5.73$, $p = 0.02$, $\phi_c = 0.39$], also for high anxiety at day 1 [$\chi^2(2) = 4.8$, $p = 0.03$, $\phi_c = 0.39$], high depression [$\chi^2(2) = 6.2$, $p = 0.13$, $\phi_c = 0.51$] and high sleep quality condition [$\chi^2(2) = 5.9$, $p = 0.15$, $\phi_c = 0.41$]. No other subgroup achieved target selections significantly above the chance level ($1.11 < (\chi^2(2) < 2.13$, all $ps > 0.29$).

The same analysis was applied to the Without perpetrator group. We observed no significant differences for the target selection between high and low anxiety score on day 8 [$\chi^2(1) = 0.001$, $p = 0.98$, $V = 0.01$], high and low anxiety score on day 1 [**Figure 4B**, $\chi^2(1) = 0.79$, $p = 0.38$, $V = 0.16$], high and low depression score [**Figure 4C**, $\chi^2(2) = 0.11$, $p = 0.74$, $\phi_c = 0.06$], high and low sleep quality score [**Figure 4D**, $\chi^2(2) = 1.87$, $p = 0.17$, $\phi_c = 0.24$]. None of the high/low subgroups of the three symptom scales achieved target selections above the chance level in the Final lineup [$0 < (\chi^2(2) < 0.85$, all $ps > 0.36$].

Symptomatology Scales and Episodic Memory Recall

We found that those participants with high sleep quality, perform better in the chronological order task than those with low sleep quality (**Figure 5A**, $Mdn = 3$, $Mdn = 3$, respectively. Mann-Whitney *U*-test, $U = 501.50$, $p = 0.03$, $RB = 0.23$).

In addition, we observed that those participants with high sleep quality, provide a greater number of incorrect details during the Initial recall than those with low sleep quality (**Figure 5B**, $Mdn = 0$, $Mdn = 0$, respectively. One-tailed Mann-Whitney *U*-test, $U = 513.50$, $p = 0.03$, $RB = 0.21$). We further found that those subjects who exhibited a high degree of depressive symptoms, reported a smaller drop in the number of details, between the Initial recall and the Final recall with respect to those who showed a low depressive level (**Figure 5C**, $M = -0.68$, $SD = 9.77$, $M = 5.23$, $SD = 11.47$, respectively. Two-tailed *t*-test (70) = 2.25, $p = 0.02$, $d = 0.26$). No other significant differences were found for the recall variables grouped by the levels of the symptom scales ($-1.31 < t(70) < 1.29$, all $ps > 0.08$).

DISCUSSION

This study was a first step toward understanding how lockdown by COVID-19 pandemic context may influence eyewitness identifications and episodic memory formation. We first found that participants in the With perpetrator group who exhibited high anxiety on the first day, selected the target in the Final lineup above the chance level. On the contrary, those who showed a low degree of anxiety on day 8, selected the target in the Final lineup above the chance level. These results go in line with the model proposed by Schwabe et al. (2012), pointing out that stress has a differential effect on each memory phase, it facilitates memory acquisition and consolidation but impairs memory recall. However, these results were not observed for the Without perpetrator group which showed no significant difference to the chance level in the selection of the target in the final lineup independent of the level of anxiety. Thus, the level of anxiety seems to moderate encoding and recall only when the target is present in the initial lineup.

Regarding the recall variables, we observed no significant differences between the anxiety level for neither the free recall nor for the Chronological order task. The differences shown by our data in the anxiety modulation between target selection in the lineup and the episodic memory could be due to the influence of uncontrolled variables such as cognitive overload and test expectancy which could be impacting in a different way the different types of recall (Hall et al., 1976; Flindall et al., 2016).

We further observed that those participants within the With perpetrator group who showed a high degree of depression selected the target above the chance level, but participants with low depression did not. A similar result was obtained for the recall variables, where those participants with a high degree of depressive symptoms had a lower decay in memory change between the Initial recall and the Final recall, than those with low depression. Taken together, these results could be explained as the product of a biased processing in favor of negative content

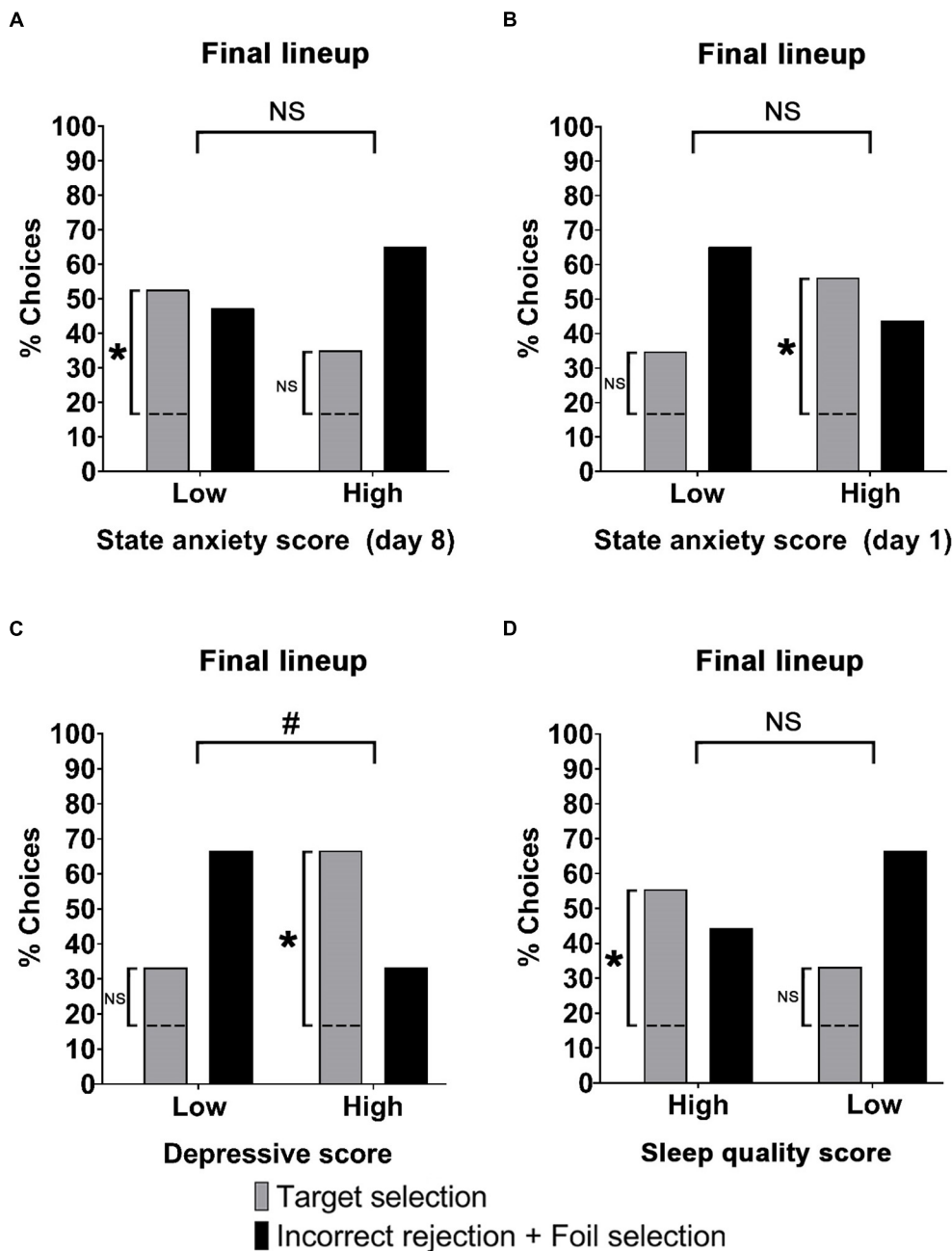
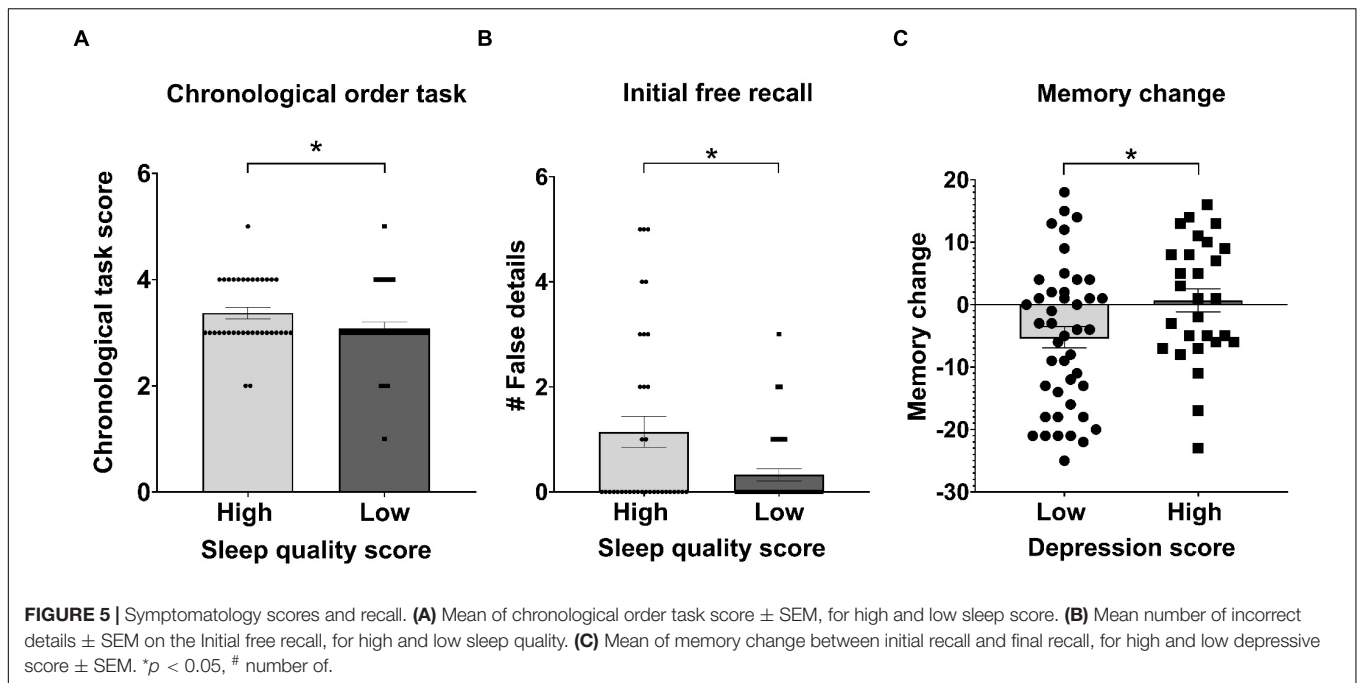


FIGURE 4 | Symptomatology scores and percentage of target selection (correct) and incorrect rejection + foil selection (incorrect) for the With perpetrator group at Final lineup. **(A)** Percentage of correct and incorrect recognitions for high and low anxiety scores on day 8. Low anxiety: correct ($N = 10$) and incorrect ($N = 9$). High anxiety: correct ($N = 7$), incorrect ($N = 13$). **(B)** Percentage of correct and incorrect recognitions for high and low anxiety scores on day 1. Low anxiety: correct ($N = 9$), incorrect ($N = 7$). High anxiety: correct ($N = 8$), incorrect ($N = 15$). **(C)** Percentage of correct and incorrect recognitions for high and low depression scores. Low depressive: correct ($N = 9$), incorrect ($N = 18$). High depressive: correct ($N = 8$), incorrect ($N = 4$). **(D)** Percentage of correct and incorrect recognitions for high and low sleep quality scores. High sleep quality: correct ($N = 10$), incorrect ($N = 8$). Low sleep quality: correct ($N = 7$), incorrect ($N = 14$). * $p < 0.05$; NS, $p > 0.05$. The dashed lines stand for the chance level (17% of correct responses). The perpetrator was included in the Final Lineup of both groups.

(Watkins et al., 1996). Thus, the participants with high depression score would tend to remember the video of a perpetrator entering a conference better than those with low score, given their tendency to strengthen their own negative vision of the world.

Regarding the temporal order, it has been shown that patients with depressive disorder showed an impairment in the temporal order of their episodic memories (Habermas et al., 2008). However, we found no significant differences between



high and low scores of depression for the Chronological order task. Although, this discrepancy could be due to differences in the methodology, since our participants exhibit different degrees of depressive symptoms but none of them reached a pathological level.

In addition, we found that those participants within the With perpetrator group who had high quality of sleep selected the target significantly above change in the Final lineup but this result was not found for the participants in the low condition. This is supported by several studies showing that sleep improves memory acquisition and consolidation (Rasch and Born, 2013). However, we did not find any differences for the Without perpetrator group. It is important to highlight that there are only a few studies analyzing the role of sleep on eyewitness lineup identifications (Stepan et al., 2017; Morgan et al., 2019) and they even showed contradictory results. On one hand, Stepan et al. (2017) found that participants that slept between the training and testing sessions rejected the lineup when the perpetrator was absent significantly more than if they stayed awake. They found no significant difference for the condition where the perpetrator was present in the lineup independently of the sleep/wake condition. On the other hand, in a similar procedure, Morgan et al. (2019) did not observe differences between groups of participants that either slept after the training or remained awake. Thus, the differences between our and their studies could be mainly explained by the methodology used. In those studies, a short and controlled period of sleep deprivation is used, while our work is based on prolonged periods of low sleep quality, which arises spontaneously as a consequence of environmental conditions of the COVID-19 pandemic.

Regarding the recall variables, we observed that participants with high sleep quality recalled more incorrect details on day 1

than participants with low quality. We would expect that a low quality of sleep will be related to a decrease in source monitoring inducing more false memories (Fenn et al., 2009). However, it is important to consider that the more information is remembered the greater the probability to form false memories (Otgaar et al., 2019). Although we did not observe a significant increase in correct details on day 1 for the high-quality sleep condition, the mean was higher. Thus, we suggest that this could be affecting false memory formation on day 1.

In addition, we found that those subjects with high sleep quality obtained better results in the Chronological order task. This highlights the widely accepted fact that adequate sleep is conducive to learning (Rasch and Born, 2013). Both, synaptic homeostasis, and active memory consolidation, may explain the better performance of those individuals who sleep adequately (it is likely that both factors act in combination). In this way, an adequate sleep regimen favors not only the acquisition of new information, but also its correct storage and its persistence in time. Furthermore, it has been observed that sleeping after learning emotional stories favors the consolidation of temporal order (Groch et al., 2011).

The results suggest that simply being exposed to an innocent suspect in an intervening lineup, whether that innocent suspect is identified by the witness or not, increases the probability of misidentifying the innocent suspect and decreases the probability of correctly identifying the true perpetrator in a subsequent test lineup.

It has been largely demonstrated that multiple lineups would increase the chance to identify an innocent as a suspect and decrease the probability of correctly identifying the true perpetrator (Hinz and Pezdek, 2001). Here we replicated those findings showing that in the Final lineup, the Without perpetrator

group had significantly fewer target selections than the With perpetrator group. This is not surprising, considering that both groups tended to the same extent to repeat their choices in both lineups, but the Without perpetrator almost invariably performed foil selections in the Initial lineup. These results can be explained as a product of the “compromise effect,” since most of the subjects in the Without perpetrator group failed during the Initial lineup (choosing a foil, rather than rejecting the lineup), and 61% of them chose the same foil in the Final lineup (even in front of the real perpetrator), maybe for a compromise with their previous choice. Another possible explanation is the “transference effect.” In our double lineup, a case similar to those observed in Mugshot studies could occur, where the participants remember more vividly the face selected in the Initial lineup (whether it is the perpetrator or not), and then they will tend to repeat their choice during the Final lineup, based on a memory of doubtful origin (not clear if they remember the face of the original event, or the Initial lineup). It is also possible that, in the case of the Without perpetrator group, some features of the foils present during the Initial lineup, which were like those of the perpetrator, generated a prediction error, allowing memory labilization, and causing an updating during reconsolidation, incorporating erroneous information from the faces present in the lineup into the original memory.

When considering the low overall performance of both groups, the possibility of explaining the results because of low encoding level, derived from contextual conditions, should also be considered. It has been extensively documented how an individual's state at the time of learning can affect their ability to acquire new information, both positively and negatively (Tyng et al., 2017). In this regard, the current context of lockdown due to the COVID-19 pandemic can be considered mentally and physiologically demanding. If, because of this disturbance, the participants of this study arrive at the Initial lineup with weakly encoded information, a larger number of new details would be incorporated, as if it were a second round of learning (in case the original encoding was minimal, we would really be facing a new learning). However, the Without perpetrator group would only have foils available to encode, and this would explain the difference in the performances observed in the Final lineup.

Concerning the confidence-accuracy relationship, the curve did not seem to follow a clear trend (Sauerland and Sporer, 2007). Although previous studies have observed that lack of sleep can negatively affect the strength of the relationship between confidence and accuracy (Blagrove and Akehurst, 2000) our results are not sufficient evidence to conclude that the relationship between precision and confidence may be affected by the set of negative changes in people's mental health, like changes in the quality of sleep, increased anxiety, and depression.

Among the limitations of our study, it stands out that the size of the sample could have prevented a more detailed and reliable analysis that would take advantage of more subdivisions of the variables. The approach in real context, offered us a

more ecological model, but a less controlled environment. As a result, the intervention of variables not contemplated should not be completely ruled out when considering the results. Additionally, unlike our experimental situation, in a real-life episode, where an individual experiences a violent crime, the sleep disturbances are not likely to appear until after the event, so they would not have an impact on the encoding. Finally, not having pre-lockdown measures in our specific population forces us to speculate based on the trend observed in other populations, and although there are strong reasons to think that anxiety, depression, and the quality of sleep were modified during this period, it is not possible to prove it undoubtedly.

The phenomenon of false memories is complex and multi-determined. It is possible that this work has been able to reflect a portion of that complexity, by showing how different contextual and individual variables interact dynamically to end up in a complex result. From this point, it is essential to move toward a more careful handling of each of the multiple factors mentioned. In particular, regarding identification performance, the variation of the exposure times to the initial stimulus, as well as the manipulation of the degree of similarity between the faces that makes up the lineup, would be a promising horizon toward which to advance in future experiments.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article is available in <https://zenodo.org/record/5548028#.YVsjLJrMLIV>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Institute of Translational Medicine Research Alberto C. Taquini Buenos Aires University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FU, CL, MB, and CF made substantial contributions to the conception and design of the work. FU ran the experiments. FU and PF-K performed the statistical analyses. FU, CL and CF contributed by drafting the work. FU, CL, MB, PF-K, and CF contributed to revising it critically. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a Research Prize from the Instituto Tecnológico de Buenos Aires (ITBA) given to CF.

REFERENCES

- Barros, M., Lima, M., Malta, D., Szwarcwald, C., Azevedo, R., Romero, D., et al. (2020). Relato de tristeza/depressão, nervosismo/ansiedade e problemas de sono na população adulta brasileira durante a pandemia de COVID-19. *Epidemiol. Serv. Saúde* 29:e2020427. doi: 10.1590/S167949742020000400018
- Beck, A., Steer, R., and Brown, G. (1996). *Manual for the Beck Depression Inventory II (BDI-II)*. San Antonio: Psychological Corporation. doi: 10.1037/t00742-000
- Blagrove, M., and Akehurst, L. (2000). Effects of sleep loss on confidence-accuracy relationships for reasoning and eyewitness memory. *J. Exp. Psychol. Appl.* 6, 59–73. doi: 10.1037/1076-898X.6.1.59
- Born, J., and Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychol. Res.* 76, 192–203. doi: 10.1007/s00426-011-0335-6
- Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-Trace Theory and False Memory. *Curr. Dir. Psychol.* 11, 1–9. doi: 10.1111/1467-8721.00192
- Brigham, J. C., Meissner, C. A., and Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Appl. Cogn. Psychol.* 13, S73–S92.
- Brigham, J. C., Ready, D. J., and Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic Appl. Soc. Psychol.* 11, 149–163. doi: 10.1207/s15324834basps1102_3
- Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., et al. (2020). El impacto psicológico de la cuarentena y cómo reducirla: revisión rápida de las pruebas. *Lancet* 395, 912–920.
- Buyse, D. J., Reynolds, C. F. III, Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 28, 193–213. doi: 10.1016/0165-1781(89)90047-4
- Clifasefi, S. L., Garry, M., and Loftus, E. F. (2007). *Setting the Record (or video camera) Straight on Memory: the Video Camera Model of Memory and Other Memory Myths*. New York: Oxford University Press. 60–75.
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., and McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law Hum. Behav.* 28, 687–706. doi: 10.1007/s10979004-0565-x
- Diekmann, S., and Born, J. (2010). The memory function of sleep. *Nat. Rev. Neurosci.* 11, 114–26. doi: 10.1038/nrn2762
- Dobson, M., and Markham, R. (1992). Individual differences in anxiety level and eyewitness memory. *J. Gen. Psychol.* 119, 343–50. doi: 10.1080/00221309.1992.9921177
- Edelstein, R. S., Alexander, K. W., Goodman, G. S., and Newton, J. W. (2004). “Emotion and Eyewitness Memory” in *Series in Affective Science. Memory and Emotion*. eds D. Reisberg and P. Hertel (Oxford: Oxford University Press).
- Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., and Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learn. Mem.* 16, 509–13. doi: 10.1101/lm.1500808
- Flindall, I. R., Leff, D. R., Pucks, N., Sugden, C., and Darzi, A. (2016). The preservation of cued recall in the acute mentally fatigued state: a randomised crossover study. *World J. Surg.* 40, 56–65. doi: 10.1007/s00268-015-3317-9
- Forcato, C., Klinzing, J. G., Carbone, J., Radloff, M., Weber, F. D., Born, J., et al. (2020). Reactivation during sleep with incomplete reminder cues rather than complete ones stabilizes long-term memory in humans. *Commun. Biol.* 3:733. doi: 10.1038/s42003-020-01457-4
- Gotlib, I. H., and Joormann, J. (2010). Cognition and depression: current status and future directions. *Annu. Rev. Clin. Psychol.* 6, 285–312. doi: 10.1146/annurev.clinpsy.121208.131305
- Groch, S., Wilhelm, I., Diekmann, S., Sayk, F., Gais, S., and Born, J. (2011). Contribution of norepinephrine to emotional memory consolidation during sleep. *Psychoneuroendocrinology* 36, 1342–50. doi: 10.1016/j.psyneuen.2011.03.006
- Habermas, T., Ott, L. M., Schubert, M., Schneider, B., and Pate, A. (2008). Stuck in the past: negative bias, explanatory style, temporal order, and evaluative perspectives in life narratives of clinically depressed individuals. *Depress. Anxiety* 25, E121–E132. doi: 10.1002/da.20389
- Hall, J. W., Grossman, L. R., and Elwood, K. D. (1976). Differences in encoding for free recall vs. recognition. *Mem. Cogn.* 4:5. doi: 10.3758/BF03213211
- Hinz, T., and Pezdek, K. (2001). The effect of exposure to multiple lineups on face identification accuracy. *Law Hum. Behav.* 25, 185–98. doi: 10.1023/A:1005697431830
- Jahrami, H., BaHammam, A. S., Bragazzi, N. L., Saif, Z., Faris, M., and Vitiello, M. V. (2021). Sleep problems during the COVID-19 pandemic by population: a systematic review and meta-analysis. *J. Clin. Sleep Med.* 17, 299–313. doi: 10.5664/jcsm.8930
- Kaida, K., Niki, K., and Born, J. (2015). Role of sleep for encoding of emotional memory. *Neurobiol. Learn. Mem.* 121, 72–9. doi: 10.1016/j.nlm.2015.04.002
- Krug, K. (2007). The relationship between confidence and accuracy: current thoughts of the literature and a new area of research. *Appl. Psychol. Crim. Just.* 3, 7–41.
- Leibovich de Figueroa, N. B. (1991). “Ansiedad: algunas concepciones teóricas y su evaluación,” in *Teoría y Técnicas de Evaluación Psicológica*, eds M. M. Casullo, N. B. Leibovich de Figueroa, and M. Aszkenazi (Buenos Aires: Psicoteca), 123–155.
- Lin, W., Strube, M. J., and Roediger, H. L. (2019). The effects of repeated lineups and delay on eyewitness identification. *Cogn. Res. Princ. Implic.* 4:16. doi: 10.1186/s41235-019-0168-1
- Lindberg, M. A., Jones, S., Collard, L. M., and Thomas, S. W. (2001). Similarities and differences in eyewitness testimonies of children who directly versus vicariously experience stress. *J. Genet. Psychol.* 162, 314–33. doi: 10.1080/00221320109597486
- Loftus, E. (2003). Our changeable memories: legal and practical implications. *Nat. Rev. Neurosci.* 4, 231–4. doi: 10.1038/nrn1054
- Loftus, E. F. (1976). Unconscious transference in eyewitness identification. *Law Psychol. Rev.* 2, 93–98.
- Malpass, R. S., and Devine, P. G. (1983). “Measuring the fairness of eyewitness identification lineups” in *Evaluating Witness Evidence*. eds S. Lloyd-Bostock and B. Clifford (London: Wiley & Sons). 81–102.
- Malpass, R. S., and Lindsay, R. C. (1999). Measuring lineup fairness. *Appl. Cogn. Psychol.* 13, S1–S7.
- Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *J. Appl. Res. Mem. Cogn.* 4, 93–102. doi: 10.1016/j.jarmac.2015.01.003
- Morgan, C. A. III, Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., et al. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *Int. J. Law Psychiatry* 27, 265–279.
- Morgan, C. A. III, Southwick, S., Steffian, G., Hazlett, G. A., and Loftus, E. F. (2013). Misinformation can influence memory for recently experienced, highly stressful events. *Int. J. Law Psychiatry* 36, 11–7. doi: 10.1016/j.ijlp.2012.11.002
- Morgan, D. P., Tamminen, J., Seale-Carlisle, T. M., and Mickes, L. (2019). The impact of sleep on eyewitness identifications. *R. Soc. Open Sci.* 6:170501. doi: 10.1098/rsos.170501
- Newman, E. J., and Lindsay, D. S. (2009). False memories: what the hell are they for? *Appl. Cogn. Psychol.* 23, 105–1121. doi: 10.1002/acp.1613
- Norris, F. H., and Kaniasty, K. (1994). Psychological distress following criminal victimization in the general population: cross-sectional, longitudinal, and prospective analyses. *J. Consul. Clin. Psychol.* 62, 111–23. doi: 10.1037/0022-006X.62.1.111
- Otgaa, H., Howe, M. L., Muris, P., and Merckelbach, H. (2019). Associative activation as a mechanism underlying false memory formation. *Clin. Psychol. Sci.* 7:2019. doi: 10.1177/2167702618807189
- Ozamiz-Etxebarria, N., Dosil-Santamaria, M., Picaza-Gorrochategui, M., and Idoaga-Mondragon, N. (2020). Niveles de estrés, ansiedad y depresión en la primera fase del brote del COVID-19 en una muestra recogida en el norte de España. *Cadernos De Saúde Pública* 36:e00054020. doi: 10.1590/0102311X00054020
- Payne, J. D. (2014). Seeing the forest through the trees. *Sleep* 37, 1029–1030. doi: 10.5665/sleep.3750
- Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L. W., Wamsley, E. J., Tucker, M. A., et al. (2009). The role of sleep in false memory formation. *Neurobiol. Learn. Mem.* 92, 327–34.
- Pezdek, K., Abed, E., and Cormia, A. (2020). Elevated stress impairs the accuracy of eyewitness memory but not the confidence-accuracy relationship. *J. Exp. Psychol. Appl.* 27, 158–169. doi: 10.1037/xap0000316

- Rasch, B., and Born, J. (2013). About sleep's role in memory. *Physiol. Rev.* 93, 681–766. doi: 10.1152/physrev.00032.2012
- Reyna, V. F., Corbin, J. C., Weldon, R. B., and Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *J. Appl. Res. Mem. Cogn.* 5:1–9. doi: 10.1016/j.jarmac.2015.12.003
- Robinson, L. (1990). Stress and anxiety. *Nurs. Clin. North Am.* 25, 935–43.
- Ross, D. R., Ceci, S. J., Dunning, D., and Toglia, M. P. (1994). Unconscious transference and mistaken identity: when a witness misidentifies a familiar but innocent person. *J. Appl. Psychol.* 79, 918–930. doi: 10.1037/0021-9010.79.6.918
- Rounding, K., Jacobson, J. A., and Lindsay, R. C. L. (2014). Examining the effects of changes in depressive symptomatology on eyewitness identification. *J. Soc. Clin. Psychol.* 33, 495–511. doi: 10.1521/jscp.2014.33.6.495
- Sauerland, M., and Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychol. Crime Law* 13, 611–625.
- Schwabe, L., Joëls, M., Roozendaal, B., Wolf, O. T., and Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neurosci. Biobehav. Rev.* 36, 1740–9. doi: 10.1016/j.neubiorev.2011.07.002
- Shah, S., Mohammad, D., Qureshi, M., Abbas, M., and Aleem, S. (2021). Prevalence, Psychological Responses and associated correlates of depression, anxiety and stress in a global population, during the coronavirus disease (COVID-19) pandemic. *Community Ment. Health J.* 57:1. doi: 10.1007/s10597020-00728-y
- Sheridan, R. L., Johnson, S., Barnett, P., Steare, T., Needle, J. J., Carr, S., et al. (2021). Early impacts of the COVID-19 pandemic on mental health care and on people with mental health conditions: framework synthesis of international experiences and responses. *Soc. Psychiatr. Epidemiol.* 56, 13–24. doi: 10.1007/s00127-020-01924-7
- Spielberger, C. D., Gorsuch, R., and Lushene, R. (1970). *Manual for the Trait-state Anxiety Inventory*. Palo Alto: Consulting Psychologists.
- Stebly, N. K., and Dysart, J. E. (2016). Repeated eyewitness identification procedures with the same suspect. *J. Appl. Res. Mem. Cogn.* 5, 284–289. doi: 10.1016/j.jarmac.2016.06.010
- Stebly, N. K., Tix, R. W., and Benson, S. L. (2013). Double exposure: the effects of repeated identification lineups on eyewitness accuracy. *Appl. Cogn. Psychol.* 27, 644–654. doi: 10.1002/acp.2944
- Stepan, M. E., Dehnke, T. M., and Fenn, K. M. (2017). Sleep and eyewitness memory: fewer false identifications after sleep when the target is absent from the lineup. *PLoS One* 12:e0182907.
- Thorley, C. (2013). The effects of recent sleep duration, sleep quality, and current sleepiness on eyewitness memory. *Appl. Cogn. Psychol.* 27:2938. doi: 10.1002/acp.2938
- Tononi, G., and Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Res. Bull.* 62, 143–50.
- Tyng, C. M., Amin, H. U., Saad, M. N., and Malik, A. S. (2017). The influences of emotion on learning and memory. *Front. Psychol.* 8:1454. doi: 10.3389/fpsyg.2017.01454
- Valentine, T., Davis, J. P., Memon, A., and Roberts, A. (2012). Live showups and their influence on a subsequent video line-up. *Appl. Cogn. Psychol.* 26, 1–23. doi: 10.1002/acp.1796
- Valentine, T., and Mesout, J. (2009). Eyewitness identification under stress in the London Dungeon. *Appl. Cogn. Psychol.* 23, 151–161. doi: 10.1002/acp.1463
- Van Der Werf, Y. D., Altena, E., Schoonheim, M. M., Sanz-Arigita, E. J., Vis, J. C., De Rijke, W., et al. (2009). Sleep benefits subsequent hippocampal functioning. *Nat. Neurosci.* 12, 122–123. doi: 10.1038/nn.2253
- Walker, M. P. (2009). The role of sleep in cognition and emotion. *Ann. N. Y. Acad. Sci.* 1156, 168–197. doi: 10.1111/j.1749-6632.2009.04416.x
- Watkins, P. C., Vache, K., Verney, S. P., Muller, S., and Mathews, A. (1996). Unconscious moodcongruent memory bias in depression. *J. Abnorm. Psychol.* 105, 34–41. doi: 10.1037/0021843X.105.1.34
- Weigold, A., Weigold, I. K., and Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychol. Methods* 18, 53–70. doi: 10.1037/a0031607
- Wells, G. L. (1984). The psychology of lineup identifications 1. *J. Appl. Soc. Psychol.* 14, 89–103. doi: 10.1111/j.1559-1816.1984.tb02223.x
- Wells, G. L., Leippe, M. R., and Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law Hum. Behav.* 3, 285–293. doi: 10.1007/BF01039807
- Yuille, J. C., Davies, G., Gibling, F., Marxsen, D., and Porter, S. (1994). Eyewitness memory of police trainees for realistic role plays. *J. Appl. Psychol.* 79, 931–936. doi: 10.1037/0021-9010.79.6.931

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Urreta Benítez, Leon, Bonilla, Flores-Kanter and Forcato. 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) and the copyright owner(s) 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.



Mediators and Moderators of Reinforced Self-Affirmation as a Method for Reducing the Memory Misinformation Effect

Malwina Szpitalak[†] and Romuald Polczyk^{*†}

Institute of Psychology, Jagiellonian University, Kraków, Poland

OPEN ACCESS

Edited by:

Rui Paulo,
Bath Spa University, United Kingdom

Reviewed by:

Karlos Luna,
National University of Colombia,
Colombia

Dara Mojtahedi,
University of Huddersfield,
United Kingdom

*Correspondence:

Romuald Polczyk
romuald.polczyk@uj.edu.pl

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Forensic and Legal Psychology,
a section of the journal
Frontiers in Psychology

Received: 10 February 2021

Accepted: 22 October 2021

Published: 23 November 2021

Citation:

Szpitalak M and Polczyk R (2021)
Mediators and Moderators
of Reinforced Self-Affirmation as
a Method for Reducing the Memory
Misinformation Effect.
Front. Psychol. 12:666707.
doi: 10.3389/fpsyg.2021.666707

The misinformation effect occurs when an eyewitness includes information in his or her account that is incongruent with the event he or she witnessed, and stems from being exposed to incorrect external sources. This is a serious threat to the quality of witness testimony and to the correctness of decisions reached by courts. However, few methods have been developed to reduce the vulnerability of witnesses to misinformation. This article presents such a method, namely, reinforced self-affirmation (RSA), which, by increasing memory confidence of witnesses, makes them less inclined to rely on external sources of information and more on their own memory. The effectiveness of this method was confirmed in three experiments. It was also found that memory confidence, but not general self-confidence, is a mediator of the impact of RSA on misinformation effect (ME), and that contingent self-esteem and feedback acceptance, but not sense of self-efficacy or general self-esteem, are moderators of this impact. It is concluded that RSA may be a promising basis for constructing methods, which can be used by forensic psychologists in real forensic settings.

Keywords: feedback, memory, misinformation effect, reducing suggestibility, reinforced self-affirmation, witness testimony

INTRODUCTION

Misinformation of various kinds is very commonplace in our lives, and it is difficult to undo its influence (Walter and Murphy, 2018). Misinformation is also influential in the context of eyewitness testimony (Luna and Martín-Luengo, 2012). Given the dramatic effects, which distorted testimony can have on judicial decisions, including wrongful convictions and acquittals of real perpetrators, it is mandatory to construct methods that can make eyewitnesses more resistant to misinformation.

The present paper explores one such method: reinforced self-affirmation (RSA). This is a way of reducing the memory misinformation effect (ME), which consists in including testimony information, which does not stem from a given event but from other sources. ME is typically studied within a three-stage experimental paradigm (seminal research: Pezdek, 1977; Loftus et al., 1978) in which participants are first exposed to some original material. It can be a video clip (e.g., Cohen and Harnick, 1980), a series of slides (e.g., Loftus et al., 1978), an audio recording (e.g., Szpitalak and Polczyk, 2010), or a text to be read (e.g., Hertel et al., 1980). After some time, participants are exposed to post-event material; for example, they read a description of the original

material, which, in the experimental group, contains some information, which is inconsistent with the original event (e.g., Zaragoza and Lane, 1994), or they answer a series of questions, which, in the experimental group, contain some incorrect premises (e.g., Loftus et al., 1978). In some research, live confederates presented misinformation while interacting with the real participants (e.g., Hope et al., 2008; Mojtahedi et al., 2019a,b). Subsequently, the participants answer a series of questions about the original event, including critical questions relating to the misinformation. In almost all experiments of this kind, it has been found that participants in the experimental group perform worse on the final memory test as they usually include some of the misinformation in their answers (for a review, see Zaragoza et al., 2007).

Many theoretical explanations of the nature of ME for the misinformation effect have been proposed, starting with the classical theories stating that misinformation overwrites the original memory trace (Loftus, 1975) or, in a way, “integrates” into the original memory (Loftus et al., 1978). Another explanation, rooted in the activation-based memory model, stated that, as a result of the misinformation, there are *two* memory traces attached to the critical event, one for the original and one for the misleading information, and activation is shared by the traces, so either could be given as a response (Ayers and Reder, 1998). Another explanation was based on the retrieval-based explanation of forgetting and stated that original information and the misinformation coexist in memory, the latter making the former more difficult to retrieve (Bekerian and Bowers, 1983; Bowers and Bekerian, 1984). Yet another explanation was based on the fuzzy-trace theory and posited that false memories occur primarily because gist memories are falsely ascribed to experience (Reyna and Lloyd, 1997).

Nowadays, it seems that the most popular theoretical explanation of the misinformation effect is the source monitoring theory, which posits that the participants confuse information stemming from the postevent material with their real memories of the original event; in other words, they misattribute the source of their information (e.g., Lindsay and Johnson, 1989; Zaragoza and Koshmider, 1989; Zaragoza and Lane, 1994; Cann and Katz, 2005; Higham et al., 2011). One of the most sophisticated versions of the source-monitoring accounts, including the model Composite Holographic Associative Recall Model (CHARM), was presented by Dodhia and Metcalfe (1999). It explains source monitoring errors in terms of the implications of retrieving a superimposed representation that contains both the original events superimposed on the misleading suggestion (the van).

All the above-mentioned explanations of the misinformation effect share the core assumption that there is some kind of memory malfunction caused by misinformation. However, there is strong empirical evidence confirming that this is not necessarily the case, and people can give memory accounts consistent with misinformation even if there is nothing wrong with their memory. First of all, McCloskey and Zaragoza (1985) presented a strong theoretical and empirical case, showing that, for the misinformation effect to occur, it is enough that two fractions of participants are present: (1) those who, at the moment of the final memory test, do not remember the original information

(for example, because it was never encoded) and/or (2) those who remember *both* the original and misleading information, and answer in accordance with the latter. McCloskey and Zaragoza (1985) presented their participants with two options in the final memory test: the one is consistent with the original information; the second one is inconsistent neither with the original nor with the postevent information (instead of an option consistent with misinformation). No differences were present between the misled and control conditions, which undermine any explanations of the misinformation effect that are based on the memory impairment hypothesis.

Moreover, there are experimental data directly confirming that ME can occur even if participants do remember the correct information from the original event but still give accounts consistent with the misinformation, probably due to lack of confidence in their own memories (Blank, 1998; Szpitalak and Polczyk, 2010; Polczyk, 2017). For example, in research by Polczyk (2017), the participants were administered the standard procedure for testing for the misinformation effect; *afterward*, they were debriefed and given full explanations about the procedure, and again asked what they saw in the original film and read in the postevent material. Many of those who gave answers consistent with misinformation were perfectly able to correctly indicate what was in both sources; thus, they yielded to misinformation in spite of being aware of the discrepancies between the original and postevent materials. In a broader sense, this is a manifestation of informational influence (Mojtahedi et al., 2019a,b).

Some participants yield to misinformation even if they are allowed to access the original and post-event sources while answering the questions from the final memory test; thus, they simply cannot misremember the content of the sources nor can they misattribute them (Polak et al., 2016).

As the basis for the present research, the theory by Blank (1998) was adopted, called an integrative framework for the analysis of memory and performance (I MP) in eyewitness suggestibility experiments. In short, I MP basically assumes that subjects taking part in experiments concerning the misinformation effect are facing a problem-solving process. When answering the questions on the final memory test, they have to find a solution to a memory task. The solution is based on memory states – available information in memory and on the internal representation of the memory task. In particular, this theory posits that there are participants who have information about the content and the source of the original event, as well as about the postevent material. Such participants are fully aware of the discrepancies between both sources. Provided that they assume consistency – they do not assume that they are deliberately misled – they may adopt different strategies to resolve the perceived discrepancies. In particular, some of them may answer in accordance with the postevent material, for example, because they do not trust their own memories.

There is surprisingly little research on the development of methods that aim to undo the suggestive influence of misinformation or to immunize against it, despite the fact that such research may be extremely useful in real forensic settings. One of the most often explored methods is simply

warning the participants against possible discrepancies between the original and post-event materials (Greene et al., 1982). The efficacy of this method varied considerably; in some research, it was not effective at all (e.g., Greene et al., 1982; Neuschatz et al., 2001); in some others, it seemed to reduce the misinformation effect almost completely (e.g., Highhouse and Bottrill, 1995; Oeberst and Blank, 2012; see also the meta-analysis: Blank and Launay, 2014).

Apart from warning, not many other methods have been researched. In some research, a memory-enhancing technique, the cognitive interview reduced the vulnerability to misinformation among children (Holliday and Albon, 2004) and among elderly people (Holliday et al., 2012). A technique similar to the Cognitive Interview, Self-Administered Interview, also seemed to reduce the misinformation effect (Gabbert et al., 2012). However, such results were not present in research by Centofanti and Reece (2006).

As for other methods, English and Nielson (2010) found that triggering arousal reduced yielding to misinformation. Clifasefi et al. (2007) as well as Parker et al. (2008) showed that a placebo presented to participants as a substance that seemingly enhanced cognitive processes improved their ability to resist misinformation. However, this result was not replicated in research by Nastaj et al. (2019). Wagstaff et al. (2011) found that focused meditation reduced interrogative suggestibility (Gudjonsson, 1997), although no clear results were obtained in the case of the standard three-stage procedure. Another experiment suggested that horizontal saccadic eye movements (but not vertical ones) reduced susceptibility to misinformation in the three-stage paradigm (Parker et al., 2009), and Szpitalak and Polczyk (2014) showed that mental warm-up reduces this susceptibility, while mental fatigue increases it.

As can be seen, there were not many techniques developed for reducing the misinformation effect, and the existing ones were not explored extensively (apart from warning). Moreover, many of them (apart from warning and cognitive interview) are not applicable in real forensic settings; one cannot arouse real witnesses by presenting them with disturbing videos, giving them medicaments, or asking them to make eye movements or meditate. Therefore, an exploration of methods reducing the misinformation effect or undoing the effects of misinformation is still warranted. The present research aims at this direction by exploring one such method: RSA (Szpitalak, 2012).

The basic premise of RSA was the assumption that there is a proportion of participants, which, in fact, do remember the correct original information while performing the final memory test. We further assume that a proportion of such participants gives answers that are consistent with the external misinformation but are inconsistent with their own correct memory due to their lack of confidence in it (Blank, 1998; Van Bergen et al., 2010). It was, therefore, assumed (Szpitalak, 2012) that increased confidence in one's own memories should decrease the tendency to rely on the post-event material in the case of participants who, in fact, are aware of the discrepancies between the original and post-event material but believe their memories regarding the former are wrong and, therefore, prefer to rely on the latter.

The idea that self-confidence may be beneficial in the context of eyewitness memory was based on existing data, which suggest that it is advantageous in various areas. For example, it seems to improve leader performance (Hollenbeck and Hall, 2004), results on reasoning tests (Beckmann et al., 2009), or even intelligence tests (Stankov and Crawford, 1997), other cognitive competences (Kleitman and Stankov, 2007), school achievement (Srivastava, 2013), or oral presentation competences (Al-Hebaish, 2012). Most interestingly, self-confidence proved to be a predictor of reliance on oneself as a source of information (Barber, 2008), and of resisting social pressure (MacBride and Tuddenham, 1965). Also, there is research suggesting a direct link between self-confidence and resistance to suggestion in the context of witness testimony (Vrij and Bush, 2000) and memory conformity (Thorley and Kumar, 2017).

Reinforced self-affirmation is based on two elements: self-affirmation and positive feedback on memory functioning. Self-affirmation is induced by means of having participants write down their greatest achievements in life (see the detailed description in the method below). Such a method has proved effective in inducing self-affirmation in existing research (Schimel et al., 2004). In turn, a positive impact of self-affirmation on self-confidence was also found in research experiments (Petruzzello and Corbin, 1988; Compte and Postlewaite, 2004; Sherman and Cohen, 2006; Takai, 2011). As for positive feedback, there is research suggesting that it increases self-confidence (McCarthy, 1986; Fishbach et al., 2010) and reduces interrogative suggestibility (Tata and Gudjonsson, 1990). In sum, both self-affirmation and positive feedback are promising methods of increasing self-confidence, which, in turn, is expected to reduce the tendency to rely on external sources and, instead, to give reports based on one's own memories.

The efficacy of RSA in reducing ME has been repeatedly confirmed and replicated (Szpitalak, 2012; Szpitalak and Polczyk, 2013, 2015b, 2019a,b). The aim of the present paper is to further replicate its efficacy and provide data concerning the possible mechanisms of its impact. Therefore, the first hypothesis is that RSA will reduce ME. Additionally, some mediators and moderators of this main effect will be studied.

First of all, the main hypothesis concerning RSA is the assumption that it increases self-confidence, which, in turn, results in an enhanced tendency to rely on one own's memories instead of information included in post-event material. If this is so, then a mediation should be present: RSA should affect ME *via* increased self-confidence. Moreover, as the task included in the ME procedure concerns memory, and feedback in RSA also concerns memory, it can be expected that, especially, self-confidence related to memory is involved. Therefore, it was hypothesized that this mediation will be present in the case of memory-related self-confidence but not in the case of general self-confidence.

Some moderators of the impact of RSA on ME were also analyzed. The first one was self-esteem. Individuals with high self-esteem might already have access to a wide range of positive self-feelings (Steele et al., 1993; Dodgson and Wood, 1998; Sherman and Cohen, 2006; Pietersma and Dijkstra, 2012). As such, RSA

might confer little advantage to these individuals in terms of encouraging them to rely on their own memory. By contrast, individuals with low self-esteem might have a more limited array of positive self-feelings that are readily available to them when faced with threatening information. Accordingly, an explicit self-affirmation manipulation might provide an important means of boosting self-esteem for these individuals (see also Spencer et al., 2001; Düring and Jessop, 2015). Thus, self-esteem may be a moderator of the impact of RSA on ME.

The term “contingent self-esteem” (Deci and Ryan, 1995) seems to be very useful, too. Contingent self-esteem “...refers to feelings about oneself that result from – indeed, are dependent on matching some standard of excellence or living up to some interpersonal or intrapsychic expectations” (Deci and Ryan, 1995, p. 32). Contingent self-esteem is dependent on matching standards and is directly linked and dependent on perceived successes and failures (Kernis et al., 1993; Park et al., 2004). Therefore, people with contingent self-esteem should be particularly prone to procedures that aim to increase self-confidence, like RSA.

In light of these considerations, it seems that both the level of self-esteem and its stability should moderate the impact of RSA on ME. Persons with stable, reinforcement-independent self-esteem may be less susceptible to RSA than those with contingent self-esteem, which is dependent on external feedback. It was, therefore, hypothesized that RSA would mainly be effective among participants with contingent self-esteem. Also, it was postulated that RSA would be more effective in the case of low general self-esteem because people with high self-esteem may benefit from RSA less – they are probably already self-confident, and efforts to additionally increase this self-confidence may be less effective.

In a very similar vein, a second moderator was postulated, namely, sense of self-efficacy. Self-efficacy (Schwarzer, 1993) refers to self-perceived general efficacy in coping with various tasks and achieving goals. It was assumed that people who perceive their self-efficacy as high would benefit from increased self-confidence less than those whose sense of self-efficacy is low. It should be so because people who feel that they are effective across a range of tasks and goals may be more relying on themselves and perhaps having higher and more stable self-esteem. In the case of such people, increasing self-confidence may not be particularly effective as this self-confidence is probably already relatively high. In contrast, people perceiving their self-efficacy as low may tend to have lower self-confidence and, in turn, benefit from RSA more. In sum, this would cause self-efficacy to be a moderator of the impact of RSA on ME.

The third moderator analyzed in the present study is acceptance of positive feedback. Ilgen et al. (1979) defined feedback acceptance as “the recipient’s belief that the feedback is an accurate portrayal of his or her performance” (p. 356). There are examples of experiments on positive feedback, which show that its acceptance is, by no means, universal and guaranteed and that such acceptance may influence the results obtained. For example, it was found that the efficacy of feedback when avoiding “harmful” food proved dependent on its acceptance (Scoboria et al., 2008; Bernstein et al., 2011; Mantonakis et al., 2013). Similarly, Anseel et al. (2009) showed that feedback

acceptance influences attitude change. In the present research, it was hypothesized that feedback acceptance would moderate the impact of RSA on ME; this impact will be higher for the participants who believed the feedback.

Three experiments were performed. In each one, the existence of the misinformation effect and the efficacy of RSA were analyzed. In addition, in Experiment 1, memory confidence and general confidence were analyzed as mediators of the impact of RSA on yielding to misinformation. In Experiment 2, both these mediators were analyzed again, and contingent and general self-esteem, as well as the sense of self-efficacy, was studied as a possible moderator. In Experiment 3, memory confidence was analyzed as the mediator, and the efficacy of feedback in RSA, as a possible moderator.

POWER AND SAMPLE SIZE CONSIDERATIONS

Power and sample size analysis was performed by means of the software G*POWER 3.1.9 (Faul et al., 2007). The required sample size was calculated for the power 95% for three effect sizes commonly assumed in such analysis, namely, Cohen $f = 0.10$ (small effect), 0.25 (medium), and 0.40 (large). Denominator $df = 1$ and four groups were assumed (the design in all three experiments was 2×2 , see description below). For the main effects, as well as the interaction, the necessary sample sizes were 1,302, 210, and 84, respectively. Given the resources available, a sample size of about 210 was assumed, sufficient to detect medium and large effects, but the small one. In Experiments 2 and 3, the sample size in the experimental misled groups was increased, as these experiments focused on hypotheses, which could only be analyzed in the misled groups.

GENERAL STRATEGY OF ANALYZING THE DATA

In each of the three experiments, there were three general aims: (1) to replicate the misinformation effect; (2) to replicate the efficacy of RSA; and (3) to explore mediators and moderators of the impact of RSA on yielding to misinformation. The first aim was analyzed by means of the main effect of misinformation in the analysis of variance and the second one by analyzing the interaction between misinformation and RSA and appropriate simple main effects, following the existing guidelines (Rosenthal and Rosnow, 1985; Rosnow and Rosenthal, 1989). The analyses concerning the third aim were performed in the group of misled participants only. The number of answers consistent with misinformation was the dependent variable, reflecting the individual susceptibility to misinformation. RSA was the predictor, and mediators and moderators of its impact on yielding to misinformation were analyzed.

EXPERIMENT 1

The aims of Experiment 1 were to confirm the mediating effect of self-confidence in the impact of RSA on yielding to

misinformation and to replicate the misinformation effect itself. Also, it was expected that self-confidence related to memory would be a statistically significant mediator of the impact of RSA on yielding to misinformation, whereas general self-confidence would not. The latter hypothesis will be tested with the Bayesian approach as it is difficult to prove the non-existence of effects with classical NHST methodology.

The following hypotheses were tested in Experiment 1:

1. The misinformation effect will be present: the number of answers consistent with misinformation will be higher for misinformed participants than for those in the control condition.
2. RSA will be effective: the number of answers consistent with misinformation will be lower for misinformed participants who undergo the RSA procedure than for those who do not.
3. Memory confidence will be a mediator of the impact of RSA on the misinformation effect.
4. General self-confidence will not be a mediator of the impact of RSA on the misinformation effect.

Method – Experiment 1

Participants

Two hundred and thirteen subjects took part in the experiment (125 women, 88 men). Their mean age was 17.4 years ($SD = 0.8$; range: 15–19). The experiment took place during school classes. Various schools were chosen randomly; none of them included participants of previous studies on the misinformation effect (and different schools were chosen for the three experiments). No compensation was given for participation. The consent of the parents was not collected – it was not required in the schools. Two participants failed to complete the memory confidence questionnaire, and one participant failed to complete the general confidence questionnaire.

Materials

The study used a 2- and 1/2-min audio recording of some seemingly planned higher education reforms, prepared by the authors and recorded by a professional actor. These materials have been successfully used in other studies (Szpitalak, 2012). The post-event material was a written description summarizing the audio recording; in the misinformed group, it contained 10 details that were different from or additional to the original material. The final memory test consisted of 19 forced-choice questions; 10 of which were critical: the participants had to choose between the correct option or the option consistent with the misinformation in the form of a Yes/No test. Additionally, a short questionnaire, created by the authors of this study, was applied to measure memory confidence and general self-confidence. It consisted of five questions relating to the current quality of memory, and another five relating to self-confidence, e.g., “At the moment, I am assessing my memory”; “I am assessing my self-confidence at the moment.” (This questionnaire and all other materials are provided in **Supplementary Material**). Answers were given on a 7-point Likert-like scale, from 1: very low to 7: very high.

Procedure

The experiment was conducted during school classes. The participants were told that its purpose was to check opinions of students about the planned reforms in higher education. At the beginning, the participants listened to an audio recording (original material). In the recording, a single person advertises the new reform (see **Supplementary Material**). The instruction asked the participants to listen carefully, without giving additional information. After that, they were asked to answer a few questions about what they thought of the reform proposals that had been presented to them; this was done to support the cover story and was adopted from similar research by Apsler and Sears (1968). The questions did not relate to critical items. Then, after about 12 min, during which filler questionnaires were applied, the participants, under the pretext of refreshing the content of their memories, read the post-event material. Immediately after this, the RSA procedure took place. The first part aimed at inducing self-affirmation. The participants in the RSA group were asked to write down their greatest life achievements, while the other half (the control group) were asked to describe their ways home from school. Afterward, faked positive feedback on their memory quality was provided: all the participants were given a surprise memory task consisting in memorizing as many nouns as possible from a list of 60 nouns in a time period of 2 min. After these 2 min, the lists were removed, and the participants were asked to write down all the nouns they could remember. In the RSA group, the participants wrote the nouns in numbered slots so that they knew exactly how many nouns they were able to remember. In the control group, the slots were not numbered. Next, in the RSA group, the participants were told the “average mean number of nouns usually remembered.” This number was false; it was approximately 1.5 SD lower than the real average. In this way, most participants in this group “learned” that their memory was better than average. In the control group, no feedback was provided. In the next stage, the participants completed a questionnaire to check their general and specific self-confidence regarding the functioning of their memory in order to verify the efficacy of RSA. Next, the final memory test concerning the original material was administered in order to analyze the misinformation effect. At the end, the participants were debriefed.

Thus, the experimental design included two between-subjects factors: misinformation (no misinformation or misinformation present) $\times$ RSA (present or absent).

Results and Discussion – Experiment 1

In order to verify the efficacy of RSA, the differences between the groups in which it was applied as compared to the control groups were analyzed, with memory confidence and general self-confidence as dependent variables. Memory confidence was significantly higher in the RSA group as compared with the control group [$M = 4.83$, $SD = 1.44$ vs. $M = 4.12$, $SD = 1.22$; $F(1,209) = 14.49$, $p < 0.001$, $\eta^2 = 0.07$]. In the case of general self-confidence, no significant effect of RSA was present [$M = 4.53$, $SD = 1.11$ vs. $M = 4.47$, $SD = 0.83$; $F(1,210) = 0.20$, $p = 0.653$,

TABLE 1 | Means (SDs, number of participants) of yielding to misinformation across experimental conditions in Experiment 1.

RSA	No misinformation	Misinformation	Total
Absent	4.44 (1.50, 50)	6.45 (2.11, 51)	5.46 (2.09, 101)
Present	4.08 (1.82, 52)	4.50 (2.59, 60)	4.30 (2.27, 112)
Total	4.25 (1.68, 102)	5.40 (2.57, 111)	4.85 (2.25, 213)

$\eta^2 < 0.01$]. This confirms the efficacy of RSA in the case of memory-related self-confidence.

Descriptive results across experimental conditions are presented in **Table 1** and **Figure 1**.

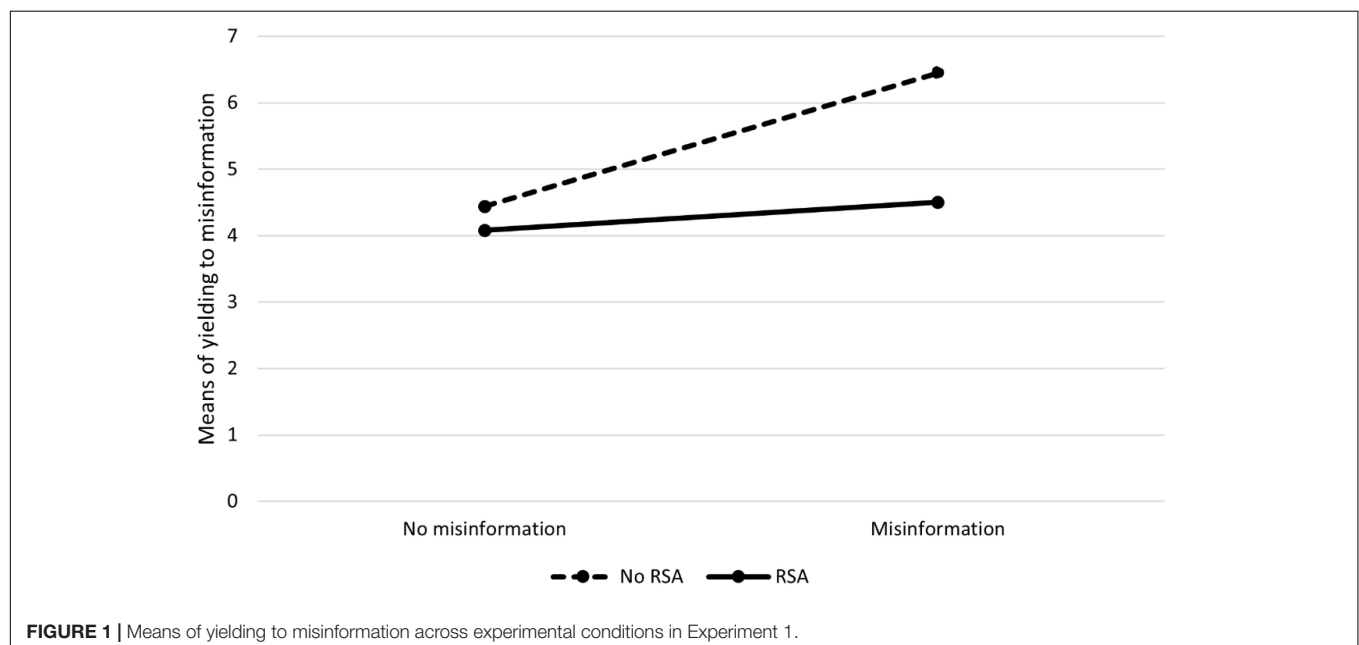
A 2×2 analysis of variance revealed a significant main effect of misinformation, with misinformed participants giving more answers consistent with the misinformation [$F(1,209) = 18.23$, $p < 0.001$, $\eta^2 = 0.08$]. This confirms Hypothesis 1, which concerns the presence of the misinformation effect. The main effect of RSA was also significant, with the participants in the RSA group giving less answers consistent with the misinformation [$F(1,209) = 16.48$, $p < 0.001$, $\eta^2 = 0.07$]. The interaction of the misinformation factor with RSA was also significant [$F(1,209) = 7.76$, $p = 0.006$, $\eta^2 = 0.04$], and the inspection of relevant means in **Table 1** suggests that the fall in the mean number of answers consistent with the misinformation was greater in the misinformed group than in the control group. Indeed, analysis of simple effects confirmed that there was no significant difference between the RSA and no-RSA groups in the condition without misinformation [$F(1,209) = 0.78$, $p = 0.378$, $\eta^2 < 0.01$]. This makes sense as RSA is directed and expected to be effective only in the group of misinformed participants. In this group, its impact was significant [$F(1,209) = 24.38$, $p < 0.001$, $\eta^2 = 0.10$]. This confirms Hypothesis 2; according to which,

a significant fall in the number of answers consistent with the misinformation was expected in the RSA group compared with the group without RSA.

In order to verify Hypothesis 3, a mediation analysis was performed. Bootstrap-generated confidence intervals were calculated to verify the existence of the mediation, as recommended by Hayes (2018). An effect is considered significant when its confidence intervals do not include zero. The PROCESS program (Hayes, 2018) was used to perform this analysis, which was performed in the group of misinformed participants.

The results indicated that a significant impact of RSA on memory confidence was found in the preliminary analysis concerning the manipulation check [$B = 0.89$, $SE = 0.29$, 95% CI (0.32, 1.46)]. The effect of memory confidence on yielding to misinformation was negative and significant [$B = -0.98$, $SE = 0.12$, 95% CI (-1.22, -0.75)]. The indirect effect of RSA on ME via memory confidence was also significant: [$B = -0.87$, $SE = 0.29$, 95% CI (-1.82, -0.34)]. This confirms Hypothesis 3, which states that memory confidence mediates the impact of RSA on ME. Interestingly, the direct effect of RSA on ME was also significant [$B = -1.08$, $SE = 0.37$, 95% CI (-1.82, -0.34)]. This suggests that RSA affects ME not only *via* increased memory-related self-confidence but also through some different mechanisms.

In the case of general self-confidence, its mediating effect was not statistically significant [$B = -0.10$, $SE = 0.11$, 95% CI (-0.35, 0.08)]. As bootstrapping is not the best method of proving the null hypothesis, quasi-Bayesian confidence intervals were also calculated by means of the *brms* package (Bürkner, 2017), running under the R Environment (R Core Team, 2020). The average causal mediation effect (ACME) was -0.10 with 95% confidence intervals: (-0.37, 0.07). This indicates a lack of a mediation effect in accordance with Hypothesis 4.

**FIGURE 1** | Means of yielding to misinformation across experimental conditions in Experiment 1.

In sum, in Experiment 1, all hypotheses were confirmed: the misinformation effect and the efficacy of RSA in reducing it were replicated. Memory confidence proved to mediate the impact of RSA on ME, in congruence with existing data (Szpitalak and Polczyk, 2019b), while general confidence did not. The mediating effect of memory confidence was partial; the direct effect of RSA on ME was significant. This encourages looking for other reasons why RSA may be effective in reducing the misinformation effect, apart from the postulated and confirmed mediating effect of memory confidence.

EXPERIMENT 2

The first three aims of Experiment 2 were similar to those of Experiment 1: replicating the misinformation effect, replicating the efficacy of RSA, and analyzing the mediating role of memory confidence and general confidence in the relationship between RSA and ME. Apart from this, analyses were performed in order to verify whether contingent self-esteem, general self-esteem, and self-efficacy moderate the impact of RSA on ME. The following hypotheses were tested:

1. The misinformation effect will be present: the number of answers consistent with the misinformation will be higher for the misinformed participants than for those in the control condition.
2. RSA will be effective: the number of answers consistent with the misinformation will be lower for the misinformed participants who undergo the RSA procedure than for those who do not.
3. Memory confidence will be a mediator of the impact of RSA on the misinformation effect.
4. General self-confidence will not be a mediator of the impact of RSA on the misinformation effect.
5. Contingent self-esteem will be a moderator of the impact of RSA on ME.
6. General self-esteem will be a moderator of the impact of RSA on ME.
7. Self-efficacy will be a moderator of the impact of RSA on ME.

Method – Experiment 2

Participants

One hundred and seventy-two participants who are students at various schools were tested (125 women and 47 men). Their mean age was 17.3 years ($SD = 0.79$, range 16–19 years). No compensation was given for participation.

Materials and Procedure

The same materials and procedure for the analysis of the misinformation effect and RSA were used as in Experiment 1. In addition, the following tests were applied:

Self-Liking – Competence Scale – Revised (SLCS-R; Tafarodi and Swann, 2001; Polish adaptation: Szpitalak and Polczyk, 2015a). This is a 16-item questionnaire measuring two dimensions of self-esteem: self-competence (e.g., “I am a

capable person”) and self-liking [e.g., “I do not have enough respect for myself” (R)]. Answers are given on a 5-point Likert scale. Higher results mean higher self-esteem and self-confidence, respectively. In the present research, the internal consistencies of both scales as measured by Cronbach alpha were 0.91 and 0.77, respectively.

Contingent Self-Esteem Scale (CSES; Paradise and Kernis, 1999; Polish adaptation: Szpitalak et al., 2018). This is a unidimensional questionnaire consisting of 15 items, e.g., “My overall feelings about myself are heavily influenced by how much other people like and accept me.” The questions are answered on a 5-point Likert scale, from “Not at all like me” to “Very much like me.” Higher results mean that self-esteem is more dependent on external cues. Internal consistency of this scale was 0.87.

Generalized Self-Efficacy Scale (GSES; Schwarzer, 1993; Polish adaptation: Juczyński, 2009). This is a tool designed to measure a general sense of perceived self-efficacy: the belief that one can perform novel or difficult tasks and cope with adversity (e.g., “I can always manage to solve difficult problems if I try hard enough”). It includes 10 items scored on a 4-point scale. Higher results mean that the person perceives them as more capable to cope effectively with tasks and problems. In this experiment, its internal consistency was 0.84.

The procedure was the same as in Experiment 1, but, instead of filler questionnaires, the above-described tools were applied. As previously, the main experimental design included two between-subjects factors: misinformation (no misinformation or misinformation present) $\times$ RSA (present or absent).

Results and Discussion – Experiment 2

Similarly, as in Experiment 1, the groups in which RSA was applied and the control group without it were compared as regards the results of a short questionnaire, measuring memory confidence. The mean memory confidence was significantly higher in the RSA group than in the group without RSA [$M = 5.14$, $SD = 1.54$ vs. $M = 4.18$, $SD = 0.98$; $F(1,170) = 24.46$, $p < 0.001$, $\eta^2 = 0.13$]. This confirms the efficacy of RSA in increasing self-confidence relating to memory quality. No significant effect of RSA was present in the case of general self-confidence [$M = 6.23$, $SD = 1.45$ vs. $M = 5.97$, $SD = 1.34$; $F(1,170) = 1.43$, $p = 0.234$, $\eta^2 = 0.01$]. This confirms the efficacy of the manipulation.

Descriptive results across the experimental condition in Experiment 2 are presented in **Table 2** and **Figure 2**.

A 2×2 ANOVA revealed a significant main effect of misinformation: the number of answers consistent with the misinformation was higher in the misled group than in the

TABLE 2 | Means (SDs, number of participants) of yielding to misinformation across experimental conditions in Experiment 2.

RSA	No misinformation	Misinformation	Total
Absent	3.00 (1.74, 34)	6.51 (1.67, 57)	5.20 (2.40, 91)
Present	3.65 (1.84, 31)	4.72 (2.70, 50)	4.31 (2.45, 81)
Total	3.31 (1.80, 65)	5.67 (2.38, 107)	4.78 (2.46, 172)

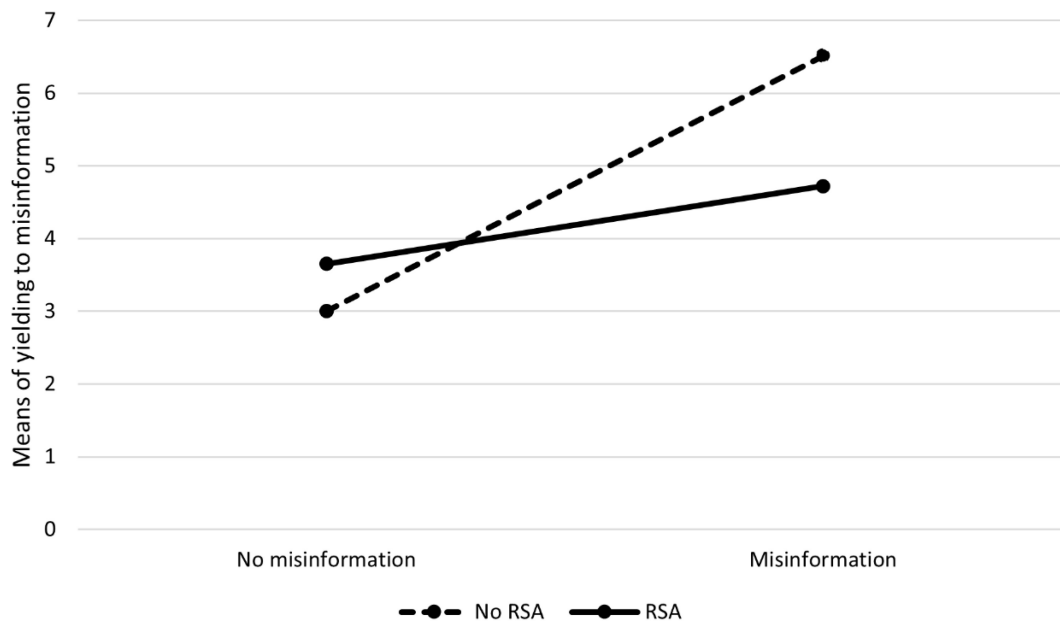


FIGURE 2 | Means of yielding to misinformation across experimental conditions in Experiment 2.

control group [$F(1,168) = 49.73, p < 0.001, \eta^2 = 0.23$]. The main effect of RSA was not significant [$F(1,168) = 3.10, p = 0.080, \eta^2 = 0.02$], but its interaction with the misinformation was significant [$F(1,168) = 14.02, p < 0.001, \eta^2 = 0.08$]. Analysis of the simple effects revealed that the difference between the RSA and non-RSA groups was significant in the case of misled participants [$F(1,168) = 20.01, p < 0.001, \eta^2 = 0.11$]. In the case of non-misled participants, the effect of RSA was not significant [$F(1,168) = 1.59, p = 0.210, \eta^2 = 0.01$]. In sum, these results confirm the existence of the misinformation effect and the efficacy of RSA in reducing it.

To verify Hypothesis 3, which concerns the mediating effect of memory confidence, the same mediation analysis was performed as in Experiment 1. The impact of RSA on memory confidence was significant [$B = 0.65, SE = 0.28, 95\% CI (0.10, 1.21)$], as was the negative effect of memory confidence on yielding to misinformation [$B = -1.10, SE = 0.11, 95\% CI (-1.31, -0.89)$]. The indirect effect was significant [$B = -0.72, SE = 0.33, 95\% CI (-1.39, -0.08)$]. This confirms Hypothesis 3. As in Experiment 1, the direct effect of RSA on yielding to misinformation was also significant [$B = -1.07, SE = 0.31, 95\% CI (-1.68, -0.46)$].

In the case of general self-confidence, the mediation was not statistically significant as the bootstrap 95% confidence intervals included zero: $B = -0.01, SE = -0.15, 95\% CI (-0.29, 0.33)$. The ACME was <0.01 with quasi-Bayesian 95% confidence intervals $(-0.32, 0.33)$. This confirms Hypothesis 4, which states that general self-confidence is not a significant mediator of the impact of RSA on ME.

To verify Hypothesis 5, which states that RSA would be effective mainly in the case of the participants with contingent self-esteem, a moderation analysis was performed with RSA as the predictor, CSES results as the continuous moderator, and yielding

to misinformation as the dependent variable. This analysis was done only in the group of the misinformed participants and was performed by means of the PROCESS software (Hayes, 2018).

The moderating effect of CSES proved significant ($B_{int} = -0.20, SE = 0.04, 95\% CI (-0.27, -0.12)$). To further explore the moderation, Johnson–Neyman cut points were calculated. It turned out that the impact of RSA on ME was significant and positive from the lowest result on CSES up to the value of CSES = 38.9: the participants who underwent RSA scored higher on the ME (that is, they were more suggestible) test than those who did not. In the range of CSES from 40 points to 51, the effect of RSA was not significant. It started to be significant again from the value of CSES = 50.4 and was negative. In sum, these results indicated that, in accordance with Hypothesis 4, RSA is, indeed, helpful in the case of people with high contingent self-esteem. In the case of medium contingent self-esteem, RSA proved not useful; interestingly, in the case of low contingent self-esteem, i.e., stable self-esteem, RSA even increased the ME.

To verify the sixth hypothesis, the potential moderating effects of general self-esteem, as measured by SLCS-R, were analyzed. The moderation was not significant in the case of self-liking [$B_{int} = 0.04, SE = 0.06, 95\% CI (-0.08, 0.15)$] and self-competence [$B_{int} = 0.02, SE = 0.08, 95\% CI (-0.15, 0.18)$].

The seventh hypothesis concerned the moderating effect of self-efficacy. The analysis was performed in the same way as in the case of Hypothesis 6. The moderating effect of self-efficacy was not significant [$B_{int} < 0.01, SE = 0.10, 95\% CI (-0.19, 0.19)$]; thus, the hypothesis was not confirmed.

In sum, the misinformation effect was replicated, and the efficacy of RSA was confirmed. The mediating role of memory confidence was also confirmed; in accordance with

the hypothesis, general self-confidence was not a significant mediator. The effect of RSA was moderated by contingent self-esteem, but not by self-efficacy or general self-esteem.

EXPERIMENT 3

Apart from replicating the ME and the efficacy of RSA, the main aim of Experiment 3 was to analyze the hypothesis that an important moderator of the impact of RSA on ME is feedback acceptance. As elaborated in section “Introduction,” feedback, which is not accepted, cannot be effective. Therefore, it was expected that the efficacy of RSA would be higher in the group of the participants who accepted the feedback. Apart from this, the mediating role of memory confidence was analyzed. The following hypotheses were tested:

1. The misinformation effect will be present: the number of answers consistent with misinformation will be higher for the misinformed participants than for those in the control condition.
2. The RSA will be effective: the number of answers consistent with misinformation will be lower for the participants undergoing the RSA procedure than for those who do not.
3. Memory confidence will be a mediator of the impact of RSA on the misinformation effect.
4. Feedback acceptance will be positively related to the effects of RSA on ME.

In Experiment 3, we decided to increase power to detect mediations and moderations as much as possible. As mediation and moderation analyses are only meaningful in the group of the misled participants, we decided to increase the sample size for the misled condition as much as possible and to use a smaller control group. The latter was only needed to establish the existence of the misinformation effect. Given the resources available, 452 participants were included in the misled group and 94 in the control one.

Method – Experiment 3

Participants

Five hundred and forty-six participants took part in Experiment 3 – 404 women and 142 men; their mean age was 16.8 years ($SD = 1.2$, range: 15–31 years). Most of the participants were students at various high schools. No compensation was given for participation. Two participants failed to complete the memory confidence questionnaire.

Materials and Procedure

The materials, procedure for RSA, and the main experimental design were the same as in the previous experiments. In order to ensure better generalizability of this research, the original material that was used to analyze the misinformation effect was new: it was a video clip presenting a burglary and a robbery, with a duration of about 4.5 min (it was adopted from the movie “Heist” by D. Mamet). The participants were asked to watch it, without any additional information. A description of the film was presented as post-event material “in order to refresh

the memory”; it included six details that were incongruent with the content of the video clip. After the post-event material, the RSA was administered in the same way as in Experiments 1 and 2, followed by a question measuring feedback acceptance: “Does your score accurately reflect your memory capabilities?” The answers were given on a 7-point Likert-like scale, from “Definitely not” to “Definitely yes.” The final memory test consisted of 12 open-ended questions, six of them relating to misled items.

Results – Experiment 3

As in Experiments 1 and 2, it was found that RSA, indeed, increased memory confidence; its means in the groups in which RSA was and was not applied were $M = 4.54$ ($SD = 1.33$) and $M = 4.20$ ($SD = 0.85$), respectively [$F(1,541) = 12.46$, $p < 0.001$, $\eta^2 = 0.02$]. As an additional analysis, the correlation between feedback acceptance and memory confidence was calculated and proved significant: $r = 0.43$, $p < 0.001$. This also confirms the existence of a relationship between the efficacy of experimental manipulations and memory confidence.

Descriptive results in all experimental conditions are presented in **Table 3** and **Figure 3**.

The general effect of misinformation was significant and large [$F(1,542) = 193.63$, $p < 0.001$, $\eta^2 = 0.26$]. The general effect of RSA and its interaction with the misinformation were not significant [$F(1,542) = 0.75$, $p = 0.386$, $\eta^2 < 0.01$ and $F(1,542) = 2.61$, $p = 0.107$, $\eta^2 < 0.01$, respectively]. However, as the hypothesis concerning RSA only applies to misinformed people, planned comparisons were more appropriate. As in Experiments 1 and 2, these comparisons revealed that the participants in the RSA subgroup of the misled group yielded to misinformation significantly less than those in the subgroup without RSA [$F(1,542) = 9.78$, $p = 0.002$, $\eta^2 = 0.02$]. In the non-misinformed group, the difference between the participants who were and were not exposed to RSA was not significant [$F(1,542) = 0.17$, $p = 0.684$, $\eta^2 < 0.01$]. However, the misinformation effect was present both in the group without RSA [$F(1,542) = 164.30$, $p < 0.001$, $\eta^2 = 0.23$] and with RSA [$F(1,542) = 59.76$, $p < 0.001$, $\eta^2 = 0.10$]; although, in the latter case, it was smaller.

In the group of the misinformed participants, the mediation effect of memory confidence was significant [$B = -0.17$, $SE = 0.05$, 95% CI (-0.27 , -0.06)]. As in the previous experiments, RSA increased memory confidence [$B = 0.35$, $SE = 0.11$, 95% CI (0.13 , 0.56)]. Memory confidence reduced yielding to misinformation [$B = -0.48$, $SE = 0.06$, 95% CI (0.59 , -0.37)]. The direct effect of RSA on ME was not significant [$B = -0.25$, $SE = 0.14$, 95% CI (0.51 , 0.02)].

TABLE 3 | Means (SDs, number of participants) of yielding to misinformation across experimental conditions in Experiment 3.

RSA	No misinformation	Misinformation	Total
Absent	0.07 (0.25, 62)	2.62 (1.73, 222)	2.06 (1.86, 284)
Present	0.19 (0.40, 32)	2.21 (1.29, 230)	1.97 (1.38, 262)
Total	0.11 (0.31, 94)	2.41 (1.53, 452)	2.02 (1.65, 546)

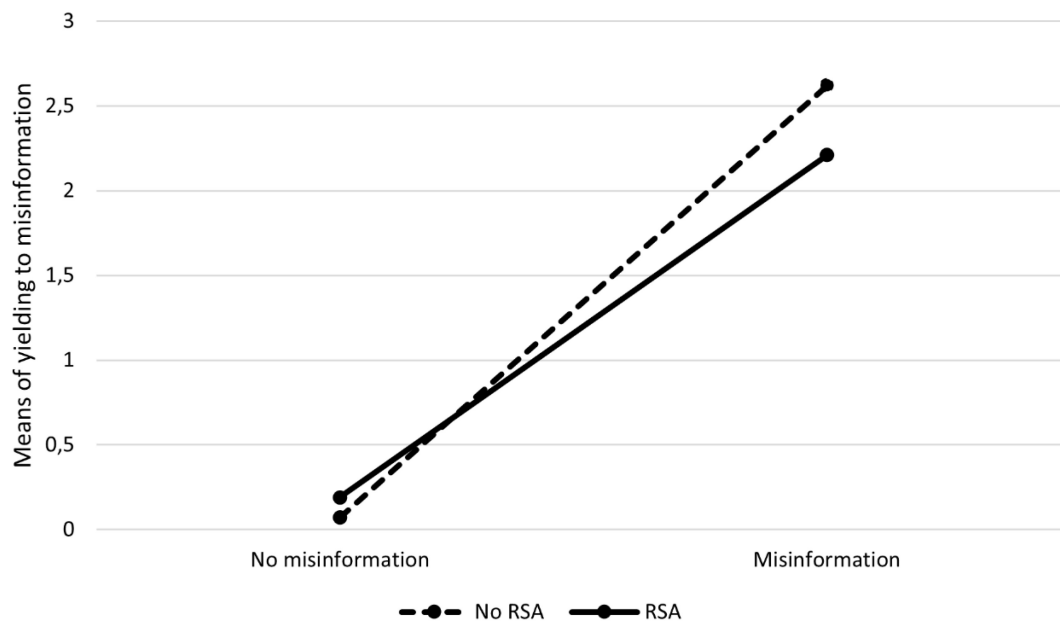


FIGURE 3 | Means of yielding to misinformation across experimental conditions in Experiment 3.

Hypothesis 4 postulated that RSA is effective when feedback is accepted. Feedback acceptance could only be scored in the group with RSA; therefore, no moderation analysis that included RSA was possible, and Hypothesis 4 was analyzed in the group of the misinformed participants who underwent the RSA procedure by means of computing the correlation between the level of feedback acceptance and yielding to misinformation. Notably, there was considerable variance in the measure of feedback acceptance. The answers to the question “Does your score accurately reflect your memory capabilities?” were given on a 7-point scale with the following frequencies: 1 (definitely not) –5.4%; 2–7.2%; 3–12.5%; 4–23.7%; 5–36.2%; 6–11.8%; 7 (definitely yes) –3.2%. The results of the correlational analysis confirmed the hypothesis: Pearson’s r was -0.54 ($p < 0.001$), which indicates that the higher the feedback acceptance, the lower the yielding to misinformation.

In sum, all four hypotheses tested in Experiment 3 were confirmed. However, a caveat is needed here: the lack of interaction between the factors: misinformation and RSA mean that the efficacy of RSA is not certain here, even if the analysis of simple effects confirms this efficacy.

GENERAL DISCUSSION

The main aim of the three experiments presented in this paper was to present further data on RSA, which is a method of reducing the tendency to rely on misinformation when giving memory reports. This tendency, known as the memory misinformation effect, was present in all three experiments. This confirms the robustness and replicability of this effect. This is a warning for justice systems, as the misinformation effect may be an important cause of incorrect testimonies and their consequences.

Not many methods of reducing the misinformation effect have been described. The method presented in this paper, namely, RSA, is intended for witnesses who, in fact, do remember the correct original information yet prefer to rely on external sources even if the information stemming from them contradicts the original information. It was assumed that the reason for such behavior is lack of confidence in one’s own memory. Therefore, RSA aims to increase memory confidence. It proved effective in all three experiments described in the present study. This is a replication of numerous existing studies on its efficacy (Szpitalak, 2012, 2015; Szpitalak and Polczyk, 2013, 2015b, 2019a,b). RSA may be a promising way to develop techniques that are suitable for use in the context of real interrogations.

To be exact, in Experiments 1 and 2, the efficacy of RSA was proved both in the light of its significant interactions with misinformation and simple effects, while, in Experiment 3, the interaction was not significant, although appropriate planned comparisons were significant and consistent with the hypothesis. In Experiment 3, the main original material was different from Experiments 1 and 2. The change was applied in order to ensure better generalizability of the results but may also explain the slightly different results concerning RSA. Recall that the main hypothesis stated that RSA is effective mainly among persons who do realize the differences between the original and postevent materials. Perhaps, there were less such participants due to the change of materials.

In the present study, some possible mediators and moderators of the impact of RSA on ME were studied. First of all, it was assumed that memory confidence would mediate the effect of RSA; this was confirmed in all three experiments and is congruent with other existing data (Szpitalak and Polczyk, 2019b). This was the core hypothesis: RSA should increase confidence in one’s own

memories and, therefore, increase the tendency to rely on them instead of other sources of information. Obviously, this reasoning assumes that some participants remember both the original and the misleading post-event information. As mentioned in section “Introduction,” there are now sufficient empirical data to assume so (Blank, 1998; Szpitalak and Polczyk, 2010; Polczyk, 2017). Interestingly, RSA has already been shown to be effective, particularly among persons who are aware of discrepancies between original and post-event information (Szpitalak and Polczyk, 2015b, Experiment 2).

Apart from acting *via* memory confidence, RSA showed a direct effect on ME. This result should be treated with caution as it only appeared in two out of the three analyses. Nevertheless, apart from mechanisms, which consist in increasing self-confidence, the result encourages considering other possible mechanisms of RSA. As described in section “Introduction,” high self-confidence is beneficial in a wide range of situations. For example, it is possible that it encourages more careful and scrupulous searching of memory. This should be analyzed in further research.

As for moderators, it was hypothesized (and successfully shown) that self-esteem matters as regards the efficacy of RSA. To be exact, our hypothesis concerned both contingent self-esteems, i.e., self-esteem that is highly dependent on external confirmation and general self-esteem. It was assumed that RSA would not be effective among people with stable self-esteem as they are not dependent on and do not need constant confirmation of their value. This hypothesis was confirmed. In contrast, it turned out that general self-esteem was not important for the efficacy of RSA. Overall, this is in agreement with views assuming that self-esteem is not a unitary trait and having generally high self-esteem does not necessarily generalize to all areas and abilities (e.g., Coopersmith, 1967; O’Brien and Epstein, 1988). Similarly, it was also hypothesized that the sense of self-efficacy would have similar effects: the participants with a high sense of self-efficacy might benefit from RSA less than those with lower self-efficacy because they are less dependent on external confirmations of this efficacy. This hypothesis was also not confirmed. One possible reason for this may be the fact that the tool used to measure self-efficacy, GSES (Schwarzer, 1993), was intended to measure a general trait. Such broad self-estimations may not necessarily generalize to specific abilities. It is possible that a tool measuring specific self-efficacy related to memory would moderate the impact of RSA on ME.

Another moderator of the effects of RSA on ME was feedback acceptance. It was hypothesized that, if positive feedback concerning memory is not accepted, memory-related self-confidence would not be enhanced, and the dependency on misleading information would not be reduced. This hypothesis was confirmed. In addition, our results confirm that feedback is not accepted and incorporated automatically. Apparently, it was interpreted and processed cognitively. It is possible, although our data cannot confirm this, that people treat feedback as valid only if it is congruent with their conceptions of self, at least to some degree (Markus, 1977; Swann, 1987). As Esses (1989) suggested, it may also be the case that feedback is accepted when its affective tone matches a mood state of an individual. After

receiving the feedback, a person searches his or her memory to obtain evidence that confirms or discredits the content of the feedback. In a situation in which the memory that confirms the content of the feedback is activated, the person accepts the feedback and is willing to modify his or her self-image in line with the feedback. In a situation in which the subject does not find confirmation for the content of the received feedback, he or she usually rejects its content (Swann, 1987). However, memories that support both negative and positive feedback are usually available (Esses, 1989); in which case, the mood of the person seems to play a key role.

Finally, potential problems with our study should be mentioned. It should be acknowledged the RSA, in general, may not be free from some risks, stemming exactly from increased confidence of a witness. The relationship between confidence and accuracy is complicated (Olsson, 2000; Kebbell, 2009). Increased confidence may be dangerous if a witness has an inaccurate recollection.

Also, another caveat is worth mentioning. As elaborated in section “Introduction,” RSA is expected to be effective mainly among persons who are aware of the discrepancies between the original and post-event information. But we also speculated that warning against discrepancies between both sources is effective among witnesses who are aware of the discrepancies between them. It may be that RSA would be, in a way, redundant with warning in the case of such witnesses.

Furthermore, witnesses who are confident in the quality of their memory would probably benefit from it less.

Also, we are aware that RSA in its present form is of little use for forensic practitioners. It is certainly impossible to provide a real witness with fake positive feedback; this would be impossible for ethical and, probably, also for legal reasons. Having a witness write down his or her greatest achievements in life would also be strange. The present research is, therefore, basic in its nature but can, nevertheless, inspire development of a technique that is suitable for real forensic settings. Efforts to construct such a method are currently in progress.

There may be a problem with the measurement of feedback acceptance. It consisted in asking a question: “Does your score accurately reflect your memory capabilities?” It is possible that a participant may feel that his or her memory was not good but chose to accept the feedback due to perceiving the experiment to be correct.

In Experiments 1 and 2, the final memory test required the participants to make forced responses, without the possibility to refuse an answer or to indicate uncertainty. This is a possible limitation of the present study as, in reality, witnesses normally are (or should be) asked little questions in the form of closed alternatives and are allowed and encouraged to state if they are unsure about certain information.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee at the Institute of Psychology of the Jagiellonian University; Decision Nr KE/01/092018. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MS: conceptualization, empirical research, and writing text. RP: statistical analysis and writing text. Both authors contributed to the article and approved the submitted version.

REFERENCES

- Al-Hebaish, S. M. (2012). The correlation between general self - confidence and academic achievement in the oral presentation course. *Pract. Lang. Stud.* 2, 60–65. doi: 10.4304/tps.2.1.60-65
- Anseel, F., Lievens, F., and Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organ. Behav. Hum. Decis. Process.* 110, 23–35. doi: 10.1016/j.obhdp.2009.05.003
- Apsler, R., and Sears, D. O. (1968). Warning, personal involvement, and attitude change. *J. Pers. Soc. Psychol.* 9, 162–166. doi: 10.1037/h0021248
- Ayers, M. S., and Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychon. Bull. Rev.* 5, 1–21. doi: 10.3758/BF03209454
- Barber, N. (2008). *How Self-Confidence and Knowledge Effects the Sources of Information Selected During Purchase Situations*. Ph. D. thesis. Lubbock, TX: Texas Tech University.
- Beckmann, N., Beckmann, J. F., and Elliott, J. G. (2009). Self-confidence and performance goal orientation interactively predict performance in a reasoning test with accuracy feedback. *Learn. Individ. Differ.* 19, 277–282. doi: 10.1016/j.lindif.2008.09.008
- Bekerian, D. A., and Bowers, J. M. (1983). Eyewitness testimony: Were we misled? *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 139–145. doi: 10.1037/0278-7393.9.1.139
- Bernstein, D. M., Pernat, N. L. M., and Loftus, E. F. (2011). “The false memory diet: False memories alter food preferences,” in *Handbook of Behavior, Food and Nutrition*, eds V. R. Preedy, R. R. Watson, and C. R. Martin (New York, NY: Springer), 1645–1663.
- Blank, H. (1998). Memory states and memory tasks: An integrative framework for eyewitness memory and suggestibility. *Memory* 6, 481–529. doi: 10.1080/741943086
- Blank, H., and Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *J. Appl. Res. Mem. Cogn.* 3, 77–88. doi: 10.1016/j.jarmac.2014.03.005
- Bowers, J. M., and Bekerian, D. A. (1984). When will postevent information distort eyewitness testimony? *J. Appl. Psychol.* 69, 466–472. doi: 10.1037/0021-9010.69.3.466
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *J. Statist. Softw.* 80, 1–28.
- Cann, D., and Katz, A. (2005). Habitual acceptance of misinformation: Examination of individual differences and source attributions. *Mem. Cognit.* 33, 405–417. doi: 10.3758/bf03193059
- Centofanti, A. T., and Reece, J. (2006). The cognitive interview and its effect on misleading postevent information. *Psychol. CrimeLaw* 12, 669–683. doi: 10.1080/10683160600558394
- Clifasefi, S. L., Garry, M., Harper, D. N., Sharman, S. J., and Sutherland, R. (2007). Psychotropic placebos create resistance to the misinformation effect. *Psychon. Bull. Rev.* 14, 112–117. doi: 10.3758/BF03194037

FUNDING

This study was supported by a grant from the Narodowe Centrum Nauki (National Science Centre), Poland: “Mechanisms of reinforced self-affirmation effect,” Nr UMO-2016/23/D/HS6/01677, for MS. All funding for the present study was obtained from this source. No additional external funding was received for this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.666707/full#supplementary-material>

- Cohen, R. L., and Harnick, M. A. (1980). The susceptibility of child witnesses to suggestion: An empirical study. *Law. Hum. Behav.* 4, 201–210. doi: 10.1007/BF01040318
- Compte, O., and Postlewaite, A. (2004). Confidence-Enhanced Performance. *Am. Econ. Rev.* 94, 1536–1557. doi: 10.1257/0002828043052204
- Coopersmith, S. (1967). *The Antecedents of Self-Esteem*. San Francisco, CA: Freeman.
- Deci, E. L., and Ryan, R. M. (1995). “Human agency: The basis for true self-esteem,” in *Efficacy, Agency, and Self-esteem*, ed. M. H. Kernis (New York, NY: Plenum Press), 31–50.
- Dodgson, P. G., and Wood, J. V. (1998). Self-esteem and the cognitive accessibility of strengths and weaknesses after failure. *J. Pers. Soc. Psychol.* 75, 178–197. doi: 10.1037/0022-3514.75.1.178
- Dodhia, R. M., and Metcalfe, J. (1999). False memories and source monitoring. *Cogn. Neuropsychol.* 16, 489–508. doi: 10.1080/026432999380898
- Düring, C., and Jessop, D. C. (2015). The moderating impact of self-esteem on self-affirmation effects. *Br. J. Health Psychol.* 20, 274–289. doi: 10.1111/bjhp.12097
- English, S. M., and Nielson, K. A. (2010). Reduction of the misinformation effect by arousal induced after learning. *Cognition* 117, 237–242. doi: 10.1016/j.cognition.2010.08.014
- Esses, V. M. (1989). Mood as a moderator of acceptance of interpersonal feedback. *J. Pers. Soc. Psychol.* 57, 769–781. doi: 10.1037/0022-3514.57.5.769
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191.
- Fishbach, A., Eyal, T., and Finkelstein, S. R. (2010). How positive and negative feedback motivate goal pursuit. *Soc. Personal. Psychol. Compass.* 4, 517–530. doi: 10.1111/j.1751-9004.2010.00285.x
- Gabbert, F., Hope, L., Fisher, R. P., and Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Appl. Cogn. Psychol.* 26, 568–575. doi: 10.1002/acp.2828
- Greene, E., Flynn, M. S., and Loftus, E. F. (1982). Inducing resistance to misleading information. *J. Verbal. Learning. Verbal. Behav.* 21, 207–219. doi: 10.1016/S0022-5371(82)90571-0
- Gudjonsson, G. H. (1997). *The Gudjonsson Suggestibility Scales*. Hove, UK: Psychology Press.
- Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach*, 2nd Edn. New York, NY: The Guilford Press.
- Hertel, P. T., Cosden, M., and Johnson, P. J. (1980). Passage recall: Schema change and cognitive flexibility. *J. Educ. Psychol.* 72, 133–140. doi: 10.1037/0022-0663.72.2.133
- Higham, P. A., Luna, K., and Bloomfield, J. (2011). Trace-strength and source-monitoring accounts of accuracy and metacognitive resolution in the misinformation paradigm. *Appl. Cogn. Psychol.* 25, 324–335. doi: 10.1002/acp.1694

- Highhouse, S., and Bottrill, K. V. (1995). The influence of social (mis)information on memory for behavior in an employment interview. *Organ. Behav. Hum. Decis. Process.* 62, 220–229. doi: 10.1006/obhd.1995.1045
- Hollenbeck, G. P., and Hall, D. T. (2004). Self-confidence and leader performance. *Organ. Dyn.* 33, 254–269. doi: 10.1016/j.orgdyn.2004.06.003
- Holliday, R. E., and Albon, A. J. (2004). Minimising misinformation effects in young children with cognitive interview mnemonics. *Appl. Cogn. Psych.* 18, 263–281. doi: 10.1002/acp.973
- Holliday, R. E., Humphries, J. E., Milne, R., Memon, A., Houlder, L., Lyons, A., et al. (2012). Reducing misinformation effects in older adults with cognitive interview mnemonics. *Psychol. Aging* 27, 1191–1203. doi: 10.1037/a0022031
- Hope, L., Ost, J., Gabbert, F., Healey, S., and Lenton, E. (2008). “With a little help from my friends...”: The role of co-witness relationship in susceptibility to misinformation. *Acta Psychol.* 127, 476–484. doi: 10.1016/j.actpsy.2007.08.010
- Ilgel, D. R., Fisher, C. D., and Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *J. Appl. Psychol.* 64, 349–371. doi: 10.1037/0021-9010.64.4.349
- Juczyński, Z. (2009). *Narzędzia Pomiaru w Promocji i Psychologii Zdrowia. [Measurement Tools in Promoting Health]*. Warszawa: Pracownia Testów Psychologicznych PTP.
- Kebbell, M. R. (2009). Witness confidence and accuracy: Is a positive relationship maintained for recall under interview conditions? *JIP-OP* 6, 11–23. doi: 10.1002/jip.89
- Kernis, M. H., Cornell, D. P., Sun, C.-R., Berry, A., and Harlow, T. (1993). There's more to self-esteem than whether it is high or low: The importance of stability of self-esteem. *J. Pers. Soc. Psychol.* 65, 1190–1204. doi: 10.1037/0022-3514.65.6.1190
- Kleitman, S., and Stankov, L. (2007). Self-confidence and metacognitive processes. *Learn. Individ. Differ.* 17, 161–173. doi: 10.1016/j.lindif.2007.03.004
- Lindsay, D. S., and Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Mem. Cognit.* 17, 349–358. doi: 10.3758/BF03198473
- Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cogn. Psychol.* 7, 560–572. doi: 10.1016/0010-0285(75)90023-7
- Loftus, E. F., Miller, D. G., and Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 4, 19–31. doi: 10.1037/0278-7393.4.1.19
- Luna, K., and Martín-Luengo, B. (2012). Improving the accuracy of eyewitnesses in the presence of misinformation with the plurality option. *Appl. Cogn. Psychol.* 26, 687–693. doi: 10.1002/acp.2845
- MacBride, P. D., and Tuddenham, R. D. (1965). The influence of self-confidence upon resistance of perceptual judgments to group pressure. *J. Psychol.* 60, 9–23. doi: 10.1080/00223980.1965.10545258
- Mantonakis, A., Wudarszewski, A., Bernstein, D. M., Clifasefi, S. L., and Loftus, E. F. (2013). False beliefs can shape current consumption. *Psychology* 4, 302–308. doi: 10.4236/psych.2013.43A044
- Markus, H. (1977). Self-schemata and processing information about the self. *J. Pers. Soc. Psychol.* 35, 63–78. doi: 10.1037/0022-3514.35.2.63
- McCarthy, P. A. (1986). Effects of feedback on the self-confidence of men and women. *Acad. Manage. J.* 29, 840–847. doi: 10.5465/255950
- McCloskey, M., and Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *J. Exp. Psychol. Gen.* 114, 1–16. doi: 10.1037//0096-3445.114.1.1
- Mojtahedi, D., Ioannou, M., and Hammond, L. (2019a). Intelligence, authority and blame conformity: Co-witness influence is moderated by the perceived competence of the information source. *J. Police Crim. Psychol.* 35, 1–10. doi: 10.1007/s11896-019-09361-2
- Mojtahedi, D., Ioannou, M., Hammond, L., and Synnott, J. (2019b). Investigating the effects of age and gender on co-witness suggestibility during blame attribution. *JIP OP* 16, 153–168. doi: 10.1002/jip.1533
- Nastaj, J., Szpitalak, M., and Bąbel, P. (2019). Can placebo administered in the guise of caffeine reduce the misinformation effect? *Polish Psychol. Bull.* 50, 167–173.
- Neuschatz, J. S., Payne, D. G., Lampinen, J. M., and Togli, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory* 9, 53–71. doi: 10.1080/09658210042000076
- O'Brien, E. J., and Epstein, S. (1988). *The Multidimensional Self-Esteem Inventory*. Odessa, FL: Psychological Assessment Resources.
- Oeberst, A., and Blank, H. (2012). Undoing suggestive influence on memory: The reversibility of the eyewitness misinformation effect. *Cognition* 125, 141–159. doi: 10.1016/j.cognition.2012.07.009
- Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *J. Appl. Psychol.* 85, 504–511. doi: 10.1037/0021-9010.85.4.504
- Paradise, A. W., and Kernis, M. H. (1999). *Development of the Contingent Self-Esteem Scale*. Georgia: University of Georgia.
- Park, L. E., Crocker, J., and Mickelson, K. D. (2004). Attachment styles and contingencies of self-worth. *Pers. Soc. Psychol. Rev.* 30, 1243–1254. doi: 10.1177/0146167204264000
- Parker, A., Buckley, S., and Dagnall, N. (2009). Reduced misinformation effects following saccadic bilateral eye movements. *Brain Cogn.* 69, 89–97. doi: 10.1016/j.bandc.2008.05.009
- Parker, S., Garry, M., Engle, R. W., Harper, D. N., and Clifasefi, S. L. (2008). Psychotropic placebos reduce the misinformation effect by increasing monitoring at test. *Memory* 16, 410–419. doi: 10.1080/09658210801956922
- Petruzzello, S. J., and Corbin, C. B. (1988). The effects of performance feedback on female self-confidence. *J. Sport. Exerc. Psychol.* 10, 174–183.
- Pezdek, K. (1977). Cross-modality semantic integration of sentence and picture memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 3, 515–524. doi: 10.1037/0278-7393.3.5.515
- Pietersma, S., and Dijkstra, A. (2012). Cognitive self-affirmation inclination: An individual difference in dealing with self-threats. *Br. J. Soc. Psychol.* 51, 33–51. doi: 10.1348/014466610X533768
- Polak, M., Dukala, K., Szpitalak, M., and Polczyk, R. (2016). Toward a non-memory misinformation effect: Accessing the original source does not prevent yielding to misinformation. *Curr. Psychol.* 35, 1–12. doi: 10.1007/s12144-015-9352-8
- Polczyk, R. (2017). The ‘memory’ misinformation effect may not be caused by memory failures: exploring memory states of misinformed subjects. *Pol. Psychol. Bull.* 48, 388–400. doi: 10.1515/ppb-2017-0045
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reyna, V. F., and Lloyd, F. (1997). Theories of false memory in children and adults. *Learn. Individ. Differ.* 9, 95–123. doi: 10.1016/S1041-6080(97)90002-9
- Rosenthal, R., and Rosnow, R. L. (1985). *Contrasts analysis: Focused comparisons in the analysis of variance*. Cambridge: Cambridge University Press.
- Rosnow, R. L., and Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychol. Bull.* 105, 143–146. doi: 10.1037/0033-2909.105.1.143
- Schimel, J., Arndt, J., Banko, K. M., and Cook, A. (2004). Not all self-affirmations were created equal: The cognitive and social benefit of affirming the intrinsic (vs extrinsic) self. *Soc. Cogn.* 22, 75–99. doi: 10.1521/soco.22.1.75.30984
- Schwarzer, R. (1993). *Measurement of perceived self-efficacy: Psychometric scales for cross-cultural research*. Berlin: Freie Universität Berlin.
- Scoboria, A., Mazzoni, G., and Jarry, J. L. (2008). Suggesting childhood food illness results in reduced eating behavior. *Acta Psychol.* 128, 304–309. doi: 10.1016/j.actpsy.2008.03.002
- Sherman, D. K., and Cohen, G. L. (2006). “The psychology of self-defense: Self-affirmation theory,” in *Advances in Experimental Social Psychology*, Vol. 38, eds M. P. Zanna and M. P. Zanna (San Diego, CA: Elsevier Academic Press), 183–242.
- Spencer, S. J., Fein, S., and Lomore, C. D. (2001). Maintaining one's self-image vis-à-vis others: The role of self-affirmation in the social evaluation of the self. *Motiv. Emot.* 25, 41–65. doi: 10.1023/A:1010659805978
- Srivastava, S. K. (2013). To study the effect of academic achievement on the level of self-confidence. *J. Psychos. Res.* 8, 41–51.
- Stankov, L., and Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence* 25, 93–109. doi: 10.1016/S0160-2896(97)90047-7
- Steele, C. M., Spencer, S. J., and Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *J. Pers. Soc. Psychol.* 64, 885–896. doi: 10.1037/0022-3514.64.6.885
- Swann, W. B. (1987). Identity negotiation: Where two roads meet. *J. Pers. Soc. Psychol.* 53, 1038–1051. doi: 10.1037/0022-3514.53.6.1038
- Szpitalak, M. (2012). *Motywacyjne Mechanizmy Efektu Dezinformacji [Motivational Mechanisms Of The Misinformation Effect]*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

- Szpitalak, M. (2015). *W kierunku poprawy jakości zeznań świadków. [Towards improving eyewitness testimony]*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Szpitalak, M., and Polczyk, R. (2010). Warning against warnings: Alerted subjects may perform worse. The impact of misinformation, involvement and warning on eyewitness testimony. *Pol. Psych. Bull.* 41, 105–112. doi: 10.2478/v10059-010-0014-2
- Szpitalak, M., and Polczyk, R. (2013). Promoting eyewitness testimony quality: Warning vs reinforced self-affirmation as methods of reduction of the misinformation effect. *Pol. Psych. Bull.* 44, 85–91. doi: 10.2478/ppb-2013-0009
- Szpitalak, M., and Polczyk, R. (2014). Mental fatigue, mental warm-up, and self-reference as determinants of the misinformation effect. *J. Forens. Psychiatry Psychol.* 25, 135–151. doi: 10.1080/14789949.2014.895024
- Szpitalak, M., and Polczyk, R. (2015a). *Samoocena. Geneza, struktura, funkcje i metody pomiaru. [Self-esteem. Origins, structure, functions and measurement]*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Szpitalak, M., and Polczyk, R. (2015b). Reinforced self-affirmation as a method for reducing the eyewitness misinformation effect. *Psychol. Crime Law* 21, 911–938. doi: 10.1080/1068316X.2015.1077243
- Szpitalak, M., and Polczyk, R. (2019a). How to induce resistance to the misinformation effect? Characteristics of positive feedback in the reinforced self-affirmation procedure. *Psychol. Crime Law*. 25, 771–791. doi: 10.1080/1068316X.2019.1574791
- Szpitalak, M., and Polczyk, R. (2019b). Inducing resistance to the misinformation effect by means of reinforced self-affirmation: The importance of positive feedback. *PLoS One* 14:e0210987. doi: 10.1371/journal.pone.0210987
- Szpitalak, M., Polczyk, R. I., and Dudek, I. (2018). Psychometric properties and correlates of the Polish version of the Contingent Self-Esteem Scale (CSES). *Pol. Psych. Bull.* 49, 449–457. doi: 10.24425/119514
- Tafarodi, R. W., and Swann, W. B. (2001). Two-dimensional self-esteem: Theory and measurement. *Pers. Individ. Dif.* 31, 653–673. doi: 10.1016/S0191-8869(00)00169-0
- Takai, N. (2011). Confidence-building and developmental changes in confidence from adolescence to late adulthood. *Jap. J. of Health Psychol.* 24, 45–58. doi: 10.11560/jahp.24.1_45
- Tata, P. R., and Gudjonsson, G. H. (1990). The effects of mood and verbal feedback on interrogative suggestibility. *Pers. Individ. Dif.* 11, 1079–1085. doi: 10.1016/0191-8869(90)90136-F
- Thorley, C., and Kumar, D. (2017). Eyewitness susceptibility to co-witness misinformation is influenced by co-witness confidence and own self-confidence. *Psychol. Crime Law* 23, 342–360. doi: 10.1080/1068316X.2016.1258471
- Wagstaff, G. F., Wheatcroft, J. M., Burt, Ch. L., Pilkington, H. J., Wilkinson, K., and Hoyle, J. D. (2011). Enhancing witness memory with focused meditation and eye-closure: Assessing the effects of misinformation. *J. Police Crim. Psychol.* 26, 152–161. doi: 10.1007/s11896-010-9082-7
- Van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., and Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Appl. Cogn. Psychol.* 24, 885–896. doi: 10.1002/acp.1595
- Vrij, A., and Bush, N. (2000). Differences in suggestibility between 5-6 and 10-11 year olds: The relationship with self-confidence. *Psychol. Crime Law* 6, 127–138. doi: 10.1080/10683160008410837
- Walter, N., and Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Commun. Monogr.* 85, 423–441. doi: 10.1080/03637751.2018.1467564
- Zaragoza, M. S., and Koshmider, J. W. (1989). Misled subjects may know more than their performance implies. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 246–255. doi: 10.1037/0278-7393.15.2.246
- Zaragoza, M. S., and Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 934–945. doi: 10.1037/0278-7393.20.4.934
- Zaragoza, M. S., Belli, R. F., and Payment, K. E. (2007). “Misinformation effects and the suggestibility of eyewitness memory,” in *Do Justice and Let the Sky Fall: Elizabeth Loftus and Her Contributions to Science, Law, and Academic Freedom*, eds M. Garry and H. Hayne (Mahwah NJ: Lawrence Erlbaum Associates Publishers), 35–63.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Szpitalak and Polczyk. 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) and the copyright owner(s) 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.



The Trajectory of Targets and Critical Lures in the Deese/Roediger–McDermott Paradigm: A Systematic Review

Patricia I. Coburn^{1*}, Kirandeep K. Dogra¹, Iarenjit K. Rai¹ and Daniel M. Bernstein^{1,2*}

¹ Department of Psychology, Kwantlen Polytechnic University, Surrey, BC, Canada, ² Simon Fraser University, Burnaby, BC, Canada

OPEN ACCESS

Edited by:

Sara Cadavid,
Rosario University, Colombia

Reviewed by:

Naveen Kashyap,
Indian Institute of Technology
Guwahati, India
Phillip R. Zoladz,
Ohio Northern University,
United States

*Correspondence:

Patricia I. Coburn
patricia.coburn@kpu.ca
Daniel M. Bernstein
daniel.bernstein@kpu.ca

Specialty section:

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

Received: 01 June 2021

Accepted: 26 October 2021

Published: 03 December 2021

Citation:

Coburn PI, Dogra KK, Rai IK and
Bernstein DM (2021) The Trajectory
of Targets and Critical Lures
in the Deese/Roediger–McDermott
Paradigm: A Systematic Review.
Front. Psychol. 12:718818.
doi: 10.3389/fpsyg.2021.718818

The Deese/Roediger–McDermott (DRM) paradigm has been used extensively to examine false memory. During the study session, participants learn lists of semantically related items (e.g., pillow, blanket, tired, bed), referred to as targets. Critical lures are items which are also associated with the lists but are intentionally omitted from study (e.g., sleep). At test, when asked to remember targets, participants often report false memories for critical lures. Findings from experiments using the DRM show the ease with which false memories develop in the absence of suggestion or misinformation. Given this, it is important to examine factors which influence the generalizability of the findings. One important factor is the persistence of false memory, or how long false memories last. Therefore, we conducted a systemic review to answer this research question: What is the persistence of false memory for specific items in the DRM paradigm? To help answer this question our review had two research objectives: (1) to examine the trajectory of target memory and false memory for critical lures and (2) to examine whether memory for targets exceeded false memory for critical lures. We included empirical articles which tested memory for the same DRM lists with at least two testing sessions. We discuss the results with respect to single-session delays, long-term memory recall and recognition, remember and know judgments for memory, and the effect of development, valence, warning, and connectivity on the trajectory of memory. Overall, the trajectory of targets showed a relatively consistent pattern of decrease across delay. The trajectory of critical lures was inconsistent. The proportion of targets versus critical lures across delay was also inconsistent. Despite the inconsistencies, we conclude that targets and critical lures have a dissimilar trajectory across delay and that critical lures are more persistent than targets. The findings with respect to long-term recall and recognition are consistent with both Fuzzy Trace Theory and Associative-Activation Theory of the DRM effect. The generation of false memory with brief delays (3–4 s) is better explained by Associative-Activation Theory. Examining the connectivity between target items, and critical lures, and the effect that has during study and retrieval, can provide insight into the persistence of false memory for critical lures.

Keywords: Deese/Roediger–McDermott (DRM) paradigm, critical lures, targets, trajectory, false memory, Fuzzy Trace Theory, delay, associative-activation theory

INTRODUCTION

Imagine that your partner asks you to shop for produce, providing you with a list of fruits to buy. You forget the list, and then buy what you think was on the list. When you return home with your bounty, your partner asks why you bought oranges. You say that you forgot the list, but you remembered that oranges were on it. This, however, is a false memory, because oranges were not on the list. Memory is vulnerable to errors of omission (information that was present initially but not retrieved later), as well as errors of commission (information that was absent initially but was retrieved later). In the latter category, intrusions can arise internally (self-generated; this is the type of error you made when you remembered that oranges were on the list) or externally through post-event information. Internally and externally generated errors fall in the broad category of false memory, although there is evidence that the two may be unrelated and have distinct underlying mechanisms (Ost et al., 2013; Bernstein et al., 2018; Nichols and Loftus, 2019).

One technique that has been used extensively to examine factors associated with false memory is the Deese/Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger and McDermott, 1995). Findings from research with this paradigm show that false memories develop rapidly and effortlessly (Read, 1996). In a standard DRM paradigm, during the study phase, individuals learn lists of semantically associated words (e.g., grapes, apples, lemons, melon, limes, and strawberries). At test, individuals try to remember the words from the study phase. True memories occur when participants remember the words presented at study – *targets*. Many individuals falsely remember semantically associated words that were absent at study (e.g., oranges). Researchers intentionally omit these associated words from study and refer to these as *critical lures*. Individuals remember critical lures at a higher rate than semantically unrelated words that were also absent at study (e.g., books). The latter are called *distractors* or *foils*. False memory has been tested in the DRM paradigm with different measures of memory, such as recall, recognition, or remember/know judgments. There is an extensive literature on the DRM which illustrates the importance of semantic encoding in memory. The findings also show that individuals form false memories in the absence of external suggestion or post-event information.

The ease with which individuals develop false memories in the DRM paradigm has contributed to research on the fallibility of memory, and the implications of false memory in real-world contexts. Therefore, it is important to study factors which influence the generalizability of the effect. One such factor is the persistence of false memory, or how long false memories last. In real-world contexts, an individual will be required to remember information after a delay, sometimes even after a lengthy delay. As in the example above, even if you provide the initial report quickly, you may have to provide or remember this information weeks or months later. Providing multiple reports may increase the likelihood that a memory error will occur. This is especially true if misinformation or suggestion is introduced (e.g., Loftus et al., 1978; Belli, 1989). However, the DRM paradigm shows that false memories can occur without suggestion. Furthermore, some

research shows that across delay, these false memories remain relatively stable, or even inflate, compared to true memories.

False memories in real-world settings can have serious consequences. Now imagine that 30 min into lunch with a friend, you realize that you have an itchy raised rash on your neck. You are finding it difficult to breathe so you decide to go directly to emergency. There you are quickly greeted by a doctor who gives you an injection of epinephrine. The doctor suggests you write down everything you had at lunch. The task seems relatively straightforward: You had grapes, apples, lemons, melon, limes, strawberries, oranges, and blueberries. However, your memory of eating oranges is false. In this example, falsely remembering an orange at lunch could make you avoid the wrong food in the future, while consuming foods that could induce another potentially fatal reaction.

While it can be argued that the trajectory of true memory (targets) and false memory (critical lures) has practical importance, the trajectory of targets and critical lures in the DRM paradigm also has important theoretical implications. Two dominant theoretical explanations for false memory of critical lures in the DRM are the Associative-Activation Theory (Howe et al., 2009), informed by the activation monitoring theory (Roediger et al., 2001a), and Fuzzy Trace Theory (Reyna and Brainerd, 1995). Associative-Activation Theory suggests that critical lures are activated through spreading activation among pre-existing mental networks (Colins and Loftus, 1975). That is, the activation of a target word during the study phase initiates the activation of other words, including words that were not presented during study.

Alternatively, gist theories suggest that critical lures are generated because individuals extract the underlying meaning associated with the list items; critical lures have high semantic relatedness to the list items (see Gallo, 2010). Fuzzy Trace Theory is one popular gist theory used to explain the trajectory of targets and critical lures in the DRM. According to Fuzzy Trace Theory, information encoded in memory forms two traces: Verbatim and gist (Reyna and Brainerd, 1995; Brainerd et al., 1999). The verbatim trace contains item-specific information, while the gist contains mainly underlying meaning of the information without perceptual details (Brainerd et al., 2006). Fuzzy Trace Theory suggests that false memories arising in the DRM paradigm result from representation of the gist that occurs during encoding of the semantic associates on the lists. Gist memory tends to decay more slowly than verbatim memory (Kintsch et al., 1990; Reyna and Kiernan, 1994, 1995). Hence, if false memory for critical lures is due to gist formation, memory for critical lures and targets should have distinct trajectories. Based on Fuzzy Trace Theory, target memory should decrease more quickly than critical lure memory across delay.

Conversely, research that shows similar trajectories between targets and critical lures could indicate that the Associative-Activation Theory better explains false memory of critical lures. The Associative-Activation theory argues that critical lures are activated during the encoding phase due to the pre-existing associations with the items on DRM lists. The theory generally does not propose differences between targets and critical lures that would result in distinct trajectories across delay. Given the

theory proposes that false memories for critical lures arise due to their association with targets, variation in how the items are associated could affect the persistence of targets and critical lures across a delay. This can include the number of items on the study lists, and the ease with which the critical lures are generated in free association from the targets (backward association strength), as well as the speed or automaticity of activation (Otgaar et al., 2019).

We conducted a systematic review of the empirical studies that examined the trajectory of memory across time using the DRM paradigm. We conducted our review to answer this research question: What is the persistence of false memory for a specific item in the DRM paradigm? Our review had two research objectives: (1) to examine the trajectory of target memory and false memory for critical lures and (2) to examine whether target memory exceeds false memory for critical lures. The answers to these research questions have practical and theoretical importance. Firstly, given that false memories that are generated in lab-based experiments are often generalized to real-world contexts, understanding how long false memories can occur after the encoding phase is critical to understanding the consequences of false memories (Bernstein and Loftus, 2009). Secondly, understanding the trajectory of targets and critical lures provides insight into the theoretical explanations of false memory. Experimental factors may influence the trajectory of memory in the DRM paradigm, including how researchers measure memory, the number of lists they use, the number of items per list, whether they manipulated delay within or between subject, and whether they included an immediate recall condition. We discuss the results with respect to the trajectory of memory (both true and false) from short to long-term, the trajectory of recall and recognition memory across long delays, the trajectory of remember/know judgments, the trajectory of memory in children and youth, and the effect of connectivity on the trajectory of memory.

METHOD

Two co-authors on this review (KD and IR) conducted independent searches in Google Scholar with the key terms, “Delay” AND “Deese-Roediger-McDermott.” KD searched for papers published between 1970 and 2010 and conducted this search between January 6 and January 17, 2021. IR searched for papers published between 2011 and 2021 and conducted this search between January and March 2021. The initial search yielded 1109 hits. KD and IR examined titles and abstracts to determine whether articles met our inclusion criteria: Empirical articles using the DRM paradigm with delay as a manipulated variable (either between or within subject). The first author (PC) developed these inclusionary criteria before the search started; so, we set out to answer this central question. Experiments had to have a minimum of two testing sessions with the same dependent variable at both sessions (i.e., recall followed by recall; or recognition followed by recognition), and the same study lists needed to be used during the testing sessions. We excluded experiments that had additional study sessions at subsequent

testing, because we were interested in the trajectory of memory without additional encoding. There were some articles which manipulated delay across experiments. We included these in our sample but note the limitation to this methodological approach. Our criteria resulted in 38 articles. Finally, we examined the reference lists of the included articles to ensure relevant articles were not overlooked. The latter resulted in two additional articles ($N = 40$). All authors met bi-weekly between December 2020 and May 2021 to discuss the criteria and ensure every paper met the inclusion criteria. If there was any confusion whether an article should be included, all authors read the article, and everyone discussed how it fit with the inclusion criteria. Our Results and Discussion focus on targets and critical lures; we only discuss distractors when relevant to interpreting the results.

RESULTS

Summaries of methods and results for all included articles appear in **Supplementary Materials Table 1**.

Single-Session Delay

The trajectory of memories in the DRM paradigm can yield different patterns. Memories can decrease, remain stable, or increase across delay. This can apply to true and false memories independently or in tandem. We refer to these patterns as decrease, stability, and inflation. Regardless of which trajectory true and false memories follow, the proportion of true memories compared to false memories can also vary. Individuals may report proportionally more true than false memories, more false than true memories, or equal numbers of true and false memories.

Perhaps unsurprisingly, the trajectory of false memory depends on the length of delay being manipulated. More surprising to readers unfamiliar with the DRM is that false memories can occur using this paradigm in as little as 3–4 s. And while false memories have been consistently observed with brief delays (McDermott, 1996; Flegal et al., 2010; Festini and Reuter-Lorenz, 2013), the trajectory of true and false memories is inconsistent. For example, in one study, participants viewed 4-item lists of semantically related words and were then probed with a single word (target, unrelated distractor, or critical lure) after a short-term 3–4 s delay or on a surprise long-term recognition test occurring approximately 20 min later (Flegal et al., 2010). The results showed that target recognition decreased from the short-term to the long-term tests while critical lure recognition remained stable across these tests. Despite this, target recognition remained higher than critical lure recognition across delay. Other studies have shown delayed inflation of critical lures with short delays (8–20 min). For example, Festini and Reuter-Lorenz (2013) observed delayed inflation of critical lure recognition across an 8-min delay. As well, Olszewska et al. (2015) observed inflation of critical lure recognition from short-term (3–4 s delay) to long-term (20-min delay) memory. This shows that delayed inflation of critical lures can occur between short and long-term memory.

In many of the studies examining target and critical lure memory within brief periods, participants viewed short lists of words and then completed recognition tests. Recognition tests sometimes yield higher rates of critical lures compared to recall tests (e.g., McEvoy et al., 1999; Stadler et al., 1999). Importantly, the decrease in target recognition and stability of critical lure recognition across a brief delay has also been shown in recall. For example, McDermott (1996) had participants listen to 24 lists. Following each list, participants either recalled the words immediately or after a 30-s delay. Target recall was higher than false recall of critical lures at immediate recall. However, target recall decreased from immediate to delayed recall, while critical lure recall remained stable. Moreover, after the 30-s delay, rates of target and critical lure recall were equivalent. These results are consistent with McEvoy et al. (1999; Experiment 3) who observed lower target recall and stable critical lure recall after a 1-min delay.

The studies described thus far show that false memories for critical lures develop rapidly. This could support gist, or semantic encoding even in short-term memory. These findings are somewhat surprising given that at short delays, memory for the original lists should be strong and permit individuals to recognize memory errors of commission. Recall and recognition of critical lures in such brief periods could support the Associative-Activation Theory of false memory.

Despite the inconsistency and range of methodologies used to examine the trajectory of memory with brief delays, the results are relatively consistent: Target memory begins to decline rapidly, while critical lure memory remains stable, and possibly increases with delays of 20 min or less. The dissimilar trajectory of targets and critical lures across a 20-min delay is consistent with Fuzzy Trace Theory, however, it is arguably also consistent with Association-Activation Theory. Repeated activation of a critical lure in memory, due to multiple associated targets being presented during study, could result in a stronger representation of the critical lure than targets after a delay. This could depend on the association strength between the targets and critical lure.

Long-Term Recall

When examining the trajectory of targets and critical lures in long-term memory, some research shows that target recall declines more rapidly than false recall of critical lures (Payne et al., 1996; Brainerd et al., 2003; Sherman and Kennerley, 2014; although see Pardilla-Delgado and Payne, 2017). For example, Brainerd et al. (2003) observed delayed inflation of critical lures but not targets over three testing sessions (2-min filler, 5-min test/session). Additionally, participants recalled critical lures at higher rates than targets across all testing sessions. Brainerd et al. (2003) proposed that repeated attempts at recalling semantically related lists provides an opportunity to practice the gist recall processes. With no further opportunity to study the lists, there would be no expectation to strengthen memory for the targets; target memory or true memory relies on access to the verbatim traces rather than constructive processes associated with the

gist memory processes (Brainerd et al., 2003). Alternatively, providing an opportunity to access the verbatim traces increases target memory. In a follow-up experiment, increasing the study sessions from one to three, when paired with increased testing, yielded higher rates of target memory than critical lure memory (Brainerd et al., 2003). One might expect, however, that inflation should emerge on targets as well as critical lures across three tests in a single session (21-min) because of the testing effect. Research on the testing effect shows that retrieving information (e.g., through testing) improves memory on subsequent tests (e.g., Roediger and Butler, 2011). Indeed, the testing effect has been shown to be relevant for target memory in the DRM. For example, McDermott (1996) reported that recall of targets and critical lures was higher when there was a previous testing session compared to when there was no previous testing session. However, even though the testing effect emerged, target recall and false recall of critical lures decreased across a 2-day delay. This decrease was smaller for targets and critical lures if participants had previously completed a recall test than if they had not (McDermott, 1996).

Critical lure recall has been shown to be more stable than target recall across even longer delays. For example, Toglia et al. (1999) presented participants with five auditory lists of semantically related words and instructed them to recall either immediately, 1 week, or 3 weeks later. Across the retention intervals, target recall decreased while critical lure recall remained stable. Brainerd et al. (2003) also reported that target recall decreased over a 1-week delay. These findings differ from those of Thapar and McDermott (2001, Experiment 1) who administered a surprise recall task either immediately, 2 days, or 7 days later. But while the decrease from immediate test to day 2 was steeper for target recall than critical lure recall, both decreased. There was little evidence of stability or delayed inflation of critical lures from immediate to a 7-day delay.

Work by Seamon et al. (2002) allows for the trajectory of memory to be examined beyond 1 week. Participants recalled either at a 2-week delay or a 2-month delay. In the delay from 2-weeks to 2-months, critical lure recall decreased while target recall did not. This finding suggests that although false recall of critical lures may be more stable than recall of targets across delays up to 2 weeks, lengthier delays (over 2 weeks) will cause steeper decreases for the critical lures. Even with the steeper decrease for critical lures from 2 weeks to 2 months, critical lures remained higher than targets at the 2-week and 2-month timepoints.

Recall of associated theme items has also been extended to applied settings. For example, Sherman and Kennerley (2014) presented participants with songs from popular artists. Participants recalled the songs 5 min and 1 week after study. During Test 1, participants recalled more target songs than critical lures (biggest hit by the artist, not presented at study; 5%). Recall of target songs decreased while recall of critical lures increased across time. This research shows how false memories produced with the DRM paradigm generalize to real-world settings. Moreover, this work shows that false memories may persist to the same degree as, if not more than, true memories for at least a period of 1 week.

There is also evidence that across delay, false memory for critical lures will exceed true memory for targets. For example, McDermott (1996, described above) tested word recall after a 2-day delay. Initially, participants had either recalled immediately or after completing arithmetic problems for 30 s. After the 2-day delay, critical lure recall exceeded target recall, even though both decreased from the short to long delay. The propensity for critical lure recall to be higher than target recall was also observed in research using a 1- and 2-day delay. In this study, individuals falsely recalled proportionally more critical lures than targets at both time periods (Pardilla-Delgado and Payne, 2017). False recall of critical lures has been shown to exceed target recall across 1-week delays (Thapar and McDermott, 2001; Sherman and Kennerley, 2014) and even across a 2-month delay (Seamon et al., 2002).

Overall, these results show a decline of target recall and stability or delayed inflation of critical lure recall, even after a 3-week delay (Toglia et al., 1999; Brainerd et al., 2003; Sherman and Kennerley, 2014). This stability of critical lure recall also occurs with confidence ratings; target recall confidence declines across delay, but critical lure recall confidence is unaffected by delay (Toglia et al., 1999). While much of the research with recall shows a steeper decrease for targets than critical lures across delay, the findings are mixed with respect to the stability or inflation of critical lures across delay (Thapar and McDermott, 2001, Experiment 1). Importantly there is also evidence that despite target recall being higher than critical lure recall when tested immediately after study, critical lure recall is higher than target recall across delays from 1 day up to 2 months. Thapar and McDermott (2001, Experiment 1) observed that target recall decreased more rapidly than critical lure recall over a 2-day delay. However, there was little evidence of stability or delayed inflation of critical lures as Fuzzy Trace Theory would predict (Thapar and McDermott, 2001, Experiment 1). Fuzzy Trace Theory predicts that critical lures would remain stable or increase across delays because they rely on gist. Failure to observe persistence of critical lures in Thapar and McDermott could be due to how they manipulated delay, or how they presented the stimuli.

In sum, participants tend to show decreased recall of targets and relatively stable or inflated memory for critical lures with delays up to 3 weeks. There is also evidence that despite target recall being higher than critical lure recall when tested immediately after study, critical lure recall is higher than target recall across delays from 1 day to 2 months. These findings are consistent with Fuzzy Trace Theory which predicts that the memory for gist will be stronger across a delay as the verbatim trace becomes less accessible. The results could also be consistent with Associative-Activation Theory. In instances where there is a strong association between the targets and the critical lures, repeated activation of a critical lure through presentation of multiple targets could result in a stronger memory representation for the critical lure than an individual target, across a delay.

Long-Term Recognition

Deese/Roediger–McDermott recognition findings are less consistent than DRM recall findings. Some studies have shown

that target recognition decreases more rapidly than critical lure recognition across long delays (Brainerd et al., 1995, Experiment 2; Payne et al., 1996; Thapar and McDermott, 2001; Parker et al., 2019; Houben et al., 2020, Experiment 1). Other studies have not shown this effect (Lampinen and Schwartz, 2000; Neuschatz et al., 2001; Wang et al., 2017).

For example, Houben et al. (2020) compared two time points (immediate versus 48-h delay) in two experiments. Participants learned five neutral and five negative 10-word DRM lists. Target recognition was higher at Time 1 than Time 2 but there was no effect of delay on critical lure recognition. The decrease in target recognition across a 2-day delay is consistent with Ebbinghaus's (1913) forgetting curve (see also Lampinen et al., 2005); however, the stable critical lure recognition over delay in Houben et al. is inconsistent with Lampinen et al. (2005) who observed an increase in critical lure recognition across delay. Another study by Lampinen and Schwartz (2000) showed that target recognition and critical lure recognition declined at a similar rate across a 48-h delay. In this study, participants listened to six lists before completing a 48-word recognition test immediately or 2 days afterward. Results showed a decrease in target recognition and corrected critical lure recognition across delay. Across two experiments, target recognition declined more than critical lure recognition for only non-corrected critical lures in Experiment 1. This pattern did not hold for corrected critical lure recognition in Experiment 1 or corrected or uncorrected critical lure recognition in Experiment 2 (Lampinen and Schwartz, 2000). Corrected critical lure scores address response bias by considering responding for non-related and related non-presented words. Non-corrected critical lure recognition does not account for response bias; it simply examines overall critical lure recognition. Bias-corrected scores are commonly reported when participants complete recognition tests (e.g., Brainerd and Reyna, 2018).

Whether recall is present during study likely influences the stability of false recognition across time. Importantly this could also influence the pattern of target recognition. In work by Payne et al. (1996), participants studied 16 lists. Immediately following each list, participants either recalled all the words or they completed arithmetic problems. After presentation of all the lists, half the participants completed a 384-item recognition test while the other participants returned 24-h later to complete the recognition test. Although target recognition decreased across delay, critical lure recognition remained stable. Participants were also more likely to report that a word was old (present at study) if the word belonged to a list which involved immediate recall rather than arithmetic at study. This was true for both types of recognition, but the effect was larger for target than critical lure recognition.

There is evidence that individuals will reject a critical lure in a recognition task if they can remember a specific target (tired) on the list contrary to the lure (sleepy) – *recollection rejection*. The idea of recollection rejection is consistent with Fuzzy Trace Theory, where participants reject a critical lure because they have access to a verbatim trace. In a study conducted by Lampinen et al. (2005), participants performed a think-aloud task during the study and test phase to examine what strategies

participants used to reject false memories. Recollection rejection occurred when participants rejected a word (e.g., sleep) because they remembered a different word being present (e.g., tired). Alternatively, *distinctiveness* occurred when the presence of that word should have evoked a specific memory (e.g., I would have remembered the word, needle, because I hate needles; Lampinen et al., 2005). The recognition test occurred either immediately or 48 h after the study phase and included targets, critical lures, non-presented associates that had a weaker association strength than the critical lures, and distractors from the non-studied lists. For all distractors and non-presented associates, distinctiveness was the most common strategy; however, for critical lures, recollection rejection was the most common strategy. While the use of both strategies decreased across delay, recollection rejection remained the most common strategy for critical lures.

A less frequent approach to examining the trajectory of true and false memories is to use a lexical decision task. Studies using reaction time on lexical decision tasks with DRM word lists yield mixed results. McKone (2004) had participants complete an intervening lexical decision task between a study phase and a recognition phase that occurred 3 or 10 min afterward. The percent of targets and critical lures was similar and there were no differences in delay. On the lexical decision task, targets were identified more quickly than targets from unstudied lists, but the reaction time for critical lures and lures from non-presented lists was the same. This pattern occurred across delay. McKone argued that the lexical decision task could distinguish between targets and critical lures across delay. These findings are inconsistent with Sergi et al. (2014), where participants completed a lexical decision task for targets, non-word targets, critical lures, new non-words, and unrelated new words. At an immediate test, a 3-min delay, and a 10-min delay, reaction time for targets and critical lures was shorter than the other categories and there were no differences in reaction time between targets and critical lures. Sergi and colleagues argued that the activation levels of targets and critical lures were similar, and the observed increase in reaction time for the lexical decision task was the same for both after a 10-min delay. The authors argued for an activation theory of false memory formation in the DRM. This argument is supported by the observation of latency scores on the lexical decision task being equivalent for targets and critical lures.

There is some evidence that across a delay, critical lure recognition will be equivalent (Thapar and McDermott, 2001) or higher than target recognition (Brainerd et al., 1995; Huff et al., 2012; Pardilla-Delgado and Payne, 2017). For example, Pardilla-Delgado and Payne (2017) had participants complete the recognition test either 24 h or 48 h after study. There was no effect of delay on target or critical lure recognition, but target recognition was lower than critical lure recognition at both testing points.

In sum, the evidence is mixed for DRM recognition after long delays. Some studies show target recognition decreases more rapidly than critical lure recognition across delay (Brainerd et al., 1995, Experiment 2; Payne et al., 1996; Thapar and McDermott, 2001; Houben et al., 2020). Others have not shown this effect (Lampinen and Schwartz, 2000; Brainerd et al., 2001; Neuschatz et al., 2001; Wang et al., 2017). There is evidence

that individuals use recollection rejection when judging critical lures. This strategy is consistent with the Fuzzy Trace Theory prediction that access to the verbatim trace may be used to reject gist-based critical lures. Sergi et al. (2014) used a lexical decision task to argue that targets and critical lures were activated and behaved similarly across delay, a finding more consistent with the Association-Activation Theory than Fuzzy Trace Theory. Finally, there is also some evidence that across delay, critical lure recognition exceeds target recognition. The inconsistency in findings derived from studies using recognition is likely due to the same methodology variations seen in free recall (e.g., between-subjects designs, number of lists, number of items, whether the items were recalled directly after each list).

Remember/Know Judgments

Theoretical explanations for false memory in the DRM paradigm suggest that the subjective experience of remembering should differ for targets and critical lures (Neuschatz et al., 2001). Based on Fuzzy Trace Theory, perceptual information from the verbatim trace could be integrated with the gist information, especially if the verbatim trace is no longer fully accessible. This could result in the subjective experience of the gist representation (false memory) being rich and resembling that of the verbatim trace. This would be expected to increase with delay due to verbatim decay (see Neuschatz et al., 2001). One way to study the subjective experience of memory is to have participants provide a remember or know judgment for each recognized word (Tulving, 1985; Gardiner and Java, 1990). For example, in Payne et al. (1996) participants provided more remember judgments at Time 1 than Time 2. This decrease in remember judgments occurred for lists followed by arithmetic but not for lists followed by immediate recall.

Neuschatz et al. (2001) found that after 48 h, participants could distinguish true from false items on the DRM using a memory characteristics questionnaire (Johnson et al., 1988). The memory characteristics questionnaire asks participants to rate the experience of remembering an item (e.g., what it sounded like, placement within the list, and what types of reactions the person had when the word was presented). At immediate test, participants provided more remember judgments for studied words than for critical lures, but this difference did not persist after the 48-h delay. However, participants' responses on the memory characteristics questionnaire differed for studied words and critical lures. That is, participants reported remembering more perceptual information for studied words than critical lures. This suggests that participants can distinguish between targets and critical lures based on some perceptual details for at least 2 days following the study phase. Neuschatz et al. (2001) argued that this is inconsistent with the notion that perceptual details for true memories fade more quickly than those for false or suggested memories, as proposed by Belli and Loftus (1994). It was noted that the findings may not generalize to situations where false memory persists beyond 2 days. It may be that for longer delays, perceptual information for false memory is less vulnerable to decay than that of true memory. For example, to examine false memory in an applied setting, participants watched simulated television programming with advertisements of five associated,

but interspersed products (e.g., beers, cars, and banks; Sherman et al., 2015). The researchers observed delayed inflation of critical lures after a 1-week delay. Additionally, remember judgments for target brands and filler items remained stable across time, while remember judgments for critical lures increased across time.

In sum, Fuzzy Trace Theory would predict that the subjective experience of associating rich, perceptual detail with target memory should decrease across delay. As the verbatim trace decays, the subjective experience associated with critical lures should resemble that of targets. The research to date has yet to show this pattern definitively.

Factors That Influence Persistence of Targets and Critical Lures

We are interested in the persistence of false memory in the DRM paradigm. Up to this point, our review focused on examining the trajectory of targets and critical lures with common measures of memory (recall, recognition, remember/know, and lexical decision tasks). The following sections explore other facets of the DRM that can interact with the effect of delay on targets and critical lures. We feel these facets contribute to the understanding of the trajectories as well as the theoretical explanations of how targets and critical lures persist across delay. For this reason, we included a small overview of: The effects of warning, valence, development, and connectivity on the trajectory of targets and critical lures in the DRM.

The Effect of Warning on Trajectory of Memories

Some researchers have examined the effect of warning participants about false memories in the DRM paradigm. Generally, individuals learn that the study lists contain items associated with one another. They also learn that during the memory test they will encounter words that are associated with the original study lists but were not presented during study. Participants are asked to avoid recalling or recognizing these words. In many studies examining the effect of warning, participants receive a single study session with or without warning and a single testing session (e.g., McDermott and Roediger, 1998; Gallo et al., 2001). We found only one study that included a warning between Test 1 and Test 2, following a single study session. Miller et al. (2011) had two detailed warning conditions and a no-warning control condition. In the critical lure warning condition, participants heard an explanation of critical lures and learned techniques to prevent false memories. In the criterion warning condition, participants were warned about saying old to any related words. The critical lure warning had no effect on targets or critical lure recognition; however, the criterion warning condition decreased both target and critical lure recognition from Test 1 to Test 2.

In a study by Wang et al. (2017), individuals were told that they falsely recognized a word on Test 1 before they completed Test 2. Participants then completed the compound remote associate task (CRAT). The solutions to the CRAT corresponded to the targets and critical lures from the DRM lists. While being challenged

affected CRAT solutions (those who were challenged produced fewer CRAT solutions than those who were not challenged), this was for true and false memories and did not vary across time.

In another experiment, Brainerd and Reyna (2018) had participants learn, with examples, that the test would include old words, new words that were semantically related to the old words (new similar), and new words that were semantically unrelated to the old words. These examples would essentially serve the same purpose as a warning, because participants were informed that words similar to old words would be on the test, but they were in fact new words. For lists with low association strength, participants were slightly more likely to judge new similar words as new than old but their ability to distinguish between new similar and old dissipated across a 10-day delay.

The limited research on the effect of warning or feedback between testing sessions suggests that general warnings may reduce false memories, but this may be due to a criterion shift. Warnings specifically targeting false memory may be less effective than those targeting how one responds more generally to associated items. The latter will reduce reporting of targets and critical lures to an equal extent. The decreased ability for the warning to help distinguish between old and new across a 10-day delay is consistent with Fuzzy Trace Theory, because it presumably results from decay of the verbatim trace.

Valence

The emotionality of the lists may influence the trajectory of targets and critical lures in a DRM paradigm. In a study conducted by Howe et al. (2010), participants studied six neutral lists and six negative emotional lists (Experiment 3) and completed a recognition test either immediately or after a 1-week delay. At initial test, participants recognized negative critical lures more often than neutral critical lures. Across delay, target recognition declined more for negative stimuli than for neutral stimuli; however, critical lure recognition remained stable for neutral stimuli and increased for negative stimuli. These results show delayed inflation of critical lures for negative emotional stimuli.

Similarly, in a study by Knott and Shah (2019), participants showed delayed inflation of critical lures for negative stimuli compared to neutral stimuli, when presented quickly. For critical lure recognition that had been presented slowly, participants were more likely to say old to negative words than neutral words. The results from Howe et al. (2010) and Knott and Shah differ from those of Choi et al. (2013) who found that target recognition was higher for negative stimuli than neutral stimuli and found no effect of valence on critical lure recognition. The effect of valence on target recognition was present after a 24-h delay, but critical lure recognition was more frequent for neutral than emotional stimuli (Experiment 2).

Individual differences might moderate the delayed inflation of critical lures for negative stimuli. Norris et al. (2019) found increased memory for negative lists at immediate test for target and critical lure recognition. However, after a 24-h delay, the researchers observed that those low in neuroticism no longer showed increased critical lure recognition for negative stimuli.

Valence of the stimuli may also interact with mood, and there could also be a mood congruency effect (Knott and Thorley, 2014; Packard et al., 2014). Knott and Thorley (2014) observed that after a delay, critical lure remember judgments were higher for negative stimuli than neutral stimuli, but only among participants who had watched a video aimed at eliciting a negative mood state.

Conclusions from Choi et al. (2013) and Norris et al. (2019) must be drawn cautiously because delay was compared across experiments. However, it appears that stimulus valence could influence the trajectory of critical lure memory in the DRM paradigm, with some studies showing inflated critical lures for negative words (Howe et al., 2010; Knott and Shah, 2019; Norris et al., 2019). This effect is likely moderated by several factors, including individual differences and mood congruency.

Developmental Trajectories of the Deese/Roediger–McDermott Paradigm

Many studies have observed a developmental reversal of the DRM effect, where false memories are higher for young adults compared to children (e.g., Brainerd et al., 1995, 2004; Dewhurst and Robinson, 2004; Howe et al., 2004; Howe, 2005; Lampinen et al., 2006; Dewhurst et al., 2007; Anastasi and Rhodes, 2008; Calado et al., 2019). Fuzzy Trace Theory predicts that false memories will increase with age, because children have yet to develop the same extensive level of semantic networks that adults possess (Brainerd et al., 2006).

Unlike adults, children who do not spontaneously generate gist memories should show lower levels of false memory persistence (Brainerd et al., 2006). Additionally, children may not exhibit the delayed inflation effect, because it requires processing gist memories of list themes during recall. Brainerd et al. (2006) examined the effects of immediate and delayed testing across 6- and 11-year-olds. In session one, participants studied the first eight DRM lists, followed by either 2 min of free recall or a distractor task. Participants proceeded with immediate testing, where they received a recognition test consisting of the previously studied lists and eight additional ones. After a 2- to 3-day delay, participants completed a 128-item recognition test, including half the items presented from session one. Older children showed the delayed inflation effect, while younger children did not. While false memory declined over a 2- to 3-day delay, false memory was higher for older children than younger children, regardless of immediate or delayed testing. These findings are consistent with Brainerd et al. (1995), where 5- and 8-year-olds completed recognition tests in an immediate and 1-week delayed testing session. False memories were higher in older children compared to younger children, although some false alarms and hits were persistent across this delay.

Theoretical explanations for these findings on false memory development include Fuzzy Trace Theory (Reyna and Brainerd, 1995; Brainerd and Reyna, 1998). Given that gist-based memories have been shown to be more stable than verbatim memories, true memory should decrease following a delay, while false memory should remain constant (Kintsch et al., 1990; Reyna and Kiernan, 1994, 1995). This process is known as delayed stability and has been demonstrated in children and adults (e.g., Payne et al.,

1996; Brainerd et al., 2001, 2006; Howe et al., 2010). Moreover, studies have shown an increase in levels of false memory for critical distractors (i.e., delayed inflation) in adults and children on delayed memory tests (e.g., Brainerd and Reyna, 1996; Payne et al., 1996).

Findings which show delayed inflation of critical lures in older but not younger children are consistent with Fuzzy Trace Theory. Younger children may differ from older children and adults in terms of gist processing. Therefore, it might be expected for critical lures to persist less so in younger children (Brainerd et al., 2006). However, the finding could also be consistent with Associative-Activation Theory. If younger children have less developed associative networks, this would also explain developmental differences with respect to persistence of critical lures across a delay. Additionally, some research shows that when given developmentally appropriate lists, younger children's memory of critical lures resembles that of older children and adults (Metzger et al., 2008).

List Connectivity

Norming studies show a wide range in rates of false memories produced by the different DRM lists. Stadler et al. (1999) observed that critical lure recognition varied from 27 to 84% on 36 lists and Roediger et al. (2001b) found that critical lure recognition rates varied from 11 to 84% on 55 lists. Thus, characteristics of the list items, including the number of items in a list, their connectivity to one another, and connectivity to the critical lure, may also influence the trajectory of true and false memories over delays. Connectivity levels refer to the mean connections per associate to the critical lure. Research shows that the connectivity levels of the lists may differentially affect the trajectory of target and critical lure recall across a 1-week delay (Goh and Khoo, 2007). In this study, participants viewed 24 lists, of which half were high connectivity (mean connections > 2), and the others were low connectivity (mean connections < 1). During Test 1 (immediately following study), connectivity facilitated true recall: Memory for targets was higher for high connectivity lists. However, connectivity did not affect critical lure memory. During Test 2 (1 week following), connectivity no longer influenced memory for targets; rather, it influenced memory for critical lures: False memory of critical lures was greater for lists with low connectivity than lists with high connectivity (Goh and Khoo, 2007). This is the only study that we found that used delay and connectivity strength to directly test Fuzzy Trace Theory and an alternative theory (PIER 2). Goh and Khoo (2007) argued that the findings are inconsistent with Fuzzy Trace Theory because high, not low connectivity, should result in greater gist extraction. A stronger gist should lead to inflation of critical lures. The researchers observed the inverse of this, with greater critical lure memory for low connectivity lists, after the delay.

McEvoy et al. (1999) showed that connectivity to the critical lure increased the likelihood that the critical lure will be falsely recalled. However, high connectivity between the presented words within a list, increased target recall. Likewise, critical recall decreased in lists with high connectivity between list words; the effect was consistent across a 1-min and 5-min delay. The

explanation for this finding is that access to the true presented words competes with the false recall of the critical lure and serves a protective function. Consistent with this idea, several studies have shown that *recall* rates for the target and the critical lures are inversely related (Stadler et al., 1999; Roediger et al., 2001b). However, other studies have reported positive relationships between targets and critical lures (Brainerd et al., 2003; Cody et al., 2015). Importantly, for recognition, high list connectivity resulted in more true recognition rates (hits) but also more false recognition rates (McEvoy et al., 1999). This points to different underlying mechanisms for recognition and recall of false memory in the DRM paradigm.

GENERAL DISCUSSION

We began this systematic review with one overarching question – when a person develops a false memory for an item in the DRM, how long will that false memory last? Answering that question led us to examine (1) the trajectory of memories for targets and critical lures in the DRM across delay; and (2) the proportion of true versus false memories across delay. The results of our search led to an inconsistent data pattern. However, some findings were consistent across studies. Firstly, false memories for critical lures develop rapidly. The small body of literature that examines at least two testing points within a single session shows that individuals will falsely recognize (Olszewska et al., 2015) within 3–4 s and falsely recall (McDermott, 1996) critical lures within 30 s of the study session. Secondly, target memory begins to decline rapidly. Most studies show steep declines up to about 2 days, entirely consistent with Ebbinghaus's (1913) forgetting curve. Thirdly, in delays up to 2 weeks, individuals commonly recall proportionately more critical lures than targets. Despite the inconsistency in the trajectory, there is remarkable consensus that across delays critical lures are falsely recalled at higher rates than recalled targets. This occurs with delays ranging from 1–2 days (McDermott, 1996; Thapar and McDermott, 2001; Pardilla-Delgado and Payne, 2017, Experiment 1) to 2 weeks (Toglia et al., 1999) to 2 months (Seamon et al., 2002). Finally, when recall tests are given, target memory generally declines more rapidly than false memory for critical lures.

These similarities in the data are met with an equal number of dissimilarities. There are several factors that help to explain the dissimilarities:

- (1) Whether delay was manipulated within or between subjects. For example, Toglia et al. (1999) observed that across three testing periods (immediate, 1 week, or 3 weeks) target recall decreased while critical lure recall remained stable. However, Thapar and McDermott (2001, Experiment 1), who tested participants either immediately, 2 days, or 7 days later, found that both target recall and false recall of critical lures decreased across delay. Failure to observe stability of critical lures in Thapar and McDermott could be due to delay being manipulated between subjects. Furthermore, some researchers compared across experiments to draw conclusions about delay (Choi et al.,

2013; Norris et al., 2019). Cross-experiment comparisons are not as methodologically sound as within-experiment comparisons; thus, we urge caution when interpreting results from cross-experiment comparisons.

- (2) The number of lists used. This will affect the study and test phase. More items create more interference. This may be particularly relevant when recognition is the dependent variable. For example, Payne et al. (1996) observed stability of critical lure recognition after a 24-h delay, while other researchers did not observe this stability (e.g., Lampinen and Schwartz, 2000). Payne et al. had 16 lists and a 384-item recognition test, while other studies used fewer lists.
- (3) The modality of the presentation of items at study and test. Lists presented visually may increase false recognition of critical lures in short-term memory, while lists presented auditorily may increase false recognition of critical lures in long-term memory (Olszewska et al., 2015).
- (4) Whether participants recalled items prior to delayed testing sessions. Some studies, even those which used remember/know judgments or recognition tests across two time points, had participants recall the words either directly after each list, or after all lists had been presented. The inclusion of a recall test may influence the trajectory of targets and critical lures. False recall of a specific critical lure on a test increases the likelihood that the word will be falsely recalled on a subsequent test. Increased likelihood to falsely recall critical lures at Test 2 when they were recalled at Test 1 could also be due to forgetting that comes from retrieval. Retrieving information increases subsequent memory for items that were retrieved, while impairing memory for semantically related items that were not retrieved (Anderson et al., 1994). This has been shown to occur with both targets and critical lures (Bäuml and Kuhbandner, 2003).

Is Persistence of False Memory for Critical Lures Due to a Criterion Shift?

A reviewer sagely asked whether persistence in false memory of critical lures across delay is a result of a criterion shift. That is, with a delay, individuals might be more likely to say they recognize words as having been presented at study. There are few studies which give warning or feedback between testing sessions. The available literature suggests that, compared to specific warnings about critical lures, general warnings (e.g., saying old to any related words) may be more effective at reducing false memories (Miller et al., 2011). General warnings, however, likely decrease reporting of targets and critical lures. This could suggest that individuals are developing a more liberal criterion across delay which might explain instances of delayed inflation of critical lures. If delayed inflation of critical lures results from a criterion shift, it could be argued that individuals would also be more likely to show increased recognition for targets and foils across a delay. Generally, target memory begins to decrease rapidly, even in studies where false memory for critical lures remains stable or increases across time. When examining the effect of delay on foils, there is an important consideration: While

most studies included foils, not all studies included analyses of the effect of delay on foils. Comparing patterns of critical lures to foils across delay would help DRM researchers determine whether persistence of critical lures is simply due to a criterion shift.

Studies that examined effect of delays on foils showed that individuals made more errors on related words (critical lures) than unrelated words (foils; Olszewska et al., 2015). Seamon et al. (2002) examined the trajectory of foils and reported a marginal effect ($p < 0.07$); more foils were reported at the 2-month test than the 2-week and immediate test. If this trend is to be interpreted, it could be that participants are adopting a more liberal criterion; however, it should be noted that critical lures decreased from the 2-week to 2-month period, and targets decreased from immediate to the 2-week period. If participants were more liberal in their responding, one might expect similar patterns for all item types. Seamon et al. (2002) attempted to equate baseline recall by analyzing adjusted recall proportions. To do this, they statistically adjusted both targets and critical lures to equal at baseline. They did this to address the concern that a linear scale of recall may not be appropriate for two functions that vary at baseline (e.g., Loftus, 1985; Thapar and McDermott, 2001). These adjusted values resulted in a similar pattern of results (at immediate test, target recall exceeded false recall of critical lures, but false recall of critical lures exceeded target recall at 2 weeks and 2 months). Additionally, adjusted scores showed a decrease in target recall and no change on critical lures from immediate to 2 weeks. From 2 weeks to 2 months there was a slight decrease in target recall and a steep decrease in critical lure recall. Using adjusted scores therefore yields results which are likely not due to a criterion shift.

In sum, because some studies do not analyze or report findings on the effect of delay on foils, it is difficult to rule out that delayed inflation of critical lures is not a result of a criterion shift. Warnings which instruct participants to avoid saying yes to semantically related words, have been shown to reduce both targets and critical lures. This supports a criterion shift argument. However, studies that do report the effect of delay on foils have shown critical lure memory is greater than foil memory. Additionally, it could be argued that a criterion shift would result in delayed inflation of targets as well. This is generally not observed in the literature. The existing research does not appear to support a criterion shift argument, but future research in this area should report the effect of delay on foils.

Why Are the Results More Consistent With Recall Than Recognition?

Inconsistencies most often seem to arise when dependent variables other than recall (recognition, R/K, lexical decision tasks) are used to examine memory across delay. Why is this so? As with recall, the trajectory of true memories when recognition is used to assess long-term memory is relatively consistent. If a test occurs immediately after study, target recognition is typically higher than false critical lure recognition. Target recognition then declines rapidly, within seconds. Most studies show steep declines in target recognition of up to 2 days. However, the

pattern is less consistent when examining the trajectory of long-term critical lure recognition: Some researchers report decreases, some report stability, and some report delayed inflation. As with studies using recall, studies with recognition have highly variable methodologies. Possibly, the task of recognizing targets and critical lures is more sensitive to these variations than is recall. Given that participants view possible targets and critical lures in a recognition task, the number of items at test is a factor that affects recognition but not recall. For example, after a 24-h delay, Payne et al. (1996) observed decreased target recognition and stable critical lure recognition. This study had a recall session preceding the first recognition test, which increased persistence of false memory for critical lures. Payne et al. also had 16 lists and a 384-item test, while other studies which failed to show stability of critical lure recognition had fewer lists and a shorter recognition test (e.g., Lampinen and Schwartz, 2000). Therefore, in addition to factors that influence recall (e.g., within or between subjects, modality of presentation, and whether one completes immediate recall following each list), recognition tasks might be especially sensitive to the number of items during study and test.

Does Fuzzy Trace Theory or Associative-Activation Theory Better Explain the Persistence of False Memory for Items in the Deese/Roediger-McDermott Paradigm?

Persistence of critical lures across delay has been argued as evidence for Fuzzy Trace Theory. According to Fuzzy Trace Theory, multiple testing sessions promote rehearsal of the gist without rehearsal of the verbatim trace. Increased study sessions allows one to rehearse verbatim traces, thereby countering the stability or inflation of false memories (Brainerd et al., 2003). Studies that examine delay within the same testing session show that false memories for critical lures develop rapidly. This could support gist or semantic encoding even in short-term memory. These findings are somewhat surprising given that at short delays, memory for the original lists should be strong and permit individuals to recognize memory errors of commission. According to Fuzzy Trace Theory, strong verbatim memory allows for rejection of critical lures in the DRM paradigm (Reyna et al., 2016, p. 7). With delays of 3–4 s, the verbatim trace should be available and target recognition should be high. This should decrease the reliance on gist-based processes. In some studies, participants view brief lists and then view a single recognition probe immediately and minutes later. McDermott (1996) and McEvoy et al. (1999; Experiment 3) had participants recall immediately after each list or after a 30–60 s delay. Studies which showed delayed inflation used recognition tests in which participants judged which words on the test appeared in prior lists. It could be that critical lure recognition is more persistent than critical lure recall in short-term than in long-term memory. Overall, the findings from studies examining short-term memory show different trajectories for targets and critical lures. Fuzzy Trace Theory argues that false memory for critical lures is due to reliance on the gist as the verbatim traces decay. After a delay of a few seconds, individuals should have access to the

verbatim traces and not have to rely on the gist. Critical lure recall and recognition in such brief periods could support the Associative-Activation Theory of false memory.

The trajectory of true memories when recall is used to assess long-term memory is relatively consistent. Most studies show steep declines up to about 2 days. However, the pattern is less consistent when examining the trajectory of false memory in long-term memory. Some studies show that critical lure recall decreases (McDermott, 1996; Thapar and McDermott, 2001, Experiment 1), others show critical lures remain stable (Toglia et al., 1999), and others show delayed inflation for critical lures (Brainerd et al., 2003; Sherman and Kennerley, 2014). Despite the inconsistency in the trajectory, there is remarkable consensus that across delays critical lures are falsely recalled at proportionately higher rates than recalled targets. This has been shown in delays from 1–2 days (McDermott, 1996; Thapar and McDermott, 2001; Pardilla-Delgado and Payne, 2017, Experiment 1) to 2 weeks (Toglia et al., 1999) to 2 months (Seamon et al., 2002).

The findings of rapid decline of target memory and higher false memory for critical lures compared to target memory after a delay is consistent with the predictions of Fuzzy Trace Theory; that is, the gist is more resistant to decay than is the verbatim trace. The findings could also be consistent with Associative-Activation Theory. This pattern might be especially expected when lists which have high backward association strength are used, which is common in many of the studies that examine delay. Backward association strength has been found to be one of the best predictors of false memory for critical lures (Roediger et al., 2001b). If one presents at study many lists containing targets that are highly associated with the critical lures, this would in theory result in repeated activation of specific critical lures. This pattern would also be expected during a recognition task. Presenting associated words during the retrieval phase would also result in strong activation of the critical lures, activation which might exceed that of the activation for any single target. Therefore, the critical lure could be more activated than specific targets, making the critical lures seem more familiar and more memorable after a delay. Given that the persistence of critical lures could be explained by both theories, further investigation is needed to tease apart whether the effect is due to gist extraction or association activation. Some work has been done to test this. For example, Otgaar et al. (2012) observed that disruptions in the association process through distraction resulted in fewer false memories for critical lures in children.

Findings from McKone (2004) which showed that participants could identify targets more quickly than critical lures on a lexical decision task seem to support Fuzzy Trace Theory which predicts that the verbatim trace and gist representation are similar but distinguishable. Conversely, Sergi et al. (2014) argued for an activation theory of false memory formation in the DRM. This argument is supported by the observation of latency scores on the lexical decision task being equivalent for targets and critical lures, even with a 10-min delay. The differences between McKone (2004) and Sergi et al. (2014) may be due to how the authors interpreted the data, and the inclusion of an immediate testing condition. Sergi and colleagues observed increased latency between the immediate and 10-min condition;

with no immediate condition, McKone could not observe this increase. Additionally, Sergi and colleagues' argument stems from the similarity of reaction time between targets and critical lures, which were both faster than non-words and new distractors. However, McKone compared reaction time from critical lures associated with presented lists to reaction time from critical lures associated with non-presented lists and found no advantage for critical lures associated with presented lists. So, while it appears that there may be an increase in latency from immediate test to a 10-min delay, targets and critical lures may be distinguished with a lexical decision task. Future research might consider including an immediate delay condition along with a comparison of critical lures from presented and non-presented lists to examine whether the ability to distinguish between targets and critical lures changes across delay.

There is also some evidence that the memory process is different for recognizing targets and recognizing critical lures. Brainerd and Reyna (2018) argued that when participants have the option to respond new-similar to critical lures, the findings indicate that new-similar and old words are remembered in different ways. When judging items as old or new semantically related words, participants were much more likely to correctly judge old words as old, but they judged new semantically related words as old and new semantically related at approximately equal rates. Permitting participants to respond new-similar in addition to old may provide interesting and insightful results if included in future DRM studies. There is also evidence that individuals use memory of targets to help reject critical lures, further suggesting that the two traces may be distinguishable. The idea that individuals will reject a critical lure because they can remember the target word/words is consistent with Fuzzy Trace Theory. If individuals can still access the verbatim then they will be less reliant on the gist.

Lampinen et al. (2005) examined the most common strategies for foils, related associates, and critical lures, even after a 48-h delay. They found that recollection rejection was the most common strategy used for critical lures. This would suggest that individuals still have access to the verbatim trace, for at least some of the targets. This is not necessarily contrary to Fuzzy Trace Theory, because Fuzzy Trace Theory does not stipulate complete degradation of verbatim traces across a specific time. An important finding may be that of different strategies used for related associates and critical lures. Both these items are semantically related to the targets. Therefore, it might be expected that access to the verbatim, recollection rejection, could also be used to reject the unpresented related associates. However, recollection rejection was most common for only critical lures, and distinctiveness was more common for other related associates. One difference between the two non-presented words is the strength of the association between each item and the targets; critical lures have a stronger association. It might therefore be argued that critical lures have a stronger activation than other non-presented associates, a stronger sense of familiarity, and therefore require a certain strategy (e.g., memory for the target) to counter this. Alternatively, other non-presented associates result in a

lower level of activation and familiarity, and, thus, can be rejected by strategies other than memory for the targets (e.g., distinctiveness). Therefore, the use of different rejection strategies for critical lures and other non-presented associates is potentially consistent with both Fuzzy Trace Theory and Associative-Activation Theory.

Looking at the subjective experience of remembering across time yields mixed findings. Some studies show that remember judgments for targets and critical lures decrease in a similar way (Lampinen and Schwartz, 2000); other studies show a different pattern for the two (Sherman et al., 2015). Fuzzy Trace Theory suggests that the item-specific information, the verbatim trace, decays rapidly after the study phase. Therefore, we would expect to see that the subjective experience of remembering the target words presented at study would initially differ from the gist-based representation of the critical lure; however, with a delay, the sense of remembering targets and critical lures would be more similar. This data pattern for remember/know judgments occurred in Neuschatz et al. (2001). That said, the subjective experiences for targets and critical lures were differentiated through the memory characteristics questionnaire after a 48-h delay, which is inconsistent with Fuzzy Trace Theory (Neuschatz et al., 2001).

In some studies, false memories are greater for older children than younger children regardless of immediate or delayed testing conditions (e.g., Brainerd et al., 1995, 2006). Additionally, older children will show delayed inflation of critical lures, while younger children do not show this effect. The developmental reversal of the DRM effect can be explained by children's less extensive levels of semantic networks and a lack of spontaneous, gist memory formation between list targets. Studies which fail to observe critical lure inflation in young children are consistent with Fuzzy Trace Theory. These results are arguably also consistent with the Associative-Activation Theory. Given that younger children have less developed semantic networks, it might also be expected that the presentation of targets results in less activation of critical lures in younger children compared to older children. In fact, dividing attention during the study phase may decrease the spread of activation. This has been shown to decrease false memories for children but not for adults. The explanation for this is that association activation is less automatic for children than adults, making the process more vulnerable to disruption in children (Otgaar et al., 2012, 2019).

Increased memory of critical lures compared to targets across a delay may be consistent with both Fuzzy Trace Theory and Associative-Activation Theory. A sophisticated understanding of how items are associated with one another and with the critical lure might provide more insight into the trajectory of targets and critical lures across various delays. Norming studies show that the rates of memory for critical lures vary across lists. Therefore, the characteristics of the lists themselves influence how they are remembered, the proportion of target and critical lure memories they produce, and how these memories persist across delay. There is good consensus that the strength of the associates to the critical lure increases false memories. However, when examining the association strength between target words rather than between target words and the absent

critical lure, lists that have high inter-item association strength produce lower critical lure memory, across delay (McEvoy et al., 1999; Goh and Khoo, 2007). Goh and Khoo (2007) argued that Fuzzy Trace Theory cannot fully explain this finding. Fuzzy Trace Theory predicts that lists with high connectivity, with strong semantic relatedness among the list items, should strengthen the gist representation, thereby increasing recall of the critical lures. Goh and Khoo (2007) argued that the finding of low connectivity having higher rates of critical lure memory is consistent with a specific associative-activation theory of memory (Nelson et al., 1992). Associative-Activation Theory argues that activation of a word increases the likelihood of the word being recalled. Additionally, recall of a list item can cue other list items because of their semantic relatedness. Therefore, high connectivity lists result in greater recall of targets (true memory). However, because high connectivity lists increase the likelihood that recall of one target on the list will cue another target, this decreases the likelihood that the recalled word will cue the critical lure. This is due to the increasing competition of the strong associates that were presented at study. Critical lures would be more likely to be cued through recall where competition from targets is lower, as is the case with low connectivity lists (Goh and Khoo, 2007). Although there were only two studies which examined connectivity strength and delay specifically, the findings are more consistent with Associative-Activation Theory than Fuzzy Trace Theory. Future research on list connectivity across longer delays could provide important insight about the trajectory of targets and critical lures.

In sum, results which show a rapid decrease in target memory and stability or delayed inflation for false memory for critical lures are consistent with Fuzzy Trace Theory. Fuzzy Trace Theory predicts that the verbatim trace, or memory for the presented targets at study, will fade rapidly; conversely, gist, or memory for critical lures, will persist across a delay. This has been shown with studies using recall and recognition. These results, however, may also be consistent with Associative-Activation Theory. This might be particularly true with lists that have a strong backward association strength or high connectivity between the critical lure and the targets. This could result in the critical lure being repeatedly activated to the point where the association is stronger than that of a target that appeared once during study. Studies which show that critical lures are more likely to be judged as remembered rather than known across a delay are consistent with Fuzzy Trace Theory but could also be consistent with Associative-Activation Theory for the same reason described above. Research which shows the rapid generation of critical lure memory may not be consistent with Fuzzy Trace Theory. Reporting of critical lures, or the gist, would not be expected within seconds because the verbatim traces should be available. Finally, research examining the connectivity between items provides insight into the persistence of target and critical lure memory. While strong connectivity between the critical lures and the targets promotes increased false memory for critical lures, strong inter-item connectivity results in decreased memory for critical lures. This has been argued to be inconsistent with Fuzzy Trace Theory; strong inter-item connectivity in theory

should promote gist formation, thereby increasing false memory for the critical lure.

Future Directions

The DRM effect is robust. It is likely because of this that many studies examining interactions do not report the effects on the foils relative to the critical lures. Reporting effects on foils across delay would strengthen the conclusions of the findings and potentially provide useful insight into the mechanisms underlying the DRM. Future research might meta-analyze the effect of delay on targets, critical lures, and foils in the DRM paradigm. Such a study would require authors of original research papers to re-analyze their foil data and then provide their results for the meta-analysis, which is a big ask. Future research could use artificial networks without pre-existing associations, or ones that are seemingly random, to test predictions made by Fuzzy Trace Theory and Associative-Activation Theory. If individuals can be taught, implicitly, to associate non-words to the point where false memories are formed, then this might support Associative-Activation Theory rather than Fuzzy Trace Theory. The key would be to eliminate pre-existing relatedness, and to ensure that individuals were not developing meaning for the associations. Research looking at the effects of distraction on false memory is fascinating, particularly the developmental differences that emerged; distraction was shown to reduce false memory in young children but not adults. Given that spread of activation also occurs at the retrieval phase, future research might consider using a divided-attention task during the retrieval phase of the DRM. Finally, research looking at the connectivity between all items in the paradigm will be useful. If the persistence of false memory after a delay is due to spread of activation, then even subtle activity which serves to activate the critical lure during retrieval (e.g., which item individuals recall first), needs to be examined.

CONCLUSION

We conducted our systemic review to answer a broad research question: What is the persistence of false memory for a specific item in the DRM paradigm? To help answer this question we had two research objectives: (1) to examine the trajectory of target memory and false memory for critical lures and (2) to examine whether memory for targets exceeded false memory for critical lures. With respect to (1), the bulk of studies show immediate decreases in target or true memories. The trajectory of critical lures or false memories yields mixed findings, likely due to methodological variation across studies. With respect to (2), the bulk of studies show that across delays of up to 2 months, critical

lure memories may be higher than target memories. Despite inconsistent findings, there is enough evidence to conclude that target (true) and critical lure (false) memories behave differently across delay. This pattern is consistent with the Fuzzy Trace Theory prediction that gist processes are more resistant to decay than verbatim memory traces. The pattern is also potentially consistent with Associative-Activation Theory, especially in instances of high association strength. Future research should continue to examine the effect of connectivity on the trajectory of targets and critical lures. A deeper understanding of how remembering specific items serves to cue or inhibit other items in the DRM paradigm will further our understanding of the persistence of false memories in the DRM.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PC developed the conceptualization and research question of the review. KD and IR conducted the article search. All authors summarized articles for the results table. PC and DB completed the initial writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Canada Research Chairs Program (950-232078).

ACKNOWLEDGMENTS

We would like to thank Kamaljit Bajwa for assistance with editing and formatting the Results table. **Table 1S** is submitted as **Supplementary Materials**.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.718818/full#supplementary-material>

REFERENCES

- Anastasi, J. S., and Rhodes, M. G. (2008). Examining differences in the levels of false memories in children and adults using child-normed lists. *Dev. Psychol.* 44, 889–894. doi: 10.1037/0012-1649.44.3.889
- Anderson, M. C., Bjork, R. A., and Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *J. Exper. Psychol. Learn. Mem. Cogn.* 20:1063. doi: 10.1037//0278-7393.20.5.1063
- Bäuml, K. H., and Kuhbandner, C. (2003). Retrieval-induced forgetting and part-list cuing in associatively structured lists. *Mem. Cogn.* 31, 1188–1197.
- Belli, R., and Loftus, E. F. (1994). "Recovered memories of childhood abuse: a source monitoring perspective," in *Dissociation: Clinical and Theoretical Perspectives*, eds S. J. Lynn and J. W. Rhue (New York, NY: Guilford Press), 415–433.

- Belli, R. F. (1989). Influences of misleading postevent information: Misinformation interference and acceptance. *J. Exper. Psychol. General* 118, 72–85. doi: 10.1037/0096-3445.118.1.72
- Bernstein, D. M., and Loftus, E. F. (2009). The consequences of false memory for food preferences and choices. *Perspect. Psychol. Sci.* 4, 135–139.
- Bernstein, D. M., Scoboria, A., Desjarlais, L., and Soucie, K. (2018). “False memory” is a linguistic convenience. *Psychol. Conscious. Theory Res. Prac.* 5:161. doi: 10.1037/cns0000148
- Brainerd, C. J., Forrest, T. J., Karibian, D., and Reyna, V. F. (2006). Development of the false-memory illusion. *Dev. Psychol.* 42, 962–979. doi: 10.1037/0012-1649.42.5.962
- Brainerd, C. J., Holliday, R. E., and Reyna, V. F. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child Dev.* 75, 505–522. doi: 10.1111/j.1467-8624.2004.00689.x
- Brainerd, C. J., Payne, D. G., Wright, R., and Reyna, V. F. (2003). Phantom recall. *J. Mem. Lang.* 48, 445–467. doi: 10.1016/S0749-596X(02)00501-6
- Brainerd, C. J., and Reyna, V. F. (1998). Fuzzy-trace theory and children’s false memories. *J. Exp. Child Psychol.* 71, 81–129. doi: 10.1006/jecp.1998.2464
- Brainerd, C. J., and Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Dev. Psychol.* 32, 467–478. doi: 10.1037/0012-1649.32.3.467
- Brainerd, C. J., and Reyna, V. F. (2018). Complementarity in false memory illusions. *J. Exper. Psychol. General* 147:305.
- Brainerd, C. J., Reyna, V. F., and Brandse, E. (1995). Are children’s false memories more persistent than their true memories? *Psychol. Sci.* 6, 359–364. doi: 10.1111/j.1467-9280.1995.tb00526.x
- Brainerd, C. J., Reyna, V. F., and Mojardin, A. H. (1999). Conjoint recognition. *Psychol. Rev.* 106, 160–179. doi: 10.1037/0033-295X.106.1.160
- Brainerd, C. J., Wright, R., Reyna, V. F., and Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *J. Exper. Psychol. Learn. Mem. Cogn.* 27:307.
- Calado, B., Otgaar, H., and Peter, M. (2019). Are children better witnesses than adolescents? Developmental trends in different false memory paradigms. *J. Child Custody* 15, 330–348. doi: 10.1080/15379418.2019.1568948
- Choi, H. Y., Kensinger, E. A., and Rajaram, S. (2013). Emotional content enhances true but not false memory for categorized stimuli. *Mem. Cogn.* 41, 403–415. doi: 10.3758/s13421-012-0269-2
- Cody, M. W., Steinman, S. A., and Teachman, B. A. (2015). True and false memories in social anxiety disorder: effects of speech anticipation and social content. *Cogn. Ther. Res.* 39, 797–807. doi: 10.1007/s10608-015-9712-6
- Colins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic memory. *Psychol. Rev.* 82, 407–428.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exper. Psychol.* 58, 17–22.
- Dewhurst, S. A., Pursglove, R. C., and Lewis, C. (2007). Story contexts increase susceptibility to the DRM illusion in 5-year-olds. *Dev. Sci.* 10, 374–378. doi: 10.1111/j.1467-7687.2007.00592.x
- Dewhurst, S. A., and Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychol. Sci.* 15, 782–786. doi: 10.1111/j.0956-7976.2004.00756.x
- Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*, trans. H. A. Ruger and C. E. Bussenius. New York, NY: Teachers College, Columbia University. (Original work published 1885)
- Festini, S. B., and Reuter-Lorenz, P. A. (2013). The short-and long-term consequences of directed forgetting in a working memory task. *Memory* 21, 763–777. doi: 10.1080/09658211.2012.754900
- Flegal, K. E., Atkins, A. S., and Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *J. Exper. Psychol. Learn. Mem. Cogn.* 36, 1331–1338.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Mem. Cogn.* 38, 833–848. doi: 10.3758/MC.38.7.833
- Gallo, D. A., Roediger, H. L., and McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychon. Bull. Rev.* 8, 579–586. doi: 10.3758/BF03196194
- Gardiner, J. M., and Java, R. I. (1990). Recollective experience in word and nonword recognition. *Mem. Cogn.* 18, 23–30.
- Goh, W. D., and Khoo, B. H. (2007). Effects of inter-associate connectivity on the persistence of false recall. *Proc. Annu. Meet. Cogn. Sci. Soc.* 29:1069.
- Houben, S. T. L., Otgaar, H., Roelofs, J., Smeets, T., and Merckelbach, H. (2020). Increases of correct memories and spontaneous false memories due to eye movements when memories are retrieved after a time delay. *Behav. Res. Ther.* 125, 103546–103546. doi: 10.1016/j.brat.2019.103546
- Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychol. Sci.* 16, 927–931. doi: 10.1111/j.1467-9280.2005.01638.x
- Howe, M. L., Candel, I., Otgaar, H., Malone, C., and Wimmer, M. C. (2010). Valence and the development of immediate and long-term false memory illusions. *Memory* 18, 58–75. doi: 10.1080/09658210903476514
- Howe, M. L., Cicchetti, D., Toth, S. L., and Cerrito, B. M. (2004). True and false memories in maltreated children. *Child Dev.* 75, 1402–1417. doi: 10.1111/j.1467-8624.2004.00748.x
- Howe, M. L., Wimmer, M. C., Gagnon, N., and Plumpton, S. (2009). An associative-activation theory of children’s and adults’ memory illusions. *J. Mem. Lang.* 60, 229–251. doi: 10.1016/j.jml.2008.10.002
- Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., and Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: Effects of initial recall, recognition, and the ironic effect of guessing. *J. Exper. Psychol. Learn. Mem. Cogn.* 38:1720.
- Johnson, M. K., Foley, M. A., Suengas, A. G., and Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *J. Exper. Psychol. General* 117:371.
- Kintsch, W., Welsch, D., Schmalhofer, F., and Zimny, S. (1990). Sentence memory: A theoretical analysis. *J. Mem. Lang.* 29, 133–159.
- Knott, L. M., and Shah, D. (2019). The effect of limited attention and delay on negative arousing false memories. *Cogn. Emot.* 33, 1472–1480. doi: 10.1080/02699931.2018.1556153
- Knott, L. M., and Thorley, C. (2014). Mood-congruent false memories persist over time. *Cogn. Emot.* 28, 903–912. doi: 10.1080/02699931.2013.860016
- Lampinen, J. M., Leding, J. K., Reed, K. B., and Odegard, T. N. (2006). Global gist extraction in children and adults. *Memory* 14, 952–964. doi: 10.1080/09658210601008957
- Lampinen, J. M., Meier, C. R., Arnal, J. D., and Leding, J. K. (2005). Compelling untruths: content borrowing and vivid false memories. *J. Exper. Psychol. Learn. Mem. Cogn.* 31:954.
- Lampinen, J. M., and Schwartz, R. M. (2000). The imperistence of false memory persistence. *Memory* 8, 393–400.
- Loftus, E. F., Miller, D. G., and Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *J. Exper. Psychol. Hum. Learn. Mem.* 4, 19–31.
- Loftus, G. R. (1985). Evaluating forgetting curves. *J. Exper. Psychol. Learn. Mem. Cogn.* 11, 397–406.
- McDermott, K. B. (1996). The persistence of false memories in list recall. *J. Mem. Lang.* 35, 212–230. doi: 10.1006/jmla.1996.0012
- McDermott, K. B., and Roediger, H. L. III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *J. Mem. Lang.* 39, 508–520. doi: 10.1006/jmla.1998.2582
- McEvoy, C. L., Nelson, D. L., and Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *J. Exper. Psychol. Learn. Mem. Cogn.* 25:1177. doi: 10.1037/0278-7393.25.5.1177
- McKone, E. (2004). Distinguishing true from false memories via lexical decision as a perceptual implicit test. *Aus. J. Psychol.* 56, 42–49. doi: 10.1080/00049530410001688092
- Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W., and Williams, D. (2008). Do children “DRM” like adults? False memory production in children. *Dev. Psychol.* 44, 169–181. doi: 10.1037/0012-1649.44.1.169
- Miller, M. B., Guerin, S. A., and Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *J. Exper. Psychol. Learn. Mem. Cogn.* 37, 1228–1235. doi: 10.1037/a0024539

- Nelson, D. L., Schreiber, T. A., and McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychol. Rev.* 99, 322–348. doi: 10.1037/0033-295X.99.2.322
- Neuschatz, J. S., Payne, D. G., Lampinen, J. M., and Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory* 9, 53–71.
- Nichols, R. N., and Loftus, E. F. (2019). Who is susceptible in three false memory tasks? *Memory* 27, 962–984. doi: 10.1080/09658211.2019.1611862
- Norris, C. J., Leaf, P. T., and Fenn, K. M. (2019). Negativity bias in false memory: moderation by neuroticism after a delay. *Cogn. Emot.* 33, 737–753. doi: 10.1080/02699931.2018.1496068
- Olszewska, J. M., Reuter-Lorenz, P. A., Munier, E., and Bendler, S. A. (2015). Misremembering what you see or hear: Dissociable effects of modality on short- and long-term false recognition. *J. Exper. Psychol. Learn. Mem. Cogn.* 41, 1316–1325. doi: 10.1037/xlm0000115
- Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., and Salmon, K. (2013). False Memory $\neq$ False Memory: DRM errors are unrelated to the misinformation effect. *PLoS One* 8:e57939. doi: 10.1371/journal.pone.0057939
- Otgaar, H., Howe, M. L., Muris, P., and Merckelbach, H. (2019). Associative activation as a mechanism underlying false memory formation. *Clin. Psychol. Sci.* 7, 191–195. doi: 10.1177/2167702618807189
- Otgaar, H., Peters, M., and Howe, M. L. (2012). Dividing attention lowers children's but increases adults' false memories. *J. Exper. Psychol. Learn. Mem. Cogn.* 38, 204–210. doi: 10.1037/a0025160
- Packard, P. A., Rodríguez-Fornells, A., Stein, L. M., Nicolás, B., and Fuentemilla, L. (2014). Tracking explicit and implicit long-lasting traces of fearful memories in humans. *Neurobiol. Learn. Mem.* 116, 96–104. doi: 10.1016/j.nlm.2014.09.004
- Pardilla-Delgado, E., and Payne, J. D. (2017). The Deese-Roediger-McDermott (DRM) task: A simple cognitive paradigm to investigate false memories in the laboratory. *J. Visual. Exper.* 119:54793. doi: 10.3791/54793
- Parker, A., Dagnall, N., and Abelson, A. (2019). Effects of survival processing and retention interval on true and false recognition in the DRM and category repetition paradigms. *Memory* 27, 353–367. doi: 10.1080/09658211.2018.1511808
- Payne, D. G., Elie, C. J., Blackwell, J. M., and Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *J. Mem. Lang.* 35, 261–285.
- Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychon. Bull. Rev.* 3, 105–111. doi: 10.3758/BF03210749
- Reyna, V. F., and Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learn. Individ. Differ.* 7, 1–75.
- Reyna, V. F., Corbin, J. C., Weldon, R. B., and Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *J. Appl. Res. Mem. Cogn.* 5, 1–9. doi: 10.1016/j.jarmac.2015.12.003
- Reyna, V. F., and Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Dev. Psychol.* 30, 178–191.
- Reyna, V. F., and Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor Symb. Activity* 10, 309–331.
- Roediger, H. L. III, Balota, D. A., and Watson, J. M. (2001a). "Spreading activation and the arousal of false memories," in *The nature of remembering: Essays in honor of Robert G. Crowder*, eds H. L. Roediger III, J. S. Nairne, I. Neath, and A. M. Surprenant (Washington, DC: American Psychological Association), 95–115.
- Roediger, H. L. III, Watson, J. M., McDermott, K. B., and Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407.
- Roediger, H. L. III, and Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn. Sci.* 15, 20–27. doi: 10.1016/j.tics.2010.09.003
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *J. Exper. Psychol. Learn. Mem. Cogn.* 21, 803–814.
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., et al. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Mem. Cogn.* 30, 1054–1064.
- Sergi, I., Senese, V., Pisani, M., and Nigro, G. (2014). Assessing activation of true and false memory traces: A study using the DRM Paradigm. *Psychol. Belgica* 54, 171–179. doi: 10.5334/pb.ak
- Sherman, S. M., Follows, H., Mushore, A. B., Hampson-Jones, K., and Wright-Bevans, K. (2015). Television advertisements create false memories for competitor brands. *J. Appl. Res. Mem. Cogn.* 4, 1–7. doi: 10.1016/j.jarmac.2014.06.001
- Sherman, S. M., and Kennerley, J. (2014). The organisation of musical semantic memory: Evidence from false memories for familiar songs. *Memory* 22, 852–860. doi: 10.1080/09658211.2013.839709
- Stadler, M. A., Roediger, H. L., and McDermott, K. B. (1999). Norms for word lists that create false memories. *Mem. Cogn.* 27, 494–500.
- Thapar, A., and McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Mem. Cogn.* 29, 424–432.
- Toglia, M. P., Neuschatz, J. S., and Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory* 7, 233–256.
- Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12.
- Wang, J., Otgaar, H., Howe, M. L., Smeets, T., Merckelbach, H., and Nahouli, Z. (2017). Undermining belief in false memories leads to less efficient problem-solving behaviour. *Memory* 25, 910–921. doi: 10.1080/09658211.2016.1249888

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Coburn, Dogra, Rai and Bernstein. 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) and the copyright owner(s) 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.



Novel Approaches and Cognitive Neuroscience Perspectives on False Memory and Deception

Michael P. Toglia^{1*}, Joseph Schmuller², Britni G. Surprenant³, Katherine C. Hooper², Natasha N. DeMeo⁴ and Brett L. Wallace⁵

¹ Department of Psychology, Cornell University, Ithaca, NY, United States, ² Department of Psychology, University of North Florida, Jacksonville, FL, United States, ³ Department of Psychology, University of Georgia, Athens, GA, United States, ⁴ Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, United States, ⁵ School of Psychology, Florida Institute of Technology, Melbourne, FL, United States

OPEN ACCESS

Edited by:

Eddy J. Davelaar,
Birkbeck, University of London,
United Kingdom

Reviewed by:

Enmanuelle Pardilla-Delgado,
Brigham and Women's Hospital and
Harvard Medical School,
United States
Naveen Kashyap,
Indian Institute of Technology
Guwahati, India

*Correspondence:

Michael P. Toglia
m.toglia@unf.edu

Specialty section:

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

Received: 07 June 2021

Accepted: 23 February 2022

Published: 21 March 2022

Citation:

Toglia MP, Schmuller J,
Surprenant BG, Hooper KC,
DeMeo NN and Wallace BL (2022)
Novel Approaches and Cognitive
Neuroscience Perspectives on False
Memory and Deception.
Front. Psychol. 13:721961.
doi: 10.3389/fpsyg.2022.721961

The DRM (Deese–Roediger–McDermott) paradigm produces robust false memories of non-presented critical words. After studying a thematic word list (e.g., *bed*, *rest*, and *pillow*) participants falsely remember the critical item “sleep.” We report two false memory experiments. Study One introduces a novel use of the lexical decision task (LDT) to prime critical words. Participants see two letter-strings and make timed responses indicating whether they are both words. The word pairs Night-Bed and Dream-Thweeb both prime “sleep” but only one pair contains two words. Our primary purpose is to introduce this new methodology via two pilot experiments. The results, considered preliminary, are promising as they indicate that participants were as likely to recognize critical words (false memories) and presented words (true memories) just as when studying thematic lists. Study Two actually employs the standard DRM lists so that semantic priming is in play there as well. The second study, however, uses functional near-infrared spectroscopy (fNIRS) to measure activity in the prefrontal cortex during a DRM task which includes a deception phase where participants intentionally lie about critical lures. False and true memories occurred at high levels and activated many of the same brain regions but, compared to true memories, cortical activity was higher for false memories and lies. Accuracy findings are accompanied by confidence and reaction time results. Both investigations suggest that it is difficult to distinguish accurate from inaccurate memories. We explain results in terms of activation-monitoring theory and Fuzzy Trace Theory. We provide real world implications and suggest extending the present research to varying age groups and special populations. A nagging question has not been satisfactorily answered: Could neural pathways exist that signal the presence of false memories and lies? Answering this question will require imaging experiments that focus on regions of distinction such as the anterior prefrontal cortex.

Keywords: DRM paradigm, lexical decision task, fNIRS, deception, fuzzy trace theory, activation-monitoring theory, spreading activation, cognitive neuroscience

INTRODUCTION AND HISTORICAL BACKDROP

It is well known that accurate remembering may fall under the influence of errors of omission and/or of commission. Nevertheless, memory generally serves us quite well, even with gaps in recounting an event. Gap-related failures due to errors of omission are often attributed to forgetting, a natural feature of cognitive functioning.

Our focus, however, is on errors of commission that typify faulty remembrances. These flaws in remembering have come to be known as *false memories*. They occur when we remember events in a distorted manner or come to have memories for events that never happened. The term “false memory” appeared in the early 1990s in the eyewitness memory literature (Loftus, 1993) and with Peter and Pamela Freyd’s founding of the False Memory Syndrome Foundation in 1992. The term’s acceptance solidified shortly after the publication of the seminal article by Roediger and McDermott (1995). While “false memory” has become an umbrella encompassing many forms of deficits and alterations of human memory, the psychological literature addressing the frailties of memory stretches back well over a 100 years (Ebbinghaus, 1885; Munsterberg, 1908). Explanations for how and why false memories occur emphasize the role of reconstructive processes in retrieval (see Bartlett, 1932 for cogent, early arguments). Roediger and McDermott (1995) went a step further: They made the case that all remembering, accurate or not, depends on reconstruction.

Inaccurate memories share a common trait in that they contain at least slivers of truth. Even our strongest true memories, however, have false components – as is the case in flashbulb memories (Hornstein et al., 2003). This overlapping characteristic is consistent with the notion that illusory memories, though errant, maintain some aspects of real experiences, often sharing central features of meaning as in recalling “hurt on a school playground” rather than “hurt in a city park” (see Brainerd et al., 2008 for similar arguments).

Many factors contribute to the difficulty in distinguishing true recollections from false ones (Marche et al., 2010). Further blurring the distinction is that illusory memories are highly believed, often at levels comparable to true memories (Toglia et al., 1999; Shaw, 2020).

The concept of belief raises another form of factual inaccuracy – the deliberate inaccuracy when someone knowingly lies. False memories in most instances are not, and should not, be equated with lies (Ceci et al., 1987; Loftus, 2005). Yet some similarities are evident: Liars bolster believability by inserting some degree of truth into an overall false narrative or by embedding false information in a generally true narrative. In fact, these “half-truths”, if believed, might serve as a lynchpin for the creation of a false memory (Otgaar and Baker, 2018). Empirical and theoretical studies of accuracy, inaccuracy, and deception, across many subfields of psychology as well as forensic science and neuroscience, are important because false memories and lies are common in daily life. Regarding the latter, intent is involved in fibs, white lies, and whoppers. False memories are not considered intentional (but see our Study 2 reported later).

They take shape from misinformation or from misremembering. Specifically, some real-life illusory recollections form from a blend of unintentional or purposeful misinformation, and/or from schematic knowledge represented in conceptual networks in semantic memory. These inaccurate additions to long-term declarative memory are pervasive in real-world settings. Significantly, research shows false memories to be long lasting under experimental conditions (Toglia et al., 1999; Seamon et al., 2002; Brainerd and Reyna, 2005).

The Deese–Roediger–McDermott Illusion

The DRM paradigm (Deese, 1959; Roediger and McDermott, 1995) involves standard verbal learning of lists of semantically related words. This paradigm and variants have been the foundation of thousands of false memory experiments, including the studies we report in this article. Kirkpatrick (1894) conducted and reported a DRM-like experiment generally regarded as the first false memory study. Her participants studied several groups of verbal items, each containing thematic connections. Their recall protocols revealed intrusions of words associated with the study lists. In her discussion section she wrote:

There were some incidental cases of false recall. About a week before. . . I had said ten words to the students. Many of these were evoked and placed on the lists as if they were part of it. Again, it seems that when words such as “roll,” “thimble,” and “knife” were pronounced, many students thought of “thread,” “needle,” and “fork,” which are so often associated with them. The result was that many of these words were evoked as belonging to the list. This is an excellent illustration of how things suggested to a person during an experience can be honestly reported by that person as part of that experience. (pp. 608, 609).

Note that Kirkpatrick mentions “incidental cases of false recall” and that suggestions during an experience “can be honestly reported by that person as part of that experience.” These quotes comport with our earlier comments that false recollections are not viewed as intentional nor considered lies.

One hundred years after Kirkpatrick’s study, Roediger and McDermott (1995) revived Deese’s (1959) associative memory procedure used effectively, primarily in the laboratory, for more than 25 years, to investigate an array of false memory phenomena. In a typical DRM study participants study lists of words. Each list consists of semantically related items, some more than others, that converge upon a critical item (often referred to as a “critical lure”) that does not appear during study. For instance, consider *chair* as the critical missing item for this list: *table, sit, legs, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, and bench*. List presentation sometimes precedes a brief delay to buffer the effects of short-term memory. Then participants have to recall or recognize as many of the presented words as they can and in doing so, they often “remember” the critical item as part of the study list. False recall usually occurs 50% of the time, while false recognition of critical items occurs at rates that approach 80%. These false memory effects are sometimes referred to in the literature as the “DRM illusion.” Thus, illusory memories are very compelling, so much so that their rate of occurrence very often exceeds levels of accurate (true) memory

(Toglia et al., 1999; Brainerd and Reyna, 2005; Bui et al., 2013; Pimentel and Albuquerque, 2013).

Beyond their frequency, other findings speak to the phenomenological experience of false memories. For example, participants typically express very high confidence in memories that are inaccurate (Roediger and McDermott, 1995). In addition to measuring belief in their recall or recognition of words, in some experiments, participants categorize their memories with either a “remember” judgment (a detailed, vivid independent recollection), a “know” judgment (a general sense of familiarity without any independent recollection), or a “guess” judgment. Participants often claim they “remember” critical lures as frequently as the studied words. Thus, false memories are very robust. This alarming conclusion prompted researchers to examine ways to potentially diminish the creation and occurrence of false memories.

Experimental approaches to reduce the DRM illusion include warning participants about the thematic nature of the lists to be studied by providing them with sample lists and their critical lures. These explicit examples are given to make the point that the experimenter is trying to trick participants to remember non-presented words and to urge them not to fall into this trap. These “warning” studies report mixed results, yet generally show warnings often fail to reduce the DRM illusion (Neuschatz et al., 2001).

Why the Deese–Roediger–McDermott Paradigm

The DRM approach to the study of false memory has dominated this literature over the last two decades. An easily conducted experimental procedure, it yields impressive levels of illusory memory which is at the root of this dominance. Arguably the results have wide ranging applications beyond word lists (Brainerd and Reyna, 2002). For purposes of this article, these include implications for false memory in legal contexts. Such implications have met skepticism based on ecological validity. This is a fair point. We ask readers to refrain from judging the degree of external validity until we have advanced our concluding remarks.

For the moment, consider that forensic interviewers might introduce thematically related words and phrases weaved throughout the questioning. The related material could converge on topics that can lead an interviewee to provide false memories as answers. Similarly, police frequently use aggressive methods (the Reid Technique), combined with establishing an accusatory atmosphere when interrogating suspects who they presume guilty (Kassin and Kiechel, 1996; Kassin and Gudjonsson, 2004).

When the investigating detective’s goal is to force a confession, the detective might make repeated accusations of guilt, infuse misleading information and lies, and discuss criminal themes. In the most extreme form of forced confessions known as internalized false confessions (Kassin and Wrightsman, 1985; Kassin, 2008) the innocent suspect comes to believe in his or her guilt substantiated by new, confabulated, memories that are false.

While this chapter focuses on falsely recalled and/or recognized words, in terms of generalizability the literature is

replete with DRM-consistent memory results observed across a wide range of stimulus materials. These include memory for pictures (Israel and Schacter, 1997; Koustaal et al., 2003; Foley and Foy, 2008), memory for sentences (Bransford and Franks, 1971; Bransford et al., 1972; Matzen and Benjamin, 2013), memory for text (Bower et al., 1979; Reyna and Kiernan, 1994), remembering pragmatic inferences (Brewer, 1977; Barclay et al., 1984), and remembering scenes (Dechterenko et al., 2021). Consistent results also occur in naturalistic contexts, as in memory for the contents of a graduate student’s office (Brewer and Treyns, 1981) and memory for a professor’s lecture and behaviors during its delivery (Neuschatz et al., 2002). Accompanying these many forms of behavioral data establishing the pervasive occurrence of false memories, is an ever-increasing literature that addresses the neural correlates of false memory. Expansive meta-analyses of dozens of fMRI studies have implicated several regions of the frontal, temporal, and parietal cortices in the encoding and retrieval phases of false memory, with the superior prefrontal gyrus being most significant (Kurkela and Dennis, 2016; Yu et al., 2019).

These materials and contexts, as well as behavioral and neurophysiological investigations, clearly involve a variety of paradigms that all rely on semantic bases in fomenting false memories. The semantic/thematic nature of experimental materials, even though they may vary widely, permit theories to account for most effects across these stimulus materials.

Theoretical Explanations of False Memory Effects

Any adequate theory of memory must account for both true and false memories. We now outline two theories that satisfy this requirement.

We begin with Activation-Monitoring Theory (AMT), a dual-process theory. Here the processes are cognitive operations, activation and monitoring. Activation primarily happens at encoding. Monitoring is largely a retrieval process, though both can operate at encoding and retrieval (Roediger et al., 2001). The term “activation” is from spreading activation theory which represents concepts and their properties in associative networks (e.g., Collins and Loftus, 1975). Items in long-term memory are interconnected nodes throughout the network, an extensive web-like structure. Connections between two concepts vary in their “semantic distance,” determined by the strength of the connection. Strongly related terms (e.g., chair-table) are separated by short semantic distances, while weaker-bonded items (e.g., chair-bench) are linked at longer distances. We use here the examples from the DRM list mentioned earlier to facilitate the discussion of activation as it applies to an encoding phase in a false memory experiment. According to AMT, as the participant hears a list, he or she encodes each word into a network node and activation spreads to connected words; stronger connections enjoy greater activation. Many of the presented words receive substantial activation, repeatedly so, and this supports true memory. Critical lures, however, also receive repeated activation from the semantically-related, presented list items. At retrieval participants engage in the second activity of

the dual processes, monitoring, which involves editing of whether or not to commit to recalling or recognizing a word. Highly activated items, the presented words and critical lures, are likely to cross the decision threshold (i.e., the “criterion” in recognition memory models) for output at test.

Another approach to understanding the DRM illusion is Fuzzy Trace Theory (FTT) advanced by Reyna and Brainerd (1995). They proposed that stimuli are encoded in two independent ways. FTT is thus a dual-process theory. Unlike AMT, its processes target the development, independently, of two types of memory traces. One involves verbatim encoding, producing a trace of exact, detailed information as in the list words. The other involves coding for gist, creating a trace of general characteristics, like the list words relating to a theme (e.g., chair). Both verbatim and gist representations contribute to high levels of true memory. For illusory memory, the focus is on gist processing. Participants frequently falsely recall/recognize “chair” as a list item. They cannot possibly access a verbatim trace of “chair” because it was not presented. They can, however, rely on a strong gist trace because “chair” is the best descriptor of the theme of the studied list.

We now present two pilot studies, collectively referred to as Study 1. We begin with a lexical decision task investigation, involving a student project and a follow-up experiment, the second author designed with spreading activation/AMT in mind.

STUDY 1

Introduction

Introduced by Meyer and Schvaneveldt (1971), the lexical decision task (LDT) has seen wide usage in the study of semantic priming processes in neurotypical populations of all ages (see Plaut and Booth, 2000 for theoretical arguments), in individuals with linguistic deficits (e.g., aphasia, Miberg and Blumstein, 1981; dyslexia, Martens and de Jong, 2006), and with memory impairments (e.g., Alzheimer’s disease, Ober and Shenaut, 1988). In one version of Meyer and Schvaneveldt’s (1971) procedure, the participant sees two letter-strings on each trial, one above the other. The task is to press one button if both letter-strings are words, and another if at least one-letter string is not a word. The two strings remain visible until the response, and the dependent variable is the time between the beginning of the trial and the participant’s response. One typical result is that participants correctly respond faster when the two letter-strings are semantically related words than when they are semantically unrelated words. This is consistent with spreading activation theory (Collins and Loftus, 1975).

We used this two-word procedure to fully leverage the potent priming effects of DRM list words. The DRM procedure in this study appears in a novel way: Instead of using the LDT in its standard method as an assessment for false recognition, the LDT was the semantic priming stimulus for subsequent false recognition. The words in the LDT came from DRM lists. A recognition test followed each LDT block. The test contained the critical lures from the DRM word lists primed in the preceding LDT trials. For example, an LDT trial could present

two words from the same DRM list (e.g., “night” and “bed” from the *sleep* list), two words from different DRM lists (“night” and “butter” from the *bread* list), a word and a non-word, or two non-words. **Figure 1** shows examples.

Our long-term goal is to understand *behavioral* and *neurophysiological* underpinnings of false recognition by utilizing this novel *behavioral* approach as well as a *neurophysiological* approach that leverages functional near-infrared spectroscopy (fNIRS, described in Study 2) concurrent with the LDT trials and recognition trials.

The present report only addresses a short-term objective achieved by implementing the *behavioral* approach in a classroom experiment and in a small follow-up experiment. In doing so we accomplish the overarching objective of introducing this new and novel paradigm, consistent with the purpose of the volume in which the present article appears. We predicted that semantic priming stemming from LDT trials would successfully produce false memories on a subsequent old/new recognition test.

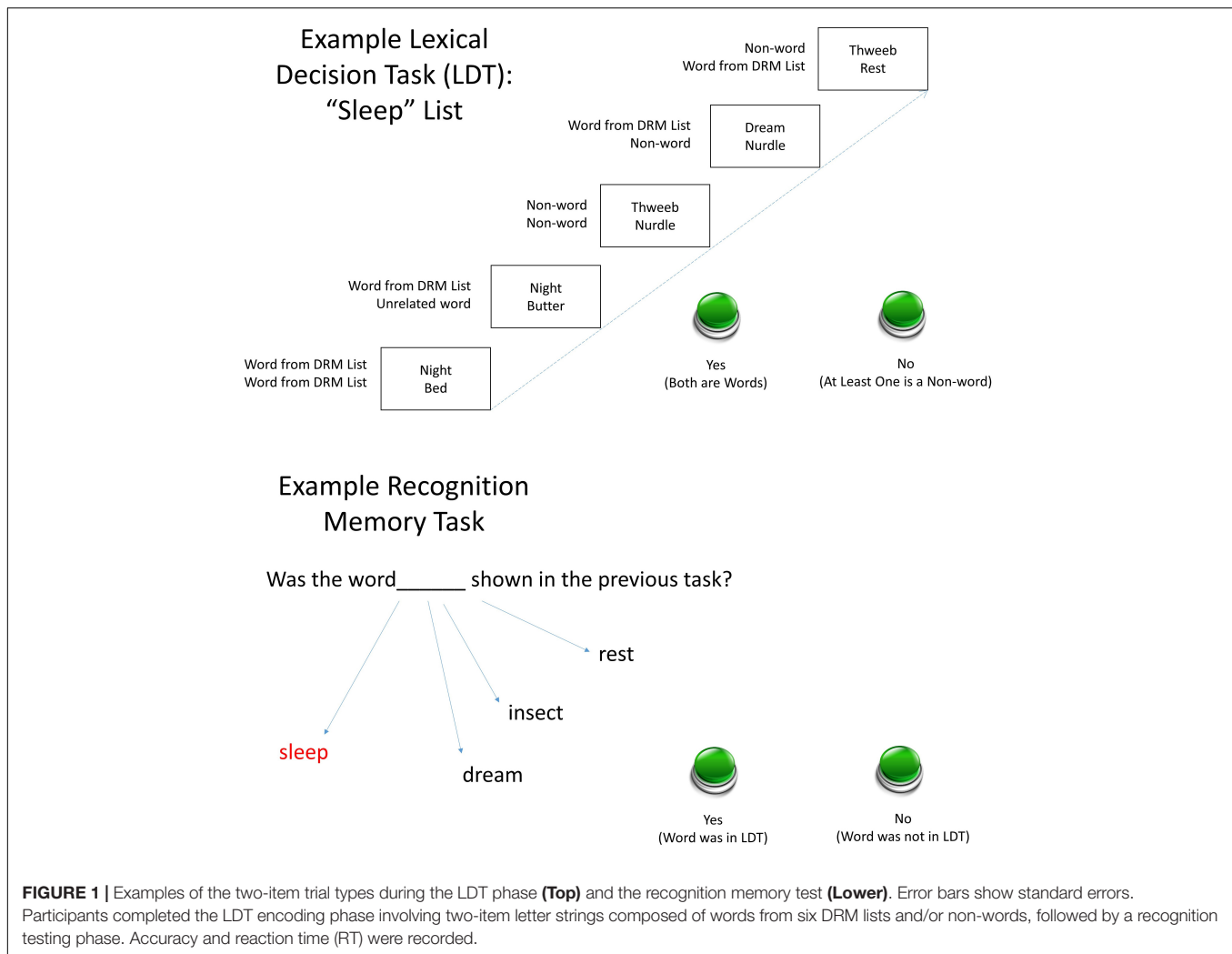
At this juncture, we have not gathered data for the *neurophysiological* approach, which will involve an already-designed separate fNIRS study. Ultimately, this subsequent fNIRS experiment will include an LDT paradigm similar to the behavioral work presented here. Results in the literature that delineate neural correlates of true versus false recognition are incomplete, as we will describe when we present Study 2. Therefore, when we launch an fNIRS experiment involving LDT we will focus on determining cortical activation patterns that reflect similarities and differences between true and false recognition when they are based on priming that occurs on LDT trials. What might we find in an LDT investigation that includes an examination of neural correlates? We can glean clues from the brain imaging experiment we report later. In this second experiment we used a relatively new brain imaging technique, functional near-infrared spectroscopy (fNIRS). We measured activity in the prefrontal cortex in an attempt to dissociate false recognition memories from true recognition.

Method

Student Project

In an initial evaluation of the LDT-priming method, eight college students (five female and three male; mean age = 21.14) participated in three blocks of LDTs with a recognition block after each LDT block. This was part of a classroom project. The procedural details we describe are complemented by **Figure 1**. Each participant worked at a Dell 780 Optiplex computer equipped with a standard keyboard, monitor, and mouse. Each computer ran under Windows 7. The experiment ran in a program called PEAK which presented the stimuli and recorded the responses and response times. Words for the LDT came from DRM lists.

We randomly chose six categories, referred to by critical lures for six DRM list themes for the classroom experiment (*anger*, *sleep*, *cold*, *girl*, *thief*, and *chair*), two categories for each LDT block. Each category had a critical lure. Two critical lures appeared in each block of the recognition task.



Each participant sat approximately two feet away from his or her monitor with fingers resting on the “1” and “2” keys of the keyboard. The experiment consisted of six blocks. Blocks one, three, and five presented the lexical decision task, blocks two, four, and six presented the recognition task.

A fixation point appeared in the center of the screen. The participant pressed the spacebar to initiate a trial. On each LDT trial, two letter-strings appeared, one above the fixation point, one below. Each pair consisted of either two related words (W-RW), two unrelated words (W-URW), a word above a non-word (W-NW), a non-word above a word (NW-W), or two non-words (NW-NW). The task was to push the “1”-key if both were words, or the “2”-key otherwise. The dependent variable was the time from the appearance of the pair until the button-push.

On each recognition trial, a word appeared in the center of the screen. Each word either occurred in the preceding LDT block or it did not. Of the words that did not, two were critical lures. The task was to push the “1”-key if the word appeared in the preceding LDT block, or the “2”-key if it did not. The dependent variables measured during the memory test were (a) the time from the appearance of the word until the button-push and (b) accuracy.

Each participant completed three LDT blocks. Each LDT block consisted of 25 trials and preceded a recognition block consisting of 20 trials. Within each block, each stimulus appeared in a different random order for each participant.

Follow-Up Experiment

Though testing only six participants (four female and two male; mean age = 21.63), the follow-up experiment to the student laboratory project expanded the scope to include 12 DRM lists across six separate LDT blocks. The follow-up doubled the number priming lists, and differed from the lab study by including new test words not thematically related to the 12 DRM lists and did not measure RT. Thus, the follow-up experiment is not strictly a replication and for this reason we could not pool the data from these two pilot efforts. For each LDT block in the second pilot study, pairs of items were (a) words from the top 12 associates in each of two ordered DRM lists (e.g., “sleep” and “rough” lists), (b) 18 non-words of similar length, and (c) six words from unrepresented DRM lists that were unrelated to either presented list.

More specifically, the six unrelated items were from six DRM lists which were not associated with any items on the subsequent recognition test. These formed 24 pairs of items: three DRM word-pairs from each list (e.g., “bed” and “rest” as a pair; six pairs total), three DRM list words from each list paired with non-words (six pairs total), three DRM list words from each list paired with unrelated words (six total), and six non-word pairs. Similar to the classroom study, participants judged whether or not the pair of items were both words, by responding yes or no, respectively.

The pairs resulted in 12 yes items and 12 no items if answered correctly, although correct responding was not recorded for the LDT encoding phase. Incorrect decisions are rare, and even the occurrence of a few mistakes would likely not interfere with exposure to the sources of priming which was the overarching purpose.

Each corresponding recognition block consisted of 24 words (presented individually). The task was to decide if each word appeared in the preceding LDT block, the same procedure when asking for old/new recognition judgments.

For each of the six recognition blocks, we randomly selected six of the 12 associates from each DRM list (“targets”), the critical item from each of the two lists (“critical lures”), two non-presented items from positions 13–15 of each of the two presented lists (“low associates” from the sleep/rough etc. lists), two critical lures from non-presented DRM lists (“pseudo critical lures”), and two low associates from positions 13–15 of each of two non-presented DRM lists (“pseudo low associates”). This resulted in 12 items that previously appeared and 12 that did not. During the recognition block, the dependent variable was accuracy: RT was not part of the follow-up, which we designed primarily to further pilot procedural details, while still allowing general comparisons with the laboratory project.

Results

Given the small sample sizes in both the project and subsequent experiment, we only report descriptive (means) information which we believe speaks for itself. These are preliminary findings and are to be interpreted with care as we have not fully tested this new paradigm and its efficacy. Again, we are introducing this paradigm into the LDT and false memory literatures. For the student project, **Figure 2** shows reaction time values for the

five different types of LDT encoding trials. As expected, the NW-NW combination was associated with the fastest times. Turning to testing, results indicated that participants were almost six times as likely to falsely recognize a critical lure than an unrelated lure (0.69 vs. 0.12). Regarding correct recognition of target words, participants were highly accurate with performance near ceiling. Unlike many studies in the literature, correct recognition exceeded false recognition rather than the opposite outcome.

Seven of the eight participants averaged 255 ms faster responses to falsely recognize a critical lure than to falsely recognize an unrelated lure (695 ms vs. 950 ms). Correctly classifying a critical lure took 373 ms longer (1,183 ms) than correctly classifying an unrelated lure (810 ms), which was about the same decision time for correctly responding to target words.

As noted, accuracy was the only dependent variable for the follow-up experiment. The proportion correct recognition for target words was 0.60, while false memory of critical lures occurred at a 0.50 rate. As in the student project, true memory exceeded false memory. Low associates from presented DRM lists were falsely recognized at a proportion of 0.22.

Pseudo test items were either “critical lures” (PCLs) or “low associates” (PLAs) selected from non-presented DRM lists. The PCLs can be equated with unrelated lures in the student study and novel lures in the fNIRS experiment reported in the next section. The proportion of errors, yes responses to PCLs was 0.13, while for PLAs participants chose yes at an errant proportion of 0.27. Critical lure illusory memory, as expected, exceeded all other forms of false recognition.

Discussion

These studies support, we believe, the idea that the LDT/Recognition paradigm presented here is a unique way of studying both the processes involved in false memory and those involved in lexical decision making. Given that these are pilots, we caution that the implications of the findings are preliminary. Yet the results are promising as our LDT procedure does prime false memory, and the extent of false memory observed is an indicator, we argue, of spreading activation that occurs during the LDT trials. In AMT terms, the LDT priming is sufficiently robust as it thwarts monitoring required to prevent creating significant amounts of illusory memories. This account does not rule out a Fuzzy Trace Theory explanation. Priming

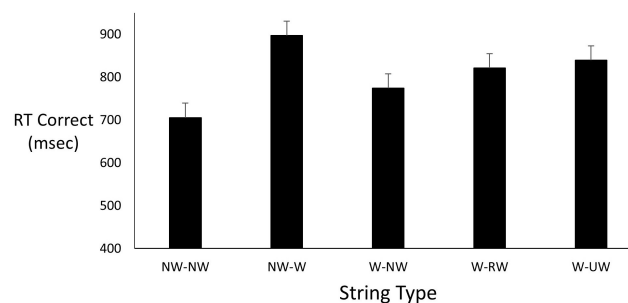


FIGURE 2 | Mean correct RT as a function of string type. String types abbreviated here are identified in the top panel of **Figure 1**.

based on verbatim processes per FTT produced very good true memory. Gist processes also aided accuracy while instilling themes supporting critical lure and low associate lure false memories. Low associates were falsely recognized far less often than critical words. This finding is predicted by FTT because these associates are much less gist consistent than critical lures that are the best exemplars of the DRM themes. Note that AMT expects this outcome as well because the critical lures are semantically linked to DRM list words more closely than are low associates. These accuracy/inaccuracy arguments are further supported by our response speed findings to which we now turn.

The fastest response times in the student project occurred on trials where both strings were non-words as participants quickly confirm neither item points to an entry in semantic memory. This is a common outcome (Meyer and Schvaneveldt, 1971). Critical lures are likely to be strongly activated due to their close links to target words compared to more semantically distant links with unrelated lures. AMT is consistent with quicker false affirmation responses to critical than to unrelated lures. Similarly, FTT predicts this pattern – during LDT trials durable theme-consistent traces of critical words are developed *via* repeated cueing of gist, while unrelated lures at test are generally gist-inconsistent.

This paradigm adds the dimension of individual differences analysis. Although only a small sample participated in the student laboratory project at least one interesting non-typical pattern emerged. In the Section “Results,” we mentioned for the lab project that 7 of the 8 participants falsely recognized a critical lure faster than an unrelated lure. The eighth participant also showed an atypical result in the LDT, correctly responding faster to W-UW pairs than to W-RW pairs. We note this case, not to speculate on its theoretical significance, but to express a caution. Specifically, rather than search for a one-size-fits-all theory, we should perhaps focus on the particular processes that individuals use. We further mention individual differences in the General Discussion section.

STUDY 2

Introduction

To elucidate the neural distinctions between real (true) memories, false memories, and lies, researchers have used various methods and experimental paradigms. Different but overlapping bilateral regions of the brain have been implicated in each of these processes, with the prefrontal cortex playing a major role in all three.

True memories activate several regions of the prefrontal cortex depending on the type of task and the material being encoded, with greater levels of activation predicting a higher likelihood of retrieval success (for review see Fletcher and Henson, 2001; for meta-analyses see Davachi et al., 2001; Barde and Thompson-Schill, 2002; Wager and Smith, 2003). False memories activate many of the same prefrontal areas that true memories do: This might reflect this region's involvement in monitoring, verifying, and reporting retrieved information (for meta-analysis see Kurkela and Dennis, 2016). Lying is an effortful

and elaborative process that requires integrating information in working memory, inhibiting one's own behavior, and monitoring others' behavior, so it is not surprising that it requires even more prefrontal resources than truth-telling and false memories (for meta-analysis see Yu et al., 2019).

For all three processes researchers report that the specific frontal areas activated depends on the methodologies used and task-related factors such as the procedure, the stimuli used, the participants' goals and intentions, and the social context (Fletcher and Henson, 2001; Barde and Thompson-Schill, 2002; Wager and Smith, 2003; Kurkela and Dennis, 2016; Yu et al., 2019; Pinti et al., 2021). Due to these variations in methodology, as well as the small sample sizes common in brain imaging studies, the specific role of particular frontal areas in memory processes is still unclear.

Functional Near-Infrared Spectroscopy

Functional near-infrared spectroscopy is a relatively new brain imaging technique that allows researchers to study memory processes in the cortex. Light emitting diodes (LEDs) placed on the scalp safely emit two wavelengths of infrared light that hemoglobin in the blood absorbs. Detectors measure the light that is refracted and the modified Beer-Lambert law enables us to quantify the level of blood oxygenation. Like the blood-oxygen-level-dependent (BOLD) response in fMRI, blood oxygenation in fNIRS is a measure of cortical activation. fNIRS has the advantage of being inexpensive, portable, not requiring participants to lay prone and still, and not requiring that trials be presented in blocks. Importantly, fNIRS has higher temporal resolution, but lower spatial resolution, than fMRI (Ferrari and Quaresima, 2012). Numerous studies, including those involving true memories, false memories, and deception, have validated fNIRS results by comparing them to those obtained with fMRI (Tian et al., 2009; Ding et al., 2013, 2014; Bhutta et al., 2015; Li et al., 2018; Pinti et al., 2021).

We used fNIRS to measure prefrontal cortical responses in a DRM false memory paradigm. Our experimental design concisely evaluates true memories, false memories, and lying in a single within-participants study. To our knowledge, only one other study, Abe et al. (2008) published more than 10 years ago, has compared all three in a single investigation.

Method

The sample was 33 college students, 28 females and 5 males, with an average age of 21.79 ± 4.46 years old. A majority of the participants were right-handed ($n = 30$) and white ($n = 19$). Each participant entered the laboratory individually. We used the NIRScout continuous-wave fNIRS system from NIRx Medical Technologies (Berlin, Germany) to measure blood oxygenation levels in the entire prefrontal cortex. We did this with a 38-channel prefrontal montage containing 16 two-wavelength (760 and 850 nm) LED sources and 12 avalanche photodiode detectors set 3 cm apart on the scalp. The step frequency of NIRScout is 62.5 Hz, thus the time-multiplexed sampling rate of each of our channels is 3.91 Hz. To record the raw data we used the NIRx NIRStar acquisition software (version 15.2 NIRx Medical Technologies LLC, Berlin, Germany) which detects refracted light from the sources and calculates oxyhemoglobin levels for each

channel in the montage. This allowed us to compare relative levels of brain activity in each prefrontal region.

Inquisit presentation software (Inquisit 5, 2016) auditorily presented DRM lists. The participant heard each word once with a 1 s delay between consecutive words. We selected the lists from a study by Stadler et al. (1999) who evaluated 36 lists for their level of identifiability. To determine if identifiability (i.e., how easily a person can detect the critical lure) was similar across phases of the study, we evaluated the backward association strength (BAS; how strongly associated the critical lure is to the list words) using normed mean BAS scores from Roediger et al. (2001). The average BAS scores (phase 1 studied words = 0.216, phase 2 studied words = 0.167, and novel non-studied words = 0.178) are fairly consistent across conditions, indicating a similar level of free recall association between the critical lure words and the list words.

In the first phase of the experiment, participants heard eight DRM lists in a randomized order, each containing fifteen words presented in the standard manner of descending order of association to the non-presented critical lure. Examples of the lists are shown in **Table 1**.

After hearing all eight lists, the participants performed a visual recognition task in which they had to decide if a test item was an old word (i.e., presented on the previous lists) or a new word (i.e., had not been heard previously). The recognition task consisted of 56 trials which included 24 words from the DRM lists (3 from each list; positions 1, 5, and 11), the critical lure words from all eight lists, and 24 words that were novel with low association with the DRM lists. Participants saw each word one at a time with options, labeled on the screen, to press “Z” to indicate the word was old and “/” to indicate the word was new. Words appeared on a white background in black print in the center of the screen. Each word remained onscreen until selection, and we recorded the decision time. Following the selection, a blank screen appeared for 5 s so that we could continue to measure the hemodynamic response. Next, participants saw a confidence rating scale which

asked them to rate their level of confidence for each of their old/new choices. The scale appeared in increments of five and ranged from *not at all confident* (0) to *very confident* (100). We averaged each trial confidence score across all forms of correct and incorrect recognition. The next trial began 500 ms later.

After the first phase was complete, we explained the nature of DRM lists and their corresponding critical lures to participants. Next, in the 15-item word lists but before each, we informed them of the critical lure word that was not on the list. They were instructed to recognize the critical lure as an old word on the memory test during the deception phase, despite knowing it was not on the list. Seven trials immediately followed each of the eight lists and included three words from the list, three novel words, and the critical lure. Participants again had to determine if these words were “old” or “new” by selecting the “Z” or “/” key and then rate their confidence on a 0–100 scale.

Finally, participants answered a brief demographic questionnaire and completed an 18-question Need for Cognition scale (Cacioppo and Petty, 1982). Individuals who have a higher need for cognition tend to commit more false memories than individuals with a lower need for cognition (Graham, 2007; Leding, 2011).

Before calculating oxyhemoglobin levels, we preprocessed each participant’s raw optical data. To do this, we used NIRx nirsLAB software (version 2019.04, NIRx Medical Technologies, LLC, Brooklyn, NY, United States). To check the data’s signal-to-noise ratio, we used a coefficient of variance filter of 7.5%. We considered any channels that exceeded this threshold as noisy and excluded them from the data. We used a band pass filter (0.01–0.2 Hz) to filter the data and capture the effects of our experimental paradigm. This reduced the effects of physiological noise, such as heart and respiration rates.

Finally, we used the modified Beer-Lambert Law to calculate oxyhemoglobin levels. The refraction path of infrared light is affected by its travel through bone and tissue, and the thickness and density of these differs with age. Accordingly, we used each participant’s age to set the differential path link factor parameter for each wavelength of light, as suggested by Scholkmann and Wolf (2013). This is important because we used the differential path link parameter to calculate oxyhemoglobin levels. We block averaged the resulting oxyhemoglobin levels across memory conditions for each channel for each participant and exported them from nirsLAB to SPSS for statistical analysis. (NIRScout also measures deoxygenated and total hemoglobin levels, but as most of the current fNIRS literature primarily focuses on less-noisy oxyhemoglobin levels, we have done the same). For each phase of the study, we compared cortical activity that occurred 1 s before (as a baseline) to the maximum oxyhemoglobin response occurring within 5 s after each event marker. Thus, for the first and deception phases of the experiment, we examined oxyhemoglobin levels within 5 s after participants were asked to indicate whether a given word was old or new. According to Vega et al. (2016), the hemodynamic response to a participant’s lie or truth occurs during this timeframe.

Because most of our participants had a few noisy channels that we removed from the data, we used the SPSS mixed procedure to estimate a two-factor (6 memory conditions × 6 brain regions)

TABLE 1 | Sample DRM lists with critical lures typed in bold and underlined.

Anger	Chair	Doctor	Mountain
Mad	Table	Nurse	Hill
Fear	Sit	Sick	Valley
Hate	Legs	Lawyer	Climb
Rage	Seat	Medicine	Summit
Temper	Couch	Health	Top
Fury	Desk	Hospital	Molehill
Ire	Recliner	Dentist	Peak
Wrath	Sofa	Physician	Plain
Happy	Wood	Ill	Glacier
Fight	Cushion	Patient	Goat
Hatred	Swivel	Office	Bike
Mean	Stool	Stethoscope	Climber
Calm	Sitting	Surgeon	Range
Emotion	Rocking	Clinic	Steep
Enrage	Bench	Cure	Ski

Lists are from Roediger and McDermott (1995).

within-subjects ANOVA (Enders, 2010). The memory conditions were Correct Old Memory (correct recollection of words on the DRM lists), Correct New Memory (correct rejection that a new word was not on the lists), Wrong Old (incorrectly stating that a word was not on the list when it was), Wrong New (incorrectly stating that a novel word was on the list when it was not), False Memory (incorrectly stating that a lure word was on a list when it was a new word), and Deception (participant stated the critical lure was on a DRM list after being instructed to do so when it was a new word). We assessed three brain regions of interest separately in the left and right frontal lobes: two channels in each anterior prefrontal cortex (APFC), eight channels in each dorsolateral prefrontal cortex (DLPFC), and five channels in each ventrolateral prefrontal cortex (VLPFC). We measured the peak oxyhemoglobin response for each channel during the 5-s period following each response and, for each region of interest, averaged their channels' peak responses. Data for Correct New Memories, Correct Old Memories, Wrong New Memories, Wrong Old Memories, and False Memories were each pooled across the recognition and deception phases. Paired-sample *t*-tests assessed specific differences in memory conditions in each region of interest. To correct for multiple pairwise comparisons, we excluded those that did not meet a threshold of $p \leq 0.01$.

Results

Behavioral Factors

We evaluated and compared reaction time, confidence, false memory, true memory, and correct rejection rates, as well as need for cognition. Average response time across participants was 7.33 s ($SD = 0.38$) during the course of the experiment. All participants had at least one false memory, with an average of 6 (79.74%). Participants on average correctly responded to 16.69 studied words, true memory (69.54%), and 19.24 non-studied, novel words by selecting "new" (80.17%; thus, only about a 20% false recognition rate for novel items). Participants on average correctly responded to 6.21 deception words (77.65%). The frequency of false memories was positively correlated with the number of true memories [$r(29) = 0.55$, $p = 0.002$]. Rates of false memories of critical lures and misremembering novel words were consistent with previous literature (Toglia et al., 1999; Prohaska et al., 2016).

Participants' average confidence on responses was 66.92 ($SD = 13.11$). A paired samples *t*-test revealed no significant difference between true memory confidence ($M = 68.55$, $SD = 13.63$) and false memory confidence ($M = 73.46$, $SD = 15.66$). The number of false memories ($M = 6.37$, $SD = 1.84$) was positively correlated with confidence [$r(31) = 0.49$, $p < 0.001$] while we found no relationship between the number of true memory or correct rejections and confidence. Additionally, the number of novel words reported as old ($M = 5.61$, $SD = 4.98$) was positively correlated to confidence as well [$r(31) = 0.44$, $p = 0.019$]. Notice that accuracy was not associated with confidence, while inaccuracy was positively related to confidence consistent with the general conclusion in the literature that confidence in memory does not guarantee its accuracy.

The number of false memories was negatively related to false memory reaction time ($M = 7.58$, $SD = 0.77$), meaning that the more false memories a participant had, the faster their response when committing a false memory of critical lures [$r(31) = -0.39$, $p = 0.02$]. This pattern did not occur for false memories of novel words. We found no difference between response time for correct rejections or true memories compared to false memory responses but participants took significantly longer to respond when committing a false memory compared to when telling an intentional lie ($M = 7.01$ ms, $SD = 0.85$), [$t(29) = -3.77$, $p < 0.001$].

Need for Cognition

We evaluated Need for Cognition (Cacioppo and Petty, 1982) by using a mid-point of zero: a positive score indicated a high need for cognition, a negative score indicated a low need for cognition. On the Need for Cognition scale, which ranged from -72 to 72, participants averaged 24.34. Higher scores were positively correlated with confidence in correct responses to studied words [$r(29) = 0.34$, $p < 0.05$], incorrect responses for studied words [$r(29) = 0.39$, $p < 0.05$], and correct responses to non-studied words [$r(29) = 0.52$, $p < 0.005$]. Need for Cognition was also positively correlated with the number of correct responses to studied words [$r(29) = 0.37$, $p < 0.05$] but was not related to false memory rate, reaction time, nor false memory confidence.

Brain Activation

A repeated measures ANOVA found main effects for brain region [$F(5,3278) = 6.03$, $p < 0.001$, $f = 0.088$] and memory condition [$F(5,3265) = 15.16$, $p < 0.001$, $f = 0.147$] and no omnibus interaction between brain region and memory conditions [$F(5,3264) = 0.92$, $p = 0.58$]. The significant pairwise comparisons ($p \leq 0.01$) for brain regions and memory conditions appear in **Table 2**.

True memories (correctly remembering that a word was on a DRM list) produced significantly less activation of prefrontal areas than did Deception, False Memories, Wrong Old memories, and Wrong New memories. Also, incorrectly stating that a new word was on a list (Wrong New) produced more activation than failure to recall that a word was on a list (Wrong Old).

During all our memory conditions, we found significantly more activation in the left dorsolateral prefrontal cortex than the right VLPFC and both left and right APFC. In addition, we measured greater activity in the right DLPFC than both left and right APFC, and greater activity in the left VLPFC compared to right APFC.

Despite lack of a global interaction, pairwise comparisons indicated several significant differences ($p \leq 0.01$) between memory conditions in specific prefrontal regions. **Table 3** presents these comparisons.

Correct Old memories ("hits" – correct recognition of words presented on the DRM lists) induced greater activation in the left DLPFC than the left APFC, right APFC, right DLPFC, and right VLPFC. Correct Old memories (hits) also produced higher activation in the right DLPFC than right APFC. In both left and right DLPFC, being correct about a critical lure produced more activity than did true memory hits.

TABLE 2 | Pairwise comparisons for memory conditions and brain regions of interest.

(I) Memory	(J) Memory	Mean difference (I–J)	df	Sig
Correct new (Correct rejection)	Correct old (Hit)	2.79E-06	3264.113	0.851
	Deception	−7.431E-5*	3267.423	<0.001
	False	−7.080E-5*	3263.795	<0.001
	Wrong new	−9.522E-5*	3263.795	<0.001
	Wrong old (miss)	−3.997E-5*	3263.795	0.007
Correct old (hit)	Deception	−7.710E-5*	3267.839	<0.001
	False	−7.359E-5*	3264.113	<0.001
	Wrong new	−9.801E-5*	3264.113	<0.001
	Wrong old	−4.276E-5*	3264.113	0.004
Deception	False	3.51E-06	3267.423	0.814
	Wrong new	−2.09E-05	3267.423	0.162
	Wrong old	3.434E-5	3267.423	0.022
False	Wrong new	−2.44E-05	3263.795	0.099
	Wrong old	3.083E-5	3263.795	0.037
Wrong new (false alarm)	Wrong old	5.525E-5*	3263.795	<0.001

(I) ROI	(J) ROI	Mean difference (I–J)	df	Sig
LDLPFC	LAPFC	5.096E-5*	3278.775	<0.001
	LVPFC	3.287E-5	3273.316	0.026
	RDPFC	1.21E-05	3263.956	0.401
	RAPFC	7.151E-5*	3278.457	<0.001
	RVPFC	3.837E-5*	3273.355	0.009
LAPFC	LVPFC	−1.81E-05	3289.922	0.235
	RDPFC	−3.884E-5*	3279.203	0.009
	RAPFC	2.06E-05	3271.672	0.178
	RVPFC	−1.26E-05	3288.178	0.409
LVPFC	RDPFC	−2.08E-05	3271.615	0.159
	RAPFC	3.864E-5*	3289.577	0.011
	RVPFC	5.50E-06	3271.86	0.714
RDPFC	RAPFC	5.939E-5*	3278.886	<0.001
	RVPFC	2.63E-05	3271.67	0.075
RAPFC	RVPFC	−3.314E-5	3287.841	0.029

LDLPFC, left dorsolateral prefrontal cortex; LAPFC, left anterior prefrontal cortex; LVPFC, left ventrolateral prefrontal cortex; RDPFC, right dorsolateral prefrontal cortex; RAPFC, right anterior prefrontal cortex; RVPFC, right ventrolateral prefrontal cortex. * indicates ($p \leq 0.01$).

Compared to Correct Old memory hits, incorrect memories (Wrong Old “misses” or Wrong New “false alarms”) led to increased activity in left DLPFC, left VLPFC, right DLPFC, right VLPFC, and right APFC. This was especially notable in the right DLPFC where each incorrect memory condition produced greater activity than true memory hits. Wrong Old memories (misses) produced the greatest activity in left and right DLPFC while a pattern was less clear for Wrong New memories (false alarms).

Compared to true memory hits, participants’ False memories led to greater activation in the left DLPFC, the right DLPFC, and the left APFC, and False memories produced greater activity

in those three regions than they did in right APFC. Compared to Wrong Old memories (misses; forgetting an old word), false memories induced greater activation in the left APFC and the left VLPFC. Similar to False memories, Deception produced more activation in the left DLPFC and the right DLPFC when compared to hits. Deception produced higher levels of activity in the left DLPFC than right APFC.

We used bivariate correlations to compare confidence scores to oxygenated hemoglobin levels for each condition within each region of interest. Confidence levels were positively correlated with hemoglobin levels in the right VLPFC when participants incorrectly stated that new words were old [Wrong New memories; false alarm, $r(29) = 0.46$, $p = 0.02$] and old words were new [Wrong Old memories; miss, $r(29) = 0.61$, $p = 0.001$]. For forgotten old words (Wrong Old memories; misses), confidence increased with oxygenated hemoglobin in the left VLPFC [$r(29) = 0.38$, $p = 0.05$]. Finally, in the left DLPFC, confidence increased with oxygenated hemoglobin in the False Memory (critical lure) condition [$r(29) = 0.38$, $p = 0.04$].

Discussion

True and False Memories

False memory data for critical lures yielded significant behavioral effects and brain activation. Higher rates of false memories correlated with true memories. This pattern is consistent with previous research showing that better recall of list items is associated with more false memories and that this relationship is stronger when word lists are thematically blocked as opposed to randomly ordered (Toglia et al., 1999). According to Fuzzy Trace Theory, a gist understanding of the word lists helps participants recognize previously presented words, but also increases the likelihood of false memories. This pattern is also consistent with Activation-Monitoring Theory which predicts that target words and critical lures are highly and repeatedly activated at encoding and pass a monitoring examination during retrieval. Individuals with better accurate (true) memory and more false memories might pay closer attention to the list presentations, and they might use strategies such as chunking which could make them more likely to misremember semantically related words. We found that participants with higher need for cognition had better recall, further supporting this idea. On the other hand, our participants with high need for cognition did not have more false memories, contrary to prior research (Graham, 2007; Leding, 2011). Those high in need for cognition did, however, have more overall confidence in their ratings regardless of the number of false memories (Kuvaas and Kaufmann, 2004).

Participants with more false memories of critical lures responded faster and were more confident in their responses. The quicker reaction time might indicate more reliance on gist trace memories rather than verbatim memory. Interestingly, response time was not related to novel false memories despite having a similar correlation with confidence. Fuzzy Trace Theory suggests that inhibitory processes repress false memory responses (Reyna and Mills, 2007), thus the faster response time might indicate that these participants are not engaging in as much inhibitory control. Our fNIRS data, however, which indicates

TABLE 3 | Significant differences ($p \leq 0.01$) between memory conditions in specific prefrontal regions.

Brain region/memory condition	Brain region/memory condition	Mean difference	<i>t</i>	df	Sig
LDLPFC correct old (hit)	RDLPFC correct old	2.97E-05	2.721	27	0.011
	LAPFC correct old	4.48E-05	2.801	25	0.010
	RAPFC correct old	5.03E-05	3.740	25	0.001
LDLPFC correct new (correct rejection)	RVL PFC correct new	5.80E-05	3.519	26	0.002
	LAPFC correct new	4.08E-05	3.712	25	0.001
	RAPFC correct new	5.89E-05	5.106	25	0.000
RDL PFC correct new (correct rejection)	RAPFC correct new	4.56E-05	3.499	25	0.002
LDLPFC correct lure (correctly reject lure)	LDLPFC correct old	2.05E-04	2.690	21	0.014
RDL PFC correct lure	RDL PFC correct old	2.25E-04	3.223	20	0.004
	RDL PFC correct new	1.81E-04	2.869	21	0.009
	LVL PFC wrong old	8.91E-05	3.709	26	0.001
LDLPFC wrong old (miss)	LAPFC wrong old	9.40E-05	3.590	25	0.001
	RAPFC wrong old	5.26E-05	2.979	25	0.006
	LVL PFC wrong old	1.01E-04	3.792	26	0.001
RDL PFC wrong old (miss)	LAPFC wrong old	1.07E-04	4.832	25	0.000
	RAPFC wrong old	6.80E-05	2.688	25	0.013
	RDL PFC correct old	1.07E-04	3.697	27	0.001
	RDL PFC correct new	7.97E-05	3.186	28	0.004
	LDLPFC correct old	1.22E-04	3.025	28	0.005
LDLPFC wrong new (false alarm)	RAPFC wrong new	1.18E-04	3.035	25	0.006
	RDL PFC correct old	1.35E-04	3.568	27	0.001
RDL PFC wrong new	RDL PFC correct new	1.05E-04	3.046	28	0.005
	RAPFC wrong new	1.03E-04	3.415	25	0.002
	LVL PFC wrong new	1.11E-04	2.715	26	0.012
LVL PFC wrong new	LVL PFC wrong old	1.27E-04	3.220	26	0.003
	LVL PFC correct new	1.14E-04	2.685	26	0.012
	RVL PFC correct new	9.56E-05	2.663	26	0.013
RVL PFC wrong new	RAPFC correct new	6.61E-05	2.702	25	0.012
RAPFC wrong new	LDLPFC correct old	6.15E-05	2.652	28	0.013
	RAPFC false memory	6.81E-05	3.584	25	0.001
LDLPFC false memory	RDL PFC correct old	7.45E-05	2.717	27	0.011
	RAPFC false memory	4.85E-05	3.187	25	0.004
RDL PFC false memory	LVL PFC wrong old	8.97E-05	2.666	26	0.013
LVL PFC false memory	LAPFC correct old	1.30E-04	3.540	25	0.002
	LAPFC wrong old	1.15E-04	3.207	25	0.004
	LAPFC correct new	1.26E-04	3.947	25	0.001
	RAPFC false memory	9.27E-05	3.107	24	0.005
LDLPFC deception	LDLPFC correct old	7.21E-05	3.069	27	0.005
	RAPFC deception	6.46E-05	3.303	25	0.003
RDL PFC deception	RDL PFC correct old	1.01E-04	2.902	26	0.007

LDLPFC, left dorsolateral prefrontal cortex; *LAPFC*, left anterior prefrontal cortex; *LVL PFC*, left ventrolateral prefrontal cortex; *RDL PFC*, right dorsolateral prefrontal cortex; *RAPFC*, right anterior prefrontal cortex; *RVL PFC*, right ventrolateral prefrontal cortex.

more activity in the prefrontal cortex during false memories compared to true memories, does not fit that interpretation. The forensic psychology literature has seen a push to distinguish accuracy from confidence (Busey et al., 2000; Chua et al., 2004; Storbeck and Clore, 2005). Often in the legal setting jurors assume that a person who is confident in their testimony must be correct, but many studies have shown little to no correlation between these constructs, a premise our study supports (Sporer et al., 1995; Ais et al., 2016). Together these results suggest that individuals are more confident despite having greater levels of false memories, which then might lead to them respond more quickly. Thus, neither confidence nor reaction time serve as proxies for accuracy. High confidence scores during false memories fall in line with AMT and the spreading activation

theory. As participants encode words they are likely creating associative networks that include the critical lures. This would make it more likely for participants to perceive these lures as target words and do so confidently.

Several areas of the prefrontal cortex were activated during all of our memory conditions but false memories (falsely remembering a critical lure) and, similarly, incorrectly thinking that other new words were old, produced more activity in the prefrontal cortex than did true memories and correct rejections. The prefrontal region most activated by true memories was the left DLPFC while false memories increased hemoglobin levels in both left and right DLPFC. Yu et al. (2019) performed a meta-analysis of 77 fMRI studies and similarly concluded that while false memories recruit several regions of the frontal,

parietal, and temporal lobes, the left DLPFC is more activated by false memories than true ones. The DLPFC, especially on the right side, according to some researchers, has been implicated in the appraisal of the value of the memory for current task performance during post-retrieval monitoring (Henson et al., 1999; Achim and Lepage, 2005; Chua et al., 2009; Wang et al., 2016). Increased DLPFC activity might also reflect the monitoring component of AMT as participants evaluate whether a lure was on the list. This would require more cognitive effort and likely some inhibition if their first instinct is to semantically associate the lure word. We also found an increase in activity in the left APFC. Using positron emission tomography (PET) with the DRM paradigm, Schacter et al. (1996) found a similar pattern of activation, and, by measuring increased blood flow in the APFC and PFC, they were able to distinguish false from true memories.

Kurkela and Dennis (2016) meta-analysis also implicated multiple prefrontal regions, mostly medial to those accessible to fNIRS, as well as the bilateral inferior frontal gyri which corresponds to the VLPFC where we found, in the left hemisphere, that false memories elicited more activity than forgetting an old word did. This was unexpected because previous research has implicated the right, not left, VLPFC in evaluating uncertain information for accuracy (Chua et al., 2009; Goel et al., 2009; Ito et al., 2012). Further research is required to untangle this discrepancy.

Because true and false memories tend to mostly activate the same cortical regions (Johnson et al., 1997; Schacter and Addis, 2007; Mitchell and Johnson, 2009; Kurkela and Dennis, 2016; Yu et al., 2019), it is difficult to determine whether there are definitive gist neural pathways or regions that respond specifically to false memory information. Our results are consistent with previous literature and are congruent with both FTT and AMT, but larger imaging studies focusing on areas of distinction such as the anterior prefrontal cortex, are needed to distinguish any subtle differences between these responses.

Deception

Numerous imaging and brain stimulation studies have confirmed that the prefrontal cortex is the generator of lies, whether the lie is simply a response to an instruction by an experimenter or is a spontaneous and deliberate attempt to deceive. Lying involves several cognitive processes that are thought, at least in laboratory situations, to require more mental effort than does telling the truth. A liar must hold remembered information in working memory, consider the consequences of lying vs. truth-telling, suppress the urge to voice the truth, concoct a narrative that is contrary to the truth but still believable, attempt to avoid “tells” by regulating eye movements, facial expressions, and body language, and infer how the lie is being received by the listener (for review see Gombos, 2006).

In our DRM study, compared to true memories, deception produced more activation in prefrontal areas, especially the left and right DLPFC. This conforms with a large meta-analysis of fMRI studies which also found that deception, whether instructed or spontaneous, increased PFC activity, especially in the left and right DLPFC (Yu et al., 2019). fNIRS has yielded similar results (Ding et al., 2013, 2014) and very recently Lin et al. (2021)

used fNIRS to measure cortical activity in the DLPFC and the APFC while participants lied while playing a poker game with an opponent. Both prefrontal areas were more active when participants lied versus when they told the truth. Interestingly, Li et al. (2018) used fNIRS to show that the left middle frontal gyrus, which overlaps with the left DLPFC, reacts most strongly to deceptive responses by participants who lie rarely compared to those who lie regularly or to those who tell the truth.

It is notable that all of our memory conditions increased hemoglobin levels in the left DLPFC. Ito et al. (2012) suggest that creating all types of memory and lies are taxing and that the left DLPFC is responsible for preparing us for both truthful and deceptive responses. They used fMRI to measure brain activity while participants prepared to tell the truth or to lie about photographs they had been shown, and then, several seconds later, during the actual lie or truth-telling. The left DLPFC was significantly more active during the preparation phase when participants knew whether they would be asked to tell the truth or to lie compared to trials in which they did not know in advance whether they would be asked to respond with truth or deception. Interestingly, during this preparation phase, the left DLPFC was equally active in both the truth and lie conditions. During the execution phase, however, they found, as we did, that the left DLPFC was more active when participants told a lie than when they told the truth.

These results support cognitive load theories of deception which suggest that executing a lie requires more cognitive resources than truth telling because telling the truth is usually an automatic response while lying involves additional steps such as constructing an alternative response while suppressing possible indicators of deception. These extra steps require the recruitment of additional cognitive resources while the lie is in progress (Zuckerman et al., 1981; Vrij et al., 1996; Walczyk et al., 2003). It is possible that in our experiment, cognitive load increased further when we asked participants to report that the critical lure was a lie. According to FTT, remembering specific words, as opposed to the lists' themes, requires reliance on verbatim trace memory and is likely to require more cognitive effort than simply remembering similar words with a semantic association. This increased cognitive load is probably also reflected in the DLPFC when we asked our participants to lie.

Further increasing cognitive load by adding additional effortful tasks can expose lies. Vrij et al. (2008) required mock suspects to tell their story to police officers in a chronologically backward order. With this additional cognitive load, the suspects could not suppress noticeable deception cues and the officers were better able to detect the liars. Our results, as well as those of others, show that fNIRS can detect differences in cognitive load (Fishburn et al., 2014) and could be used as an inexpensive, portable, and effective lie detector, at least for infrequent liars who seem to experience less cognitive load during deception (Li et al., 2018). Using fNIRS, computers and humans can be trained to distinguish truth from lies by simply viewing static images that reflect the relative changes in hemoglobin levels occurring in an interviewee's prefrontal cortex during questioning (Vega et al., 2016). Combining fNIRS with a traditional polygraph system is an even better solution that significantly improves lie detection (Bhutta et al., 2015).

Though both false memories and deception produced more activation in prefrontal areas than true memories and correct rejections, we did not find an activation difference between the two conditions. This may indicate that our low-stakes deception task, in which participants were instructed to lie about remembering words on a list, does not produce the same level of cognitive processing required of intentional face-to-face deception. Indeed, Lin et al. (2021) report that in their poker game study, deliberate spontaneous deception produces higher activity in both the DLPFC and APFC than when participants are merely told to lie. In a similar recent fNIRS study with card players that focused exclusively on the APFC, the highest levels of activity occurred only in intentional face-to-face deception (Pinti et al., 2021). This most anterior region of the prefrontal cortex, which is poorly understood, has been implicated in numerous cognitive tasks including episodic memory retrieval, social cognition, and mentalizing – all tasks that are presumably necessary to meet the demands for effective intentional deception (Gilbert et al., 2006; Burgess et al., 2007). Taken together, our findings suggest that all types of memory activate the prefrontal cortex, especially the DLPFC, but false memories and deception, which likely require more cognitive resources, result in even higher levels of activity and that the coordination of several regions of the prefrontal cortex is involved in these processes.

General Discussion

Twenty-five years of research shows that the DRM (Deese–Roediger–McDermott) paradigm impressively produces false memories of non-presented critical lures as well as inaccurate remembering of other lures. These compelling indications of the DRM illusion occurred in the two investigations we report in this article. False memory researchers continue to publish new demonstrations of the DRM illusion while concentrating on advances that fall into at least three categories: theoretical accounts, strategies for reducing false memories, and real-world implications. As we address these classifications, we begin by briefly mentioning the disparate nature of our LDT and fNIRS investigations and then turn to spelling out the relationships between these research agendas.

They are certainly different lines of research. Study1 introduces a new approach to engendering false memory by using a lexical decision task as a priming agent. The fNIRS (functional near-infrared spectroscopy) experiment measures oxygenation changes in the prefrontal cortex. Our purpose was to leverage cognitive neuroscience findings to guide analyses of cortical changes across truth-telling, deception, and false memory. In a mostly traditional DRM paradigm, we examined these three in a within-participants design, a combination that is not a common research strategy (again see Abe et al., 2008 for a similar study). Deception occurred *via* intentional lying instructions, and in all phases of this experiment we collected neural correlate data. These different approaches to studying illusory memory share many similarities which permit an interesting window into how results from one DRM priming technique (LDT) under incidental memory testing conditions generalize to findings obtained under another DRM experimental design using intentional memory instructions. Indeed, the behavioral results for true and false

memory across experiments indicate DRM priming is powerful even without intent to remember.

Response times were similar for true and false memories, a finding observed in both the laboratory project and the fNIRS study, so RT was not a distinguishing factor. Both investigations tested recognition memory with a focus on DRM-created memories in the genre of spontaneous false memories, as opposed to implanted (suggested) false remembrances. In DRM studies, including our LDT and fNIRS research, experimenters do not actively cause inaccuracies. Rather, *via* autosuggestion participants generate faulty theme-consistent memories (Brainerd and Reyna, 2005). These semantically based illusory memories are consistent with AMT's positions concerning activation at encoding and monitoring failures at retrieval, and with FTT's reasoning grounded in the formation of strong gist traces. These theories are compatible with the more-is-less pattern of increases in true memory accompanied by increases in false memory (Toglia et al., 1999) that we saw in both investigations. Though participants demonstrated very good target recognition, their overall accuracy suffered by committing high levels of false recognition memory. This constitutes an argument for memory impairment in the DRM paradigm.

Both investigations employed many of the same DRM lists. The LDT studies involve the lists of Toglia et al. (1999) which successfully fomented false memories at levels sufficient to potentially see the illusion with incidental memory assessment. In the fNIRS experiment the chosen DRM study lists corresponded to highly “identifiable” critical lures. This means that the LDT lists also referred to decidedly identifiable critical lures. Regardless, robust false memory levels occurred in the present investigations! This suggests that in daily life some false information is likely discernable as untrue and yet people may believe such easy to identify false messages, subsequently expressing them as memories deemed accurate. Thus, identifiability does not seem to serve as an implicit warning to reduce false memory, a conclusion consistent with failures of explicit warnings to attenuate illusory recollections (Neuschatz et al., 2001). Our behavioral findings on lying are relevant here because instructions to intentionally lie are the ultimate form of warning participants of the existence of the DRM illusion. Nevertheless, we saw that participants failed about 20% of the time to provide the lie that critical lures had been previously heard. Apparently, they ignored or forgot the instructions. Forgetting is quite possible because we'd expect verbatim traces of instructions to fade quickly in this context (Reyna and Brainerd, 1995). A related possibility involving loss of verbatim information, raises the notion that a memory lapse here is more complex than just not recalling an instruction. What if participants simply do not remember some of the critical lures for which they were asked to provide a lie? This scenario or a combination of the two suggest forgetting is far more plausible than mere ignorance. Thus, lying might not always occur under conditions that should permit perfect identifiability of the lure words. We also favor this forgetting explanation because it fits with the cognitive load problems exposed in the fNIRS study.

The preceding argument leads to two issues. The first is a reminder of links between false memory and deception which portend real-world applications. Our fNIRS findings, which show different degrees and/or patterns of activation in several

prefrontal regions for true memories, false memories, and lies, are consistent with the notion that these constructs are in fact different (Ceci et al., 1987; Loftus, 2005).

The second, as we mentioned, is that similarities between false memory and deception points to a positive example of ecological validity. We are aware of concerns about the external validity of the DRM procedure (Baioui et al., 2012), but the paradigm offers a straightforward way to examine issues relevant to a large body of work on schematic knowledge. Also, as noted in the Introduction, DRM experiments yield findings that line up with experimental results reported with stimulus materials regarded as more ecologically valid. These stimuli include sentences, prose passages, pictures, and visual scenes.

External validity leads us to thoughts on future paths for behaviorally-oriented DRM research and for the paradigm's role in advancing strides in cognitive neuroscience. As in many DRM studies, our participants are young adults. Extending both of our investigative techniques to testing children, middle-aged adults, and older adults will diversify our sample and add to the literature concerning the developmental trajectory of false memories (Brainerd et al., 2011). Our methodologies, especially the LDT priming, might also be useful for special populations – like individuals with intellectual disabilities such as autism spectrum disorder (Beversdorf et al., 2000). Our LDT procedure would enable us to explore individual differences such as the Need for Cognition trait that we included in the fNIRS study. Our recommendations to this juncture have been at the behavioral level.

At the neurophysiological level, considerable brain imaging research has attempted to distinguish between true and false memories and we have cited a number of studies that converge on the conclusion that these memories generally tend to show activation in the same cortical areas. Our current experiments are concordant with this conclusion but a nagging question is whether thematically-oriented pathways exist that signal the presence of false messages and lies. Answering this will require more extensive imaging experiments that

concentrate on regions of distinction such as the anterior prefrontal cortex.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: fNIRS PFC Memory and Deception, https://www.nitrc.org/projects/fnirs_drm_2021/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Office of Research & Sponsored Programs, University of North Florida, Jacksonville, FL 32224. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MT organized and wrote the manuscript. JS developed the new paradigm which is the basis of Study 1 and with MT conducted a modified follow-up experiment, wrote much of the prose for Study 1, and proofed the introductory section written by MT. BS leveraged her master's thesis advised by MT to become Study 2 in this article, conducted additional analyses, and proofed this section. KH wrote and proofed with BS the Study 2 section, edited the general discussion written by MT, and gathered the datasets in setting up the availability repository. ND worked on Study 1, proofed several sections of the present manuscript, set up the initial version of this template, and prepared the final versions of our figures. BW collected the data in the follow-up experiment within Study 1, assisted BS when she collected additional data beyond her thesis, and edited the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

- Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., et al. (2008). Neural correlates of true memory, false memory, and deception. *Cereb. Cortex* 18, 2811–2819. doi: 10.1093/cercor/bhn037
- Achim, A. M., and Lepage, M. (2005). Dorsolateral prefrontal cortex involvement in memory post-retrieval monitoring revealed in both item and associative recognition tests. *NeuroImage* 24, 1113–1121. doi: 10.1016/j.neuroimage.2004.10.036
- Ais, J., Zylberberg, A., Barttfeld, P., and Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* 146, 377–386. doi: 10.1016/j.cognition.2015.10.006
- Baioui, A., Ambach, W., Walter, B., and Vaitl, D. (2012). Psychophysiology of false memories in a Deese-Roediger-McDermott paradigm with visual scenes. *PLoS One* 7:e30416. doi: 10.1371/journal.pone.0030416
- Barclay, C. R., Toglia, M. P., and Chevalier, D. S. (1984). Pragmatic inferences and type of processing. *Am. J. Psychol.* 97, 285–296.
- Barde, L. H., and Thompson-Schill, S. L. (2002). Models of functional organization of the lateral prefrontal cortex in verbal working memory: evidence in favor of the process model. *J. Cogn. Neurosci.* 14, 1054–1063. doi: 10.1162/089892902320474508
- Bartlett, F. C. (1932). *Remembering: A Study In Experimental And Social Psychology*. Cambridge: Cambridge University Press.
- Beversdorf, D. Q., Smith, B. W., Crucian, G. P., Anderson, J. M., Keillor, J. M., Barrett, A. M., et al. (2000). Increased discrimination of 'false memories' in autism spectrum disorder. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8734–8737. doi: 10.1073/pnas.97.15.8734
- Bhutia, M. R., Hong, M. J., Kim, Y. H., and Hong, K. S. (2015). Single-trial lie detection using a combined fNIRS-polygraph system. *Front. Psychol.* 6:709. doi: 10.3389/fpsyg.2015.00709
- Bower, G. H., Black, J. B., and Turner, T. J. (1979). Scripts in text comprehension and memory. *Cogn. Psychol.* 11, 177–220. doi: 10.1016/0010-0285(79)90009-4
- Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Curr. Dir. Psychol. Sci.* 11, 164–169.
- Brainerd, C. J., and Reyna, V. F. (2005). *The Science Of False Memory: Oxford Psychology Series*. New York, NY: Oxford University Press.
- Brainerd, C. J., Reyna, V. F., and Ceci, S. J. (2008). Developmental reversals in false memory: a review of data and theory. *Psychol. Rev.* 134, 343–382. doi: 10.1037/0033-2909.134.3.343
- Brainerd, C. J., Reyna, V. F., and Zember, E. (2011). Theoretical and forensic implications of developmental studies of the DRM illusion. *Mem. Cogn.* 39, 365–380. doi: 10.3758/s13421-010-0043-2

- Bransford, J. D., and Franks, J. J. (1971). The abstraction of linguistic ideas. *Cogn. Psychol.* 2, 331–350. doi: 10.1016/0010-0285(71)90019-3
- Bransford, J. D., Barclay, J. R., and Franks, J. J. (1972). Sentence memory: a constructive versus interpretive approach. *Cogn. Psychol.* 3, 193–209. doi: 10.1016/0010-0285(72)90003-5
- Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Mem. Cogn.* 5, 673–678. doi: 10.3758/BF03197414
- Brewer, W. F., and Treyens, J. C. (1981). Role of schemata in memory for places. *Cogn. Psychol.* 13, 207–230. doi: 10.1016/0010-0285(81)90008-6
- Bui, D. C., Friedman, M. C., McDonough, I. M., and Castel, A. D. (2013). False memory and importance: can we prioritize encoding without consequence? *Mem. Cogn.* 41, 1012–1020. doi: 10.3758/s13421-013-0317-6
- Burgess, P. W., Dumontheil, I., and Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends Cogn. Sci.* 11, 290–298. doi: 10.1016/j.tics.2007.05.004
- Busey, T. A., Tunnichliff, J., Loftus, G. R., and Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychon. Bull. Rev.* 7, 26–48. doi: 10.3758/bf03210724
- Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131.
- Ceci, S. J., Ross, D. F., and Toglia, M. P. (1987). Suggestibility of children's memory: psycho-legal implications. *J. Exp. Psychol.* 116, 38–49.
- Chua, E. F., Rand-Giovannetti, E., Schacter, D. L., Albert, M. S., and Sperling, R. A. (2004). Dissociating confidence and accuracy: functional magnetic resonance imaging shows origins of the subjective memory experience. *J. Cogn. Neurosci.* 16, 1131–1142. doi: 10.1162/0898929041920568
- Chua, E. F., Schacter, D. L., and Sperling, R. A. (2009). Neural basis for recognition confidence in younger and older adults. *Psychol. Aging* 24, 139–153. doi: 10.1037/a0014029
- Collins, A. M., and Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1007/BF01074344
- Davachi, L., Maril, A., and Wagner, A. D. (2001). When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering. *J. Cogn. Neurosci.* 13, 1059–1070. doi: 10.1162/089892901753294356
- Dechtereiko, F., Lukavsky, J., and Stipl, J. (2021). False memories for scenes using the DRM paradigm. *Vis. Res.* 178, 48–59. doi: 10.1016/j.visres.2020.09.009
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* 58, 17–22. doi: 10.1037/h0046671
- Ding, X. P., Gao, X., Fu, G., and Lee, K. (2013). Neural correlates of spontaneous deception: a functional near-infrared spectroscopy (fNIRS) study. *Neuropsychologia* 51, 704–712. doi: 10.1016/j.neuropsychologia.2012.12.018
- Ding, X. P., Sai, L., Fu, G., Liu, J., and Lee, K. (2014). Neural correlates of second-order verbal deception: a functional near-infrared spectroscopy (fNIRS) study. *NeuroImage* 87, 505–514. doi: 10.1016/j.neuroimage.2013.10.023
- Ebbinghaus, H. (1885). *Memory: A Contribution To Experimental Psychology*. New York, NY: Dover.
- Enders, C. K. (2010). *Applied Missing Data Analysis*. New York, NY: Guilford Press.
- Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049
- Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8:76. doi: 10.3389/fnhum.2014.00076
- Fletcher, P. C., and Henson, R. N. (2001). Frontal lobes and human memory: insights from functional neuroimaging. *Brain* 124(Pt 5), 849–881. doi: 10.1093/brain/124.5.849
- Foley, M. A., and Foy, J. (2008). Pictorial encoding effects and memory confusions in the deese-roediger-mcdermott paradigm: evidence for the activation of spontaneous imagery. *Memory* 16, 712–727. doi: 10.1080/09658210802220054
- Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., et al. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *J. Cogn. Neurosci.* 18, 932–948. doi: 10.1162/jocn.2006.18.6.932
- Goel, V., Stollstorff, M., Nakic, M., Knutson, K., and Grafman, J. (2009). A role for right ventrolateral prefrontal cortex in reasoning about indeterminate relations. *Neuropsychologia* 47, 2790–2797. doi: 10.1016/j.neuropsychologia.2009.06.002
- Gombos, V. A. (2006). The cognition of deception: the role of executive processes in producing lies. *Genetic Soc. Gen. Psychol. Monogr.* 132, 197–214. doi: 10.3200/mono.132.3.197-214
- Graham, L. M. (2007). Need for cognition and false memory in the Deese-Roediger-McDermott paradigm. *Pers. Individ. Dif.* 42, 409–418. doi: 10.1016/j.paid.2006.07.012
- Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., and Dolan, R. J. (1999). Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. *J. Neurosci.* 19, 3962–3972. doi: 10.1523/JNEUROSCI.19-10-03962.1999
- Hornstein, S. L., Brown, A. S., and Mulligan, N. W. (2003). Long-term flashbulb memory for learning of Princess Diana's death. *Memory* 11, 293–306. doi: 10.1080/09658210244000063
- Inquisit 5 (2016). *Inquisit 5 [Computer software]*. Available online at: <https://www.millisecond.com> (accessed March 16, 2018).
- Israel, L., and Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychon. Bull. Rev.* 4, 577–581. doi: 10.3758/BF03214352
- Ito, A., Abe, N., Fujii, T., Hayashi, A., Ueno, A., Mugikura, S., et al. (2012). The contribution of the dorsolateral prefrontal cortex to the preparation for deception and truth-telling. *Brain Res.* 1464, 43–52. doi: 10.1016/j.brainres.2012.05.004
- Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L., and Curran, T. (1997). The similarity of brain activity associated with true and false recognition depends on the test format. *Psychol. Sci.* 8, 250–251.
- Kassin, S. M. (2008). False confessions, causes, consequences, and implications for reform. *Curr. Dir. Psychol. Sci.* 17, 249–253.
- Kassin, S. M., and Gudjonsson, G. H. (2004). The psychology of confessions: a review of the literature and issues. *Psychol. Sci. Public Interest* 5, 33–67. doi: 10.1111/j.1529-1006.2004.00016.x
- Kassin, S. M., and Kiechel, K. L. (1996). The social psychology of false confessions: compliance, internalization, and confabulation. *Psychol. Sci.* 7, 125–128. doi: 10.1111/j.1467-9280.1996.tb00344.x
- Kassin, S. M., and Wrightsman, L. S. (1985). "Confession evidence," in *The Psychology Of Evidence And Trial Procedure*, eds S. Kassin and L. Wrightsman (Beverly Hills, CA: Sage), 67–94.
- Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychol. Rev.* 1, 602–609.
- Koustaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., and Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: testing the semantic categorization account. *J. Exp. Psychol.* 29, 499–510. doi: 10.1037/0278-7393.29.4.499
- Kurkela, K. A., and Dennis, N. A. (2016). Event-related fMRI studies of false memory: an activation likelihood estimation meta-analysis. *Neuropsychologia* 81, 149–167. doi: 10.1016/j.neuropsychologia.2015.12.006
- Kuvas, B., and Kaufmann, G. (2004). Impact of mood, framing, and need for cognition on decision makers' recall and confidence. *J. Behav. Decision Making* 17, 59–74. doi: 10.1002/bdm.461
- Leding, J. K. (2011). Need for cognition and false recall. *Pers. Individ. Dif.* 51, 68–72.
- Li, F., Zhu, H., Xu, J., Gao, Q., Guo, H., Wu, S., et al. (2018). Lie detection using fNIRS monitoring of inhibition-related brain regions discriminates infrequent but not frequent liars. *Front. Hum. Neurosci.* 12:71. doi: 10.3389/fnhum.2018.00071
- Lin, X. A., Wang, C., Zhou, J., Sai, L., and Fu, G. (2021). Neural correlates of spontaneous deception in a non-competitive interpersonal scenario: a functional near-infrared spectroscopy (fNIRS) study. *Brain Cogn.* 150:105704. doi: 10.1016/j.bandc.2021.105704
- Loftus, E. F. (1993). The reality of repressed memories. *Am. Psychol.* 48, 518–537. doi: 10.1037/0003-066X.48.5.518
- Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn. Mem.* 12, 361–366. doi: 10.1101/lm.9470
- Marche, T. A., Brainerd, C. J., and Reyna, V. F. (2010). Distinguishing true from false memories in forensic contexts: can phenomenology tell us what is real? *Appl. Cogn. Psychol.* 24, 1168–1182. doi: 10.1002/acp.1629
- Martens, V. E. G., and de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain Lang.* 98, 140–149. doi: 10.1016/j.bandl.2006.04.003
- Matzen, L. E., and Benjamin, A. S. (2013). Older and wiser: older adults' episodic word memory benefits from sentence study contexts. *Psychol. Aging* 28, 754–767. doi: 10.1037/a0032945
- Meyer, D. E., and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* 90, 227–234. doi: 10.1037/h0031564

- Miberg, W., and Blumstein, S. E. (1981). Lexical decision and aphasia: evidence for semantic processing. *Brain Lang.* 14, 371–385.
- Mitchell, K. J., and Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychol. Bull.* 135, 638–677. doi: 10.1037/a0015849
- Munsterberg, H. (1908). *On The Witness Stand: Essays On Psychology And Crime*. New York, NY: McClure.
- Neuschatz, J. S., Lampinen, J. M., Preston, E. L., Hawkins, E. R., and Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Appl. Cogn. Psychol.* 16, 687–708.
- Neuschatz, J. S., Payne, D. G., Lampinen, J. M., and Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory* 9, 53–71. doi: 10.1080/09658210042000076
- Ober, B. A., and Sheno, G. K. (1988). Lexical decision and priming in Alzheimer's disease. *Neuropsychologia* 26, 273–286. doi: 10.1016/0028-3932(88)90080-2
- Otgaar, H., and Baker, A. (2018). When lying changes memory for the truth. *Memory* 26, 2–14. doi: 10.1080/09658211.2017.134028
- Pimentel, E., and Albuquerque, P. B. (2013). Effect of divided attention on the production of false memories in the DRM paradigm: a study of dichotic listening and shadowing. *Psicologica* 34, 285–298.
- Pinti, P., Devoto, A., Greenhalgh, I., Tachtsidis, I., Burgess, P. W., de, C., et al. (2021). The role of anterior prefrontal cortex (area 10) in face-to-face deception measured with fNIRS. *Soc. Cogn. Affect. Neurosci.* 16, 129–142. doi: 10.1093/scan/nsaa086
- Plaut, D. C., and Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychol. Rev.* 107, 786–823. doi: 10.1037/0033-295x.107.4.786
- Prohaska, V., DelValle, D., Toglia, M. P., and Pittman, A. E. (2016). Reported serial positions of true and illusory memories in the Deese/Roediger/McDermott paradigm. *Memory* 24, 865–883. doi: 10.1080/09658211.2015.1059455
- Reyna, V. F., and Brainerd, C. J. (1995). Fuzzy trace theory: an interim synthesis. *Learn. Individ. Dif.* 7, 1–75. doi: 10.1016/1041-6080(95)90031-4
- Reyna, V. F., and Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: effects of lexical familiarity, semantic content, encoding instruction, and retention interval. *Dev. Psychol.* 30, 178–191. doi: 10.1037/0012-1649.30.2.178
- Reyna, V. F., and Mills, B. A. (2007). “Interference processes in fuzzy-trace theory: aging, Alzheimer's disease, and development,” in *Inhibition In Cognition*, eds D. S. Gorfein and C. M. MacLeod (Washington, DC: American Psychological Association), 185–210.
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Roediger, H. L., Watson, J. M., McDermott, K. B., and Gallo, D. A. (2001). Factors that determine false recall: a multiple regression analysis. *Psychon. Bull. Rev.* 8, 385–407. doi: 10.3758/bf03196177
- Schacter, D. L., and Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 773–786. doi: 10.1098/rstb.2007.2087
- Schacter, D. L., Reiman, E., Curran, T., Yun, L. S., Bandy, D., McDermott, K. B., et al. (1996). Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. *Neuron* 17, 267–274. doi: 10.1016/s0896-6273(00)80158-0
- Scholkemann, F., and Wolf, M. (2013). General equation for the differential pathlength factor of the frontal human head depending on wavelength and age. *J. Biomed. Optics* 18:105004. doi: 10.1117/1.JBO.18.10.105004
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., et al. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Mem. Cogn.* 30, 1054–1064. doi: 10.3758/BF03194323
- Shaw, J. (2020). Do false memories look real? Evidence that people struggle to identify rich false memories of committing crime and other emotional events. *Front. Psychol.* 8:650. doi: 10.3389/fpsyg.2020.00650
- Sporer, S. L., Penrod, S., Read, D., and Cutler, B. (1995). Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychol. Bull.* 118:315.
- Stadler, M. A., Roediger, H. L., and McDermott, K. B. (1999). Norms for word lists that create false memories. *Mem. Cogn.* 27, 494–500. doi: 10.3758/BF03211543
- 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
- Tian, F., Sharma, V., Kozel, F. A., and Liu, H. (2009). Functional near-infrared spectroscopy to investigate hemodynamic responses to deception in the prefrontal cortex. *Brain Res.* 1303, 120–130. doi: 10.1016/j.brainres.2009.09.085
- Toglia, M. P., Neuschatz, J. S., and Goodwin, K. A. (1999). Recall accuracy and illusory memories: when more is less. *Memory* 7, 233–256. doi: 10.1080/741944069
- Vega, R., Hernandez-Reynoso, A. G., Linn, E. K., Fuentes-Aguilar, R. Q., Sanchez-Ante, G., Santos-Garcia, A., et al. (2016). Hemodynamic pattern recognition during deception process using functional near-infrared spectroscopy. *J. Med. Biol. Eng.* 36, 22–31. doi: 10.1007/s40846-016-0103-6
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265. doi: 10.1007/s10979-007-9103-y
- Vrij, A., Semin, G. R., and Bull, R. (1996). Insight into behavior displayed during deception. *Hum. Commun. Res.* 22, 544–562.
- Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274. doi: 10.3758/cabn.3.4.255
- Walczyk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Appl. Cognit. Psychol.* 17, 755–774. doi: 10.1002/acp.914
- Wang, T. H., Johnson, J. D., de Chastelaine, M., Donley, B. E., and Rugg, M. D. (2016). The effects of age on the neural correlates of recollection success, recollection-related cortical reinstatement, and post-retrieval monitoring. *Cereb. cortex (New York, N.Y. : 1991)* 26, 1698–1714. doi: 10.1093/cercor/bhu333
- Yu, J., Tao, Q., Zhang, R., Chan, C., and Lee, T. (2019). Can fMRI discriminate between deception and false memory? A meta-analytic comparison between deception and false memory studies. *Neurosci. Biobehav. Rev.* 104, 43–55. doi: 10.1016/j.neubiorev.2019.06.027
- Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). “Verbal and nonverbal communication of deception,” in *Advances In Experimental Social Psychology*, Vol. 14, ed. L. Berkowitz (New York, NY: Academic Press), 1–57.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Toglia, Schuller, Surprenant, Hooper, DeMeo and Wallace. 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) and the copyright owner(s) 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
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership