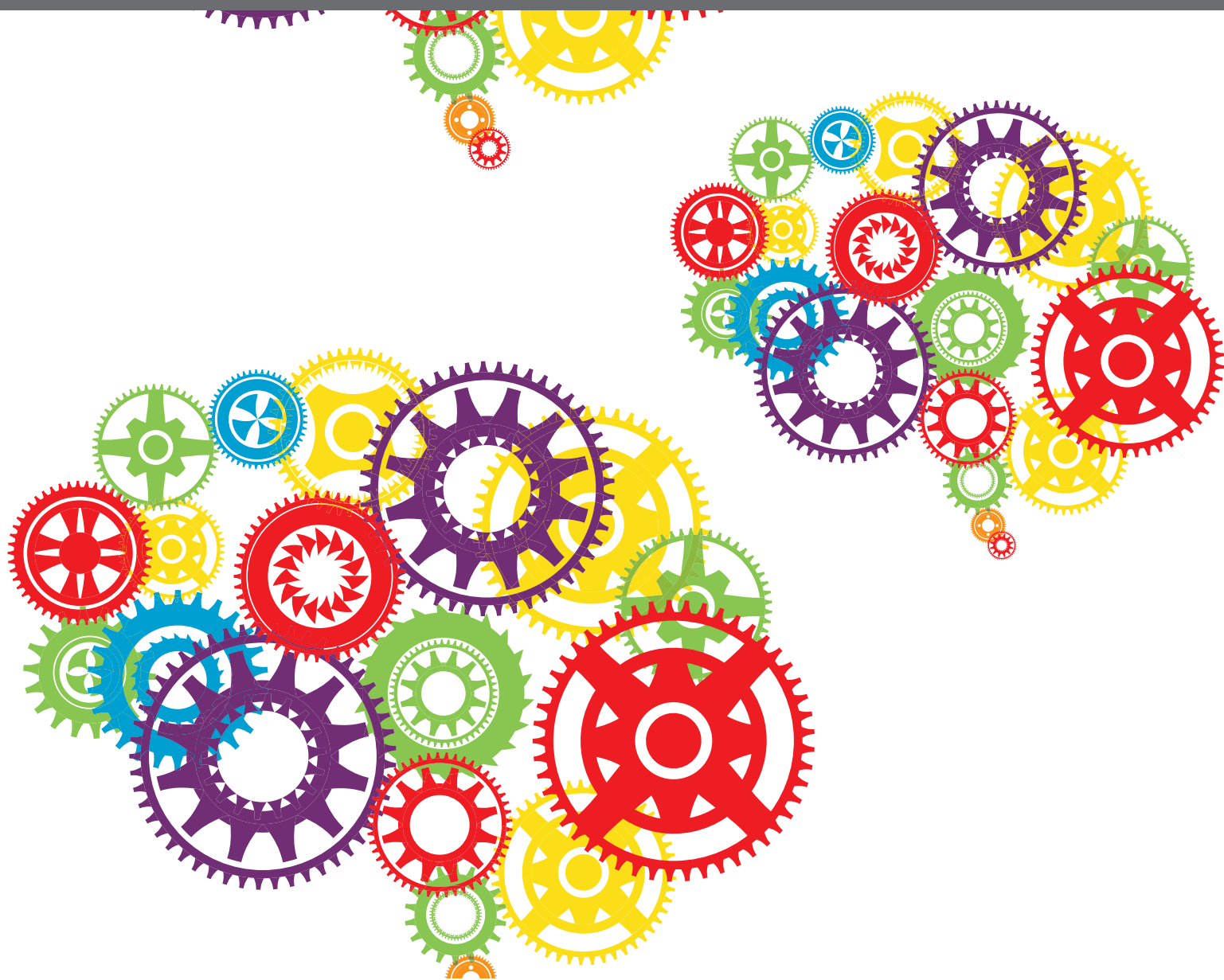




# **LANGUAGE AND MEMORY: UNDERSTANDING THEIR INTERACTIONS, INTERDEPENDENCIES, AND SHARED MECHANISMS**

EDITED BY: Melissa Duff and Vitória Piai

PUBLISHED IN: Frontiers in Human Neuroscience





# 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 978-2-88966-121-3

DOI 10.3389/978-2-88966-121-3

## About Frontiers

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

## Frontiers Journal Series

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

## Dedication to Quality

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

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

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



# LANGUAGE AND MEMORY: UNDERSTANDING THEIR INTERACTIONS, INTERDEPENDENCIES, AND SHARED MECHANISMS

Topic Editors:

**Melissa Duff**, Vanderbilt University Medical Center, United States

**Vitória Piai**, Radboud University Nijmegen, Netherlands

Language and memory have historically been studied apart, as unique cognitive abilities, and with distinct research traditions and methods. Over the past several decades, however, a growing body of evidence suggests that language and memory are heavily intertwined and may even rely on shared cognitive and neural mechanisms. Cutting across theoretical and methodological approaches, these findings offer novel insights into the interactions and interdependencies of language and memory. These advances also have considerable theoretical and clinical implications for the neurobiology of language and memory, their development, representation, and maintenance across the lifespan, the intervention and rehabilitation of disorders of language and memory, and the evolution of these two quintessential human abilities.

**Citation:** Duff, M., Piai, V., eds. (2020). Language and Memory: Understanding Their Interactions, Interdependencies, and Shared Mechanisms. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-121-3

# Table of Contents

- 05**    *Language, Memory, and Mental Time Travel: An Evolutionary Perspective*  
Michael C. Corballis
- 14**    *Downstream Behavioral and Electrophysiological Consequences of Word Prediction on Recognition Memory*  
Ryan J. Hubbard, Joost Rommers, Cassandra L. Jacobs and Kara D. Federmeier
- 27**    *How In-Group Bias Influences Source Memory for Words Learned From In-Group and Out-Group Speakers*  
Sara Iacozza, Antje S. Meyer and Shiri Lev-Ari
- 39**    *Searching for Semantic Knowledge: A Vector Space Semantic Analysis of the Feature Generation Task*  
Rebecca A. Cutler, Melissa C. Duff and Sean M. Polyn
- 52**    *Adult Age Differences in the Use of Conceptual Combination as an Associative Encoding Strategy*  
Heather D. Lucas, Resh S. Gupta, Ryan J. Hubbard and Kara D. Federmeier
- 69**    *Naming and Knowing Revisited: Eyetracking Correlates of Anomia in Progressive Aphasia*  
Molly B. Ungrady, Maurice Flurie, Bonnie M. Zuckerman, Daniel Mirman and Jamie Reilly
- 80**    *Input Complexity Affects Long-Term Retention of Statistically Learned Regularities in an Artificial Language Learning Task*  
Ethan Jost, Katherine Brill-Schuetz, Kara Morgan-Short and Morten H. Christiansen
- 93**    *Do Changes in Language Context Affect Visual Memory in Bilinguals?*  
Scott R. Schroeder
- 103**   *Bridging the Gap Between Second Language Acquisition Research and Memory Science: The Case of Foreign Language Attrition*  
Anne Mickan, James M. McQueen and Kristin Lemhöfer
- 111**   *Distinguishing Old From New Referents During Discourse Comprehension: Evidence From ERPs and Oscillations*  
Mante S. Nieuwland, Cas W. Coopmans and Rowan P. Sommers
- 135**   *Individual Differences in Verb Bias Sensitivity in Children and Adults With Developmental Language Disorder*  
Jessica E. Hall, Amanda Owen Van Horne and Thomas A. Farmer
- 154**   *Maintenance Versus Transmission Deficits: The Effect of Delay on Naming Performance in Aphasia*  
Nadine Martin and Gary S. Dell
- 167**   *Distinct Neural Processes for Memorizing Form and Meaning Within Sentences*  
Matteo Mascelloni, Roberto Zamparelli, Francesco Vespignani, Thomas Gruber and Jutta L. Mueller

- 185** *Cross-Situational Statistical Learning of New Words Despite Bilateral Hippocampal Damage and Severe Amnesia*  
David E. Warren, Tanja C. Roembke, Natalie V. Covington, Bob McMurray and Melissa C. Duff
- 198** *Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship*  
Melissa C. Duff, Natalie V. Covington, Caitlin Hilverman and Neal J. Cohen
- 215** *Verbal Working Memory as Emergent From Language Comprehension and Production*  
Steven C. Schwering and Maryellen C. MacDonald
- 234** *Post-training Load-Related Changes of Auditory Working Memory – An EEG Study*  
Helene Gudi-Mindermann, Johanna M. Rimmele, Patrick Bruns, Niels A. Kloosterman, Tobias H. Donner, Andreas K. Engel and Brigitte Röder
- 250** *Better Phonological Short-Term Memory is Linked to Improved Cortical Memory Representations for Word Forms and Better Word Learning*  
Sari Ylinen, Anni Nora and Elisabet Service



# Language, Memory, and Mental Time Travel: An Evolutionary Perspective

Michael C. Corballis\*

School of Psychology, Faculty of Science, University of Auckland, Auckland, New Zealand

Language could not exist without memory, in all its forms: working memory for sequential production and understanding, implicit memory for grammatical rules, semantic memory for knowledge, and episodic memory for communicating personal experience. Episodic memory is part of a more general capacity for mental travel both forward and backward in time, and extending even into fantasy and stories. I argue that the generativity of mental time travel underlies the generativity of language itself, and could be the basis of what Chomsky calls I-language, or universal grammar (UG), a capacity for recursive thought independent of communicative language itself. Whereas Chomsky proposed that I-language evolved in a single step well after the emergence of *Homo sapiens*, I suggest that generative imagination, extended in space and time, has a long evolutionary history, and that it was the capacity to share internal thoughts, rather than the nature of the thoughts themselves, that more clearly distinguishes humans from other species.

**Keywords:** displacement, evolution, externalization, gesture, imagination, memory, mental time travel, universal grammar

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

David E. Warren,  
University of Nebraska Medical  
Center, United States  
Thanujeni Pathman,  
York University, Canada

### \*Correspondence:

Michael C. Corballis  
m.corballis@auckland.ac.nz

**Received:** 21 February 2019

**Accepted:** 14 June 2019

**Published:** 04 July 2019

### Citation:

Corballis MC (2019) Language,  
Memory, and Mental Time Travel: An  
Evolutionary Perspective.  
*Front. Hum. Neurosci.* 13:217.  
doi: 10.3389/fnhum.2019.00217

## INTRODUCTION

Memory, in all its forms, is critical to language. Because language is sequential, we need short-term memory (working memory) as a moving window of consciousness if we are to integrate over time to make sense of sentences, and indeed stories. Long-term memory is itself divided into several components, each also serving a necessary function in linguistic communication. First is the distinction between unconscious and conscious memory. The rules of language are in large part overlearned and unconscious, and even linguists have not completely articulated how those rules work. They operate largely automatically; we know intuitively how to construct a sentence, but do not really know how we do it. Conscious memory is sometimes also referred to as declarative memory, or memory that can be declared. If part of memory is declarative memory, so part of language is memorial declaration.

Conscious memory can, in turn, be divided into semantic memory, or basic knowledge, and episodic memory, which is memory for personal episodes. Broadly speaking, semantic memory is a combined internal dictionary and encyclopedia, while episodic memory is an internal diary that records personal experiences (Tulving, 1972). Language draws on both. Semantic memory includes the large data bank of the tens of thousands of words that we use to express our thoughts, as well as providing kinds of knowledge that we can and do talk about—the political situation, the history of Ireland, differential calculus. It is episodic memory, though, that gives language many of its most distinctive properties.



Episodic memory is part of the more general capacity for *mental time travel*, a term probably first used by Tulving (1985) and elaborated by (Suddendorf and Corballis, 1997, 2007). We can travel mentally into a personal future as well as a personal past, and even create purely fictional events that need have no reference to specific time (“Once upon a time”). These are constructive acts—even episodic memory itself is better regarded as a construction than as a replay, and not always accurate. As Neisser (2008) put it, “Remembering is not like playing back a tape or looking at a picture; it is more like telling a story” (p. 88). Mental time travel is in turn founded on the understanding of space and time, with events encoded according to *what* happened, *where* it happened, and *when* it happened (the *www* criterion; Suddendorf and Corballis, 2007). I argue in this article that mental time travel provides the basis for the generative and creative aspects of language, allowing us to communicate about past and future, and indeed tell stories that need have no basis in reality.

Language, whether spoken or signed, can then be considered a device by which we share our mental travels—as Dor (2015) put it, it allows “the instruction of imagination.” Indeed, the recursive, generative nature of language may itself derive, not from the structure of language itself, but from the structure of the imaginative thoughts that underlie it.

## MENTAL TIME TRAVEL AND UNIVERSAL GRAMMAR

This view has some connection to the approach to language known as the Minimalist Program (Chomsky, 1995, 2015), but it also differs in important ways. A central tenet of the Minimalist Program is that language is structured by *universal grammar* (UG), which is common to all peoples. UG is the primary component of *I-language*, where the “I” is taken to suggest “internal,” “individual,” and “intensional.” Its main property is *merge*, a recursive operation that allows elements to be combined, and the mergers themselves to be merged, in a progressive fashion to build structures of any desired degree of complexity. The notion of UG has been criticized on the grounds that the 6,000 or so languages of the world have diverse grammars, and do not seem to conform to an overriding grammatical structure, forcing one commentary to conclude that “the emperor of UG has no clothes” (Evans and Levinson, 2009).

In his preface to the most recent edition of *The Minimalist Program*, though, Chomsky (2015) makes clear his view that UG is fundamentally a property, not of communicative language itself, but rather of thought, and is only incidental to communication. He writes: “It is a familiar fact (*sic*) that the complexity and variety of language appears to be localized overwhelmingly—and perhaps completely—in externalization (p. xi),” where “externalization” refers to the formation of specific languages from the underlying I-language. By extricating UG from communicative language itself, Chomsky appears to have sidestepped the problem of linguistic diversity. He also suggests that UG arose in a narrow window of time, shortly before the exodus of our species from Africa 50,000 to 80,000 years ago—a view endorsed by a number

of anthropologists (e.g., Hoffecker, 2007; Tattersall, 2012). Chomsky (2010) even suggests that the emergence of the operation merge occurred in a single individual whom he whimsically names “Prometheus.”

By reducing the essence of UG to the single operation of merge, Berwick and Chomsky (2016) claim also to have overcome the seemingly intractable problem of how a faculty as complex as language could have evolved in a single step, in defiance of Darwinian evolution. As they put it, “. . . narrowly focusing the phenotype in this way greatly eases the explanatory burden for evolutionary theory—we simply don’t have as much to explain, reducing the Darwinian paradox” (p. 11). They go on to write, though, that “Any residue of principles of language not reducible to Merge will have to be accounted for by some other evolutionary processes—one that we are unlikely to learn much about, at least by presently understood methods . . .” (p. 71); and they insist that “there is no room in this picture for any precursor to language” (p. 71).

My suggestion here, though, is that the recursive, generative nature of language may reside, not in a specialized I-language or UG, but in mental time travel itself, or more generally in our capacity to entertain thoughts not tied to the present. Such thoughts are the essence of imagination, defined by the Merriam-Webster Dictionary as “the act or power of forming a mental *image* of something not present to the senses or never before wholly perceived in reality.” Imaginative thoughts carry the generativity and recursiveness exemplified in our reconstructions of the past, in mental anticipations of the future, and perhaps most commonly in the fabrication of stories (McBride, 2014; Boyd, 2009). In providing the means to communicate such events, language requires the property of *displacement*, the capacity to refer to the non-present (Hockett, 1960), and arguably the most important driver of its evolution. Again, though, this capacity may reside not in language itself, but rather in the imaginative construction of mental events.

## UNIQUELY HUMAN?

Tulving (2002) view on the emergence of episodic memory echoes Chomsky’s account of the late arrival of UG itself:

*Many nonhuman animals, especially mammals and birds, possess well-developed knowledge-of-the-world (declarative, or semantic, memory) systems and are capable of acquiring vast amounts of flexibly expressible information. Early humans were like these animals, but at some point in human evolution, possibly rather recently, episodic memory emerged as an “embellishment” of the semantic memory system (p. 7).*

By extension, mental time travel has also been attributed uniquely to humans and denied to all other species (Suddendorf and Corballis, 1997, 2007).

More recently, I have argued that, on the contrary, the origins of mental time travel may go far back in evolution (Corballis, 2013; but see also Suddendorf, 2013). This change of opinion is based partly on behavioral evidence for mental

time travel in a diverse range of species, including great apes (Martin-Ordas et al., 2010; Beran et al., 2012; Janmaat et al., 2014), meadow voles (Ferkin et al., 2008), rats (Wilson et al., 2013), ravens (Kabadayi and Osvath, 2017), scrub jays (Clayton et al., 2003), and even cuttlefish (Jozet-Alves et al., 2013). In one recent study, rats remembered many different episodes over intervals of up to 45 min without any evidence of decline in performance (Panoz-Brown et al., 2016).

## Role of the Hippocampus

Evidence also comes from neuroscience, much of it focused on the hippocampus, and on parallels between human and animal hippocampal function. In humans, the hippocampus plays a critical role in declarative memory, including episodic memory and its extension to episodic future thinking. People with destruction of the hippocampus show striking difficulties in recalling past events or imagining future ones (Tulving, 2002; Wearing, 2005; Corkin, 2013), as well as in imagining fictitious scenes (Hassabis et al., 2007)—although impairment of the ability to imagine personal past or future events has also been linked to damage of the ventromedial prefrontal cortex (Bertossi et al., 2016).

Brain imaging confirms the role of the hippocampus when people are asked to recall previous episodes or to imagine future ones (Addis et al., 2011; Martin et al., 2011). Again, though, areas other than the hippocampus are also active, including the angular gyrus, the medial frontal cortex, and the posterior cingulate (Rugg and Vilberg, 2013; Karapanagiotidis et al., 2017). The particular role of the hippocampus may lie in what has been termed *scene construction* (Maguire et al., 2016), the drawing together of dispersed information for autonoetic inspection. McCormick et al. (2018) suggest that hippocampal function goes beyond mental time travel to mind-wandering more generally, and lies at “the heart of mental life” (p. 2745).

In the rat, the hippocampus is well known to play a role in spatial location. So-called “place cells” record the animal’s location in space, creating a “cognitive map” (O’Keefe and Nadel, 1978)—or a kind of internal GPS system. The population of active cells shifts as the animal moves around, recording a trajectory. It has become clear, though, that the activity of place cells is not restricted to the present, but can convey information about past trajectories or even trajectories that it did not take, perhaps representing future plans or simply exploratory movements. Such trajectories have been described as “replays,” although in many cases they might be better described as preplays or mental explorations not specifically located in time. Reviewing the evidence, Moser et al. (2015) write that:

*“the replay phenomenon may support ‘mental time travel’ ... through the spatial map, both forward and backward in time (p. 6).”*

Hippocampal activity, in conjunction with the neighboring entorhinal cortex, is also tagged in other memory-like ways. Place cells respond not only to specific locations, but also to nonspatial features of past events, such as odors (Igarashi et al., 2014), touch sensations, and the timing of events. Similar associations seem to be tagged to place cells in the human hippocampus. In one

study, human patients about to undergo surgery had electrodes implanted in cells in the medial temporal lobe, in an attempt to locate the source of epileptic seizures. They were given the task of navigating a virtual town on a computer screen and delivering items to one of the stores in town. They were then asked to recall only the items and not the location to which they were delivered. The act of recall, though, activated place cells corresponding to that location, effectively mirroring the replay of place-cell activity in the rat brain (Miller et al., 2013). In a similar study using subdural electrodes, Vaz et al. (2019) found that oscillatory activity between the medial temporal lobe and the temporal associative cortex were coupled when people retrieved memories of associated items.

The spatial function of the hippocampus is modulated by activity in the neighboring entorhinal cortex. So-called grid cells in the medial entorhinal cortex code locations corresponding to spatial features such as spatial scale and orientation, and other cells code shape and color, proximity to borders, and direction in which the head is facing (Diehl et al., 2017). These cells operate in a modular fashion, creating an enormous number of combinations reflecting the possible spatial contexts in which an animal may find itself. Moser et al. (2015) liken this to “an alphabet in which all words of a language can be generated by combining only 30 letters or less” (p. 11). This is suggestive of the generativity of language itself.

Recordings from the rat hippocampus also reveal what has been termed “time cells,” which respond in a coordinated fashion to code the relative times in which events have occurred in the past. The pattern itself changes over time as the temporal context changes (Eichenbaum, 2017). This can be observed experientially in our own memories of when things happened, gradually losing immediacy and detail, both spatial and temporal. The hippocampal coding of space, time and context in both humans and animals suggest that episodic mental travel may long predate human evolution.

The coding of episodic memories can be specified in time rather than space, and need not be visual. We might mentally replay a memory of a concert, but the ordering of individual pieces is not marked by different locations. A similar phenomenon has been reported in rats, based on their fine discrimination of different odors. Panoz-Brown et al. (2018) presented rats with sequences of specific odors in different contexts. Later, when presented with one given context, they were able to select the second from the last odor in the sequence as distinct from a different odor from the sequence, while given a different context they were able to select the fourth from the last odor in the sequence. The number of odors in the sequences varied from trial to trial, making it impossible to specify the required odor when it occurred. The animals must have held the entire sequence in memory and replayed it in order to select the required odor. Performance was well above chance even after the lapse of an hour between presentation and test and was little affected by interference. Performance dropped significantly with chemical suppression of hippocampal activity. These properties imply robust hippocampal-dependent episodic memory for sequences of events defined by the order in which they occurred and not by locations within sequences,

although the retrieval of the sequences themselves depended on spatial context.

The evidence for mental time travels in nonhuman animal raises the question of whether they are conscious. To Tulving, episodic memories are what he called *autonoetic*, or part of personal consciousness, and the same might be said of mental time travel more generally. If such travels are not exclusive to humans, contrary to what Tulving believed, can we conclude that animals too are conscious of their mental time travels? The commonality between what we know of the role of the hippocampus through electrophysiology in animals and through brain imaging, and indeed through cases of implanted electrodes in humans, seems to give little reason to doubt that in both cases the experience is conscious. Nevertheless, this is likely to remain a contentious issue.

The role of the hippocampus is not restricted to episodic information, but includes semantic information as well—indeed the replay of the past and prediction of the future is probably always a mix of the episodic and the semantic (Klein, 2013). Duff and Brown-Schmidt (2012) review evidence from studies of hippocampal amnesia that the hippocampus is critical to language itself, in binding information from different sources and supplying a flexibility of operation. Piai et al. (2016) add evidence from recording of hippocampal theta during sentence processing, and suggest that the hippocampus should be considered part of the language network, a conclusion endorsed by Covington and Duff (2016). Individuals with large-scale destruction of the hippocampus can retain the basic ability to speak, but loss of episodic memory, and of mental time travel more generally, severely restricts communicative content (Wearing, 2005; Corkin, 2013), and word learning becomes sparse and slow (Warren and Duff, 2019). The hippocampus not only contributes to the generative and integrative aspects of language, but also provides for displacement, the power of language to refer to events removed from the present in time and space.

## Expansions of Scale

A good deal of human language has to do with events or material far displaced from the present; we can tell of events from childhood, experiences in far-away places, or plans for a distant future. This suggests that mental time travel itself may have expanded in scale beyond that evident in other species, and indeed this expansion may have partly driven the evolution of language itself—although such a claim may well simply reflect what has been called the “human superiority complex” (Villa and Roebroeks, 2014, p. 1). Many animals and birds do appear to have extensive understanding of space. Dolins et al. (2014) assessed the ability of humans and chimpanzees to learn complex virtual environments and navigate through them, and found chimpanzees generally on the same level as children, and one chimpanzee (Panzee) was more accurate than human adults. Nevertheless, it is likely that the capacity for mental excursions probably expanded in both time and space, and indeed content, over the past six million years or so with the emergence of the hominins, and especially the genus *Homo*. Humans probably have

more extensive memories, plans, and fantasies than do rats and chimpanzees.

In human evolution, a critical period for such an expansion, and indeed for the pressure to communicate about it, was probably the Pleistocene, dating from some 2.8 million to 12,000 years ago, when our forebears adapted to a post-arboreal existence, with an emergent hunter-gatherer pattern. This resulted in long delays between the acquisition and the use of tools, as well as geographical distance between the sources of raw material for tools and killing or butchering sites (Gärdenfors and Osvath, 2010). The hunter-gatherer lifestyle involved frequent shifts of camp as resources were depleted, forcing the group to move on to another more abundant region—a pattern still evident in present-day hunter-gatherers (Venkataraman et al., 2017).

Migrations increased in scale during the Pleistocene, adding further to the demands of space, time, memory, and planning, and brain size also tripled during this era (Klein, 2009). The dispersals of early *Homo* from Africa reached the Loess Plateau in China by 2.1 million years ago (Zhu et al., 2018), and other widespread regions in Europe and Asia in the previous millennium (Kappelman, 2018). Later waves of migration of *Homo sapiens* out of Africa began from about 120,000 years ago (Timmermann and Friedrich, 2016), eventually inhabiting most of the globe. Of course, humans are not entirely alone in undertaking large-scale migrations. Birds, whales, wildebeest, and even butterflies migrate vast distances, but these are largely seasonal (as are some human migrations, especially of the wealthy) and based on instinct rather than planning. The Clark’s nutcracker is said to cache some 33,000 seeds in around 7,000 locations every fall and relies on spatial memory to recover them over the winter (Kamil and Balda, 1985). Evidence from scrub jays, moreover, suggests that caching behavior involves mental time travel both forward and backward in time (Clayton et al., 2003). Even so, the human ability to recapture the past and imagine the future, at least with respect to time and flexibility, probably exceeds that of any other living animal.

Perhaps the ultimate stretch is the ability to imagine events outside of the lifespan, although this is a matter of semantic rather than episodic time travel—or what Klein et al. (2010) call *known time* as distinct from *lived time*. Historical records have allowed us to create stories and movies reconstructing events long in the past, and even to imagine ourselves as spectators. Physicists have even dared to envisage the origins of the universe. We also imagine life after death. Pettitt (2018) notes that even chimpanzees follow certain *mortuary behaviors* on finding a dead conspecific, including staying by the body for many hours, giving alarm calls, and showing signs of grief, as though aware of the permanence of death. The parallels with observation of human reactions to death, he suggests, “are striking” (p. 6). In humans, this is further transformed into burial and rituals associated with it, and in the modern world most of these rituals seem to have to do with “transforming the deceased into some form of afterlife” (p. 6). Evidence for the deliberate disposal of corpses, implying a sense of one’s own mortality, has been dated from around 600,000 to 300,000 years ago (Egeland et al., 2018).



## COMMUNICATION

These expansions in time and space no doubt added to the pressure to communicate, so that experiences not restricted to the immediate environment could be shared, and indeed make up much of what we call culture. Communication of our internal thoughts is what Chomsky (2015) called *externalization*. Again, the critical period of its development was probably the Pleistocene. Beginning with the hunter-gatherer phase, and extending to more complex modes of existence through the development of farming and manufacture, life depended increasingly on cooperation and the sharing of experience, plans, and mental exploration leading to stories. The main requirement for communicating internal mental events, though, was a signaling device capable of matching the generativity and complexity of experience itself. Most animals have only a limited range of systems permitting intentional output. Neurophysiology increasingly reveals the complexity of the rat's excursions in time and space, but the animal has no obvious way to convey those excursions to others. Songbirds are something of an exception, with often complex songs, but these seem adapted to sexual or identification signaling rather than to the sharing of memories or plans. They appear not well adapted to communicating about events.

Even non-human primates have very limited vocal repertoires, dedicated for the most part to instinctive or emotional calls. Seyfarth and Cheney (2018) identify different baboon calls signaling identity, social rank, kin and various social interactions, and go so far as to suggest that the sequences of calls between different individuals constitutes a system that is "*discrete, combinatorial, and rule-governed*" (p. 28, their italics), with the implication that it may be a precursor to grammatical language itself. But as Godfrey Smith (2018) points out, the combinatorial structure is evident in the interweaving of calls between individuals, and not in individual calls themselves.

The problem of communication is largely one of production rather than reception. The understanding of spoken words can actually be quite high in nonhuman animals. Border collies have been shown to respond to verbal requests to select a particular object from an otherwise uninhabited room and returns it to a given location. One border collie called Rico has a receptive vocabulary of over 200 items (Kaminski et al., 2004), while another is said to respond to the names of 1,022 objects (Pilleary and Reid, 2011). Kanzi, a bonobo, appears able to respond appropriately to simple spoken requests, such as "Could you carry the television outdoors?" or "Could you put the pine needles in the refrigerator?" (Savage-Rumbaugh et al., 1998).

None of these species, though, can speak. A fundamental problem is that most mammals and apes, with the exception of humans, have at best limited voluntary control over voicing. Chimpanzees seem to have some ability to modify emotional calls (e.g., Slocombe and Zuberbühler, 2005) but little evident capacity to produce or learn anything like spoken words, either in number or complexity. According to Petkov and Jarvis (2012), only parrots approach humans as "high vocal learners," with songbirds not far behind, while nonhuman primates are merely "limited vocal learners." The origins of communicative language

may lie in the production of visual signals, rather than vocal ones (Corballis, 2017).

Chimpanzees and bonobos trained to use lexigrams to refer to objects and actions are able to use these, along with gestures, to make requests and even to comment on past and future events, or on other individuals (Lyn et al., 2011). In one study, the chimpanzee Panzee, who uses a keyboard containing 256 lexigrams, watched an experimenter hide objects in the woods outside her enclosure. After imposed delays of up to 16 h, she interacted with a person who did not know that an object had been hidden, pointed to the lexigram representing the object, pointed outdoors, and led the person to where the object was hidden, continually pointing as she went (Menzel, 1999). There were 34 such trials, with different objects and locations. Panzee, therefore, seems capable not only of mental time travel, but also of displacement in her ability to communicate.

Chimpanzees in the wild gesture prolifically to each other, in an intentional fashion. Byrne et al. (2017) report evidence for repertoires of at least 66 natural gestures in the chimpanzee, 68 in bonobos, 102 in gorillas, and 64 in orangutans, considerably larger than repertoires of vocal calls. Many of those observed in the wild are common to the different species, suggesting that they are based on phylogeny rather than social learning, but they are also greatly augmented in the case of apes trained to use gestures or lexigrams. The gorilla Koko, for example, is said to use and understand over 1,000 signs (Patterson and Gordon, 2001).

Gestures are also more obviously intentional than are vocal calls, and are in that sense language-like, but they are more deictic than referential (Byrne et al., 2017), and occur in short sequences of seldom more than one or two, with no evidence of syntactic structure.

The ability to generate complex sequences probably emerged in human evolution with pressure to communicate about more complex events or plans. Given our ape physiognomy, a natural way to communicate mental time travels would be through pantomime, and apes do seem capable of limited pantomime. Russon and Andrews (2001) identified 18 different pantomimes produced by orangutans in a forest-living enclave in Indonesia, 14 addressed to humans and four to fellow orangutans. These included mimed offers of fruit, enacting a haircut, and requests to have their stomachs scratched by scratching their own stomachs and then offering a stick to the prospective scratcher. A chimpanzee in the wild watched her daughter trying to use a stone to crack a nut and then enacted the operation to show her how to do it properly (Boesch, 1993). Tanner and Perlman (2017) also note that gorillas combine gestures in sequence creatively and interactively, although this seems to have more to do with play and personal display than with propositional communication, and may be the origin of music and dance rather than of language itself. Nevertheless, it seems likely that language did emerge from primate gestures rather than vocal calls. Based on studies of gestural communication in apes, Tomasello (2008, p. 55) refers to gestures as "the original font from which the richness and complexities of human communication and language have flowed."

But it was probably during the Pleistocene, with the so-called "cognitive niche" (Pinker, 2010) as an adaptation to the more



dangerous and uncertain environment, that gesture, perhaps originally as pantomime, emerged as a powerful way to share episodic events, whether past, future, or simply invented. Donald (1991) referred to the “mimetic culture” of the early Pleistocene. Pantomime involves whole-body action to represent events, but the essence of an event in space and time could be relayed more economically just using gestures of the hands and arms, which were freed from any involvement in locomotion with the advent of bipedalism. Gestural language may well have developed to resemble modern sign languages invented by deaf communities. Emerging sign languages typically begin with pantomime, but signs are then conventionalized so that many no longer provide a pictorial indication of what they stand for (Burling, 1999). Conventionalization may be at the cost of transparency but leads to greater efficiency. On an evolutionary scale, speech itself may be the end product of a conventionalization process that began with pantomime, as our forebears gained great intentional control over voicing.

Nevertheless, bodily gesture remains an integral accompaniment to speech, even in the blind (Iverson and Goldin-Meadow, 1998). They can improve the speaker’s lexical access and fluency (Rauscher et al., 1996), and even reduce the speaker’s working memory load (Goldin-Meadow et al., 2001; Wagner et al., 2004). Some have gone so far as to suggest that manual gestures were in equal partnership with vocalization throughout the evolution of language (e.g., Kendon, 2011; McNeill, 2012), but the evidence from primates suggests that manual gestures preceded vocalization in the evolution of intentional communication (Corballis, 2014). It remains something of an open question when speech evolved to the level of articulation evident in *Homo sapiens*. It is possible, even likely, that the one of our closest forebears, the Neanderthals, were capable of speech (Dediu and Levinson, 2013), but their articulation was probably relatively more restricted through non-optimal development of the vocal tract (Gokhman et al., 2019).

## CONCLUSIONS

The main thesis of this article is that imagination, initially in the form of mental travels in time and space, provide the recursive and generative properties underlying language itself. These mental travels are extensions of episodic memory and make up much of what we call imagination. Unlike Chomsky’s concept of I-language, imagination probably has a long evolutionary history, as is becoming evident from behavioral studies in a wide variety species, along with work on the role of the hippocampus and related structures in rodents. Communicative language is then the externalization of imagination, and different languages use different conventions to express the products of imagination. This approach differs from that of Chomsky and colleagues also in that both imagination and its externalization have a strong evolutionary basis in spatial understanding and bodily movement, whereas I-language is regarded as abstract and symbolic, creating what is known as the problem of *grounding* (Harnad, 1990). How can a person relate abstract symbols to events in the real world?

In the view adopted here, symbolic representation arises in the process of externalization, rather than in innate symbolic dispositions. Internal representations of objects, actions and events are for the most part similar across different peoples (corresponding to the “universal” of UG), but the symbols to represent them differ markedly. Some 6,000 languages exist in the world, each more or less incomprehensible to almost every other. As suggested earlier, those symbols probably begin as iconic, or pantomimic, but become increasingly arbitrary and abstract in the process of conventionalization. This process increases efficiency, but may also be driven by exclusiveness, acting as a barricade to outsiders. Language operates in part as a secret code. Sign languages are more transparent, but even they conventionalize differently. Nevertheless, it is also becoming clear that many speech sounds are non-arbitrary, and show similar associations across language groups (Blasi et al., 2016).

The symbols that arise in the process of externalization themselves become part of our semantic memories; we can imagine the word “dog” as easily as we can imagine the animal with which it is associated. We can play with words just as we play with toys. The use of abstract symbols may well have influenced cognition itself. Mathematics may be an extreme example, in which abstraction has developed to the point that a single symbol, say  $x$  or  $y$ , can stand for variables of wide reference. But the invention of abstract symbols was not the outcome of some singular event in our evolutionary past, but was the product of gradual evolution, perhaps leading to increased powers of reasoning and discourse.

Given that words themselves have become part of memory, the emergence of language may well have expanded our capacity for mental time travels, and perhaps especially for the imaginative extensions into fiction and storytelling. The capacity to communicate our mental excursions vastly exceeds that required for personal experience alone. To accommodate the information added through communication, including not only speech, but also the vast repertoire of information through books, films, television, and so forth, storage capacity itself must surely have expanded. The link between language and memory might, therefore, be considered bidirectional.

From the Bible to Chomsky, the emergence of language has been regarded as a singular event, bestowed uniquely on our own species. The concept of time, too, is also widely viewed as uniquely human. Donald (1991), for example, wrote that “The lives of apes are lived entirely in the present” (p. 149), and much earlier Kohler (1927), based on his studies of problem solving in chimpanzees, wrote that “the time in which chimpanzees live is limited in past and future” (p. 272). The poet Robert Browning, in his 1885 poem “A Grammarian’s Funeral,” prophetically wrote:

*“He said, What’s time? Leave Now for dogs and apes  
Man has Forever!”*

Contrary to these commonly held views, the experience of past and future probably goes far back in the evolution of animals

that move, and need to know where they are, where they have been, and where they might go next—along with what happened or might happen there. The sharing of this information, though, probably evolved later, as our forebears were forced for survival into their cognitive niche.

One-hundred and sixty years after the publication of Darwin's (1859) *Origin of Species*, it is time to work toward an evolutionarily plausible understanding of how the human mind evolved.

## REFERENCES

- Addis, D. R., Cheng, T., Roberts, R., and Schacter, D. L. (2011). Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus* 21, 1045–1052. doi: 10.1002/hipo.20870
- Beran, M. J., Perdue, B. M., Bramlett, J. L., Menzel, C. R., and Evans, T. A. (2012). Prospective memory in a language-trained chimpanzee (*Pan troglodytes*). *Learn. Motiv.* 43, 192–199. doi: 10.1016/j.lmot.2012.05.002
- Bertossi, E., Tesini, C., Cappelli, A., and Ciaramelli, E. (2016). Ventromedial prefrontal damage causes a pervasive impairment of episodic memory and future thinking. *Neuropsychologia* 90, 12–24. doi: 10.1016/j.neuropsychologia.2016.01.034
- Berwick, R. C., and Chomsky, N. (2016). *Why Only Us? Language and Evolution*. Cambridge, MA: The MIT Press.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proc. Natl. Acad. Sci. U S A* 113, 10818–10823. doi: 10.1073/pnas.1605782113
- Boesch, C. (1993). "Aspects of transmission of tool-use in wild chimpanzees," in *Tools, Language and Cognition in Human Evolution*, eds K. R. Gibson and T. Ingold (Cambridge, UK: Cambridge University Press), 171–183.
- Boyd, B. (2009). *The Origin of Stories*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Burling, R. (1999). "Motivation, conventionalization and arbitrariness in the origin of language," in *The Origins of Language: What Human Primates Can Tell Us*, ed. B. J. King (Santa FE, NM: School of American Research Press), 307–350.
- Byrne, R. W., Cartmill, E., Genty, E., Hobaiter, C., and Tanner, J. E. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Anim. Cogn.* 20, 755–769. doi: 10.1007/s10071-017-1096-4
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2010). "Some simple evo devo theses: how true might they be for language?" in *The Evolution of Human Language*, eds R. K. Larson, V. Déprez and H. Yamakido (Cambridge: Cambridge University Press), 45–62.
- Chomsky, N. (2015). *The Minimalist Program: 20th Anniversary Edition*. Cambridge, MA: MIT Press.
- Clayton, N. S., Bussey, T. J., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* 4, 685–691. doi: 10.1038/nrn1180
- Corballis, M. C. (2013). Mental time travel: a case for evolutionary continuity. *Trends Cogn. Sci.* 17, 5–6. doi: 10.1016/j.tics.2012.10.009
- Corballis, M. C. (2014). "The word according to Adam: the role of gesture in language evolution," in *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, eds M. Seyfeddinifur and M. Gullberg (Amsterdam: John Benjamins), 177–197.
- Corballis, M. C. (2017). *The Truth About Language: What it is and Where it Came From*. Chicago, IL: University of Chicago Press.
- Corkin, S. (2013). *Permanent Present Tense: The Man With No Memory and What He Taught the World*. London: Allen Lane.
- Covington, N. V., and Duff, M. C. (2016). Expanding the language network: direct contributions from the hippocampus. *Trends Cogn. Sci.* 20, 869–870. doi: 10.1016/j.tics.2016.10.006
- Darwin, C. (1859). *On The Origin of Species by Means of Natural Selection*. London: John Murray.

## DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the supplementary files.

## AUTHOR CONTRIBUTIONS

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

- Dediu, D., and Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Front. Psychol.* 4:397. doi: 10.3389/fpsyg.2013.00397
- Diehl, G. W., Hon, O. J., Leutgeb, S., and Leutgeb, J. K. (2017). Grid and nongrid cells in medial entorhinal cortex represent spatial location and environmental features with complementary coding schemes. *Neuron* 94, 83–92. doi: 10.1016/j.neuron.2017.03.004
- Dolins, F. L., Klimowicz, C., Kelley, J., and Menzel, C. R. (2014). Using virtual reality to investigate comparative spatial cognitive abilities in chimpanzees and humans. *Am. J. Primatol.* 76, 496–513. doi: 10.1002/ajp.22252
- Donald, M. (1991). *Origins of the Modern Mind*. London: Harvard University Press.
- Dor, D. (2015). *The Instruction of Imagination: Language as a Social Communication Technology*. New York, NY: Oxford University Press.
- Duff, M. C., and Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Front. Hum. Neurosci.* 6:69. doi: 10.3389/fnhum.2012.00069
- Egeland, C. P., Domínguez-Rodrigo, M., Pickering, T. R., Menter, C. G., and Heaton, J. L. (2018). Hominin skeletal part abundances and claims of deliberate disposal of corpses in the middle Pleistocene. *Proc. Natl. Acad. Sci. U S A* 115, 4601–4606. doi: 10.1073/pnas.1718678115
- Eichenbaum, H. (2017). Time (and space) in the hippocampus. *Curr. Opin. Behav. Sci.* 17, 65–70. doi: 10.1016/j.cobeha.2017.06.010
- Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–492. doi: 10.1017/S0140525X0999094X
- Ferkin, M. H., Combs, A., delBarco-Trillo, J., Pierce, A. A., and Franklin, S. (2008). Meadow voles, *Microtus pennsylvanicus*, have the capacity to recall the "what", "where" and "when" of a single past event. *Anim. Cogn.* 11, 147–159. doi: 10.1007/s10071-007-0101-8
- Gärdenfors, P., and Osvath, M. (2010). "Prospection as a cognitive precursor to symbolic communication," in *The Evolution of Human Language*, eds R. K. Larson, V. Déprez and H. Yamakido (Cambridge: Cambridge University Press), 103–114.
- Godfrey Smith, P. (2018). "Primates, cephalopods and the evolution of communication," in *The Social Origins of Language*, eds R. M. Seyfarth and D. L. Cheney (Princeton, NJ: Princeton University Press), 102–120.
- Gokhman, D., Agranat-Tamir, L., Housman, G., García-Pérez, R., Nissim-Rafini, M., and Carmel, L. (2019). Extensive regulatory changes in genes affecting vocal and facial anatomy separate modern from archaic humans. *bioRxiv* [Preprint]. doi: 10.1101/106955
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395
- Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346. doi: 10.1016/0167-2789(90)90087-6
- Hassabis, D., Kumaran, D., and Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *J. Neurosci.* 27, 14365–14374. doi: 10.1523/JNEUROSCI.4549-07.2007
- Hockett, C. F. (1960). The origins of speech. *Sci. Am.* 203, 88–96. doi: 10.1038/scientificamerican0960-88
- Hoffecker, J. F. (2007). Representation and recursion in the archaeological record. *J. Archaeol. Method Theory* 14, 359–387. doi: 10.1007/s10816-007-9041-5

- Igarashi, K. M., Lu, L., Colgin, L. L., Moser, M.-B., and Moser, E. I. (2014). Coordination of entorhinal-hippocampal ensemble activity during associative learning. *Nature* 510, 143–147. doi: 10.1038/nature13162
- Iverson, J. M., and Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature* 396:228. doi: 10.1038/24300
- Janmaat, K. R. L., Polansky, L., Ban, S. D., and Boesch, C. (2014). Wild chimpanzees plan their breakfast time, type and location. *Proc. Natl. Acad. Sci. U S A* 111, 16343–16348. doi: 10.1073/pnas.1407524111
- Jozet-Alves, C., Bertin, M., and Clayton, N. S. (2013). Evidence of episodic-like memory in cuttlefish. *Curr. Biol.* 23, R1033–R1036. doi: 10.1016/j.cub.2013.10.021
- Kabadayi, C., and Osvath, M. (2017). Ravens parallel great apes in flexible planning for tool-use and bartering. *Science* 357, 202–204. doi: 10.1126/science.aam8138
- Kamil, A. C., and Balda, R. P. (1985). Cache recovery and spatial memory in Clark's Nutcrackers (*Nucifraga columbiana*). *J. Exp. Psychol. Anim. Behav. Process.* 85, 95–111.
- Kaminski, J., Call, J., and Fischer, J. (2004). Word learning in the domestic dog: evidence for “fast mapping”. *Science* 304, 1682–1683. doi: 10.1126/science.1097859
- Kappelman, J. (2018). An early hominin arrival in Asia. *Nature* 559, 480–481. doi: 10.1038/d41586-018-05293-9
- Karapanagiotidis, T., Bernhardt, B. C., Jefferies, E., and Smallwood, J. (2017). Tracking thoughts: exploring the neural architecture of mental time travel during mind-wandering. *Neuroimage* 147, 272–281. doi: 10.1016/j.neuroimage.2016.12.031
- Kendon, A. (2011). Vocalisation, speech, gesture and the language origins debate. *Gesture* 13, 349–370. doi: 10.1075/gest.11.3.05ken
- Klein, R. G. (2009). *The Human Career*. 3rd Edn. Chicago, IL: University of Chicago Press.
- Klein, S. B. (2013). Future mental time travel: types of memory, types of selves and types of temporality. *Soc. Cogn.* 31, 417–426. doi: 10.1521/soco\_20131009
- Klein, S. B., Robertson, T. E., and Delton, A. W. (2010). Facing the future: memory as an evolved system for planning future acts. *Mem. Cognit.* 38, 13–22. doi: 10.3758/MC.38.1.13
- Kohler, W. (1927). *The Mentality of Apes*, Trans. E. Winter. London: Routledge and Kegan Paul. (Original work published in German, 1917).
- Lyn, H., Greenfield, P. M., Savage-Rumbaugh, S., Gillespie-Lynch, K., and Hopkins, W. D. (2011). Nonhuman primates do declare! A comparison of declarative symbol and gesture use in two children, two bonobos and a chimpanzee. *Lang. Commun.* 31, 63–74. doi: 10.1016/j.langcom.2010.11.001
- Maguire, E. A., Intraub, H., and Mullaly, S. L. (2016). Scenes, spaces and memory traces: what does the hippocampus do? *Neuroscientist* 22, 432–439. doi: 10.1177/1073858415600389
- Martin, V. C., Schacter, D. L., Corballis, M. C., and Addis, D. R. (2011). A role for the hippocampus in encoding simulations of future events. *Proc. Natl. Acad. Sci. U S A* 108, 13858–13863. doi: 10.1073/pnas.1105816108
- Martin-Ordas, G., Haun, D., Colmenares, F., and Call, J. (2010). Keeping track of time: evidence for episodic-like memory in great apes. *Anim. Cogn.* 13, 331–340. doi: 10.1007/s10071-009-0282-4
- McBride, G. (2014). Story telling, behavior planning and language evolution in context. *Front. Psychol.* 5:1131. doi: 10.3389/fpsyg.2014.01131
- McCormick, C., Rosenthal, C. R., Miller, T. D., and Maguire, E. A. (2018). Mind-wandering in people with hippocampal damage. *J. Neurosci.* 38, 2745–2754. doi: 10.1523/JNEUROSCI.1812-17.2018
- McNeill, D. (2012). *How Language Began: Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press.
- Menzel, C. R. (1999). Unprompted recall and reporting of hidden objects by a chimpanzee (*Pan troglodytes*) after extended delays. *J. Comp. Psychol.* 113, 426–434. doi: 10.1037/0735-7036.113.4.426
- Merriam-Webster Dictionary. Available online at: <https://www.merriam-webster.com/dictionary/imagination>
- Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., and Schulze-Bonhage, A. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. *Science* 342, 1111–1114. doi: 10.1126/science.1244056
- Moser, M. B., Rowland, D. C., and Moser, E. I. (2015). Place cells, grid cells and memory. *Cold Spring Harb. Perspect. Biol.* 7:a021808. doi: 10.1101/cshperspect.a021808
- Neisser, U. (2008). “Memory with a grain of salt,” in *Memory: An Anthology*, eds H. H. Wood and A. S. Byatt (London: Chatto and Windus), 80–88.
- O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.
- Panoz-Brown, D., Corbin, H. E., Dalecki, S. J., Gentry, M., Brotheridge, S., and Crystal, J. D. (2016). Rats remember items in context using episodic memory. *Curr. Biol.* 26, 2821–2826. doi: 10.1016/j.cub.2016.08.023
- Panoz-Brown, D., Iyer, V., Carey, L. M., Sluka, C. M., Rajic, G., and Crystal, J. D. (2018). Replay of episodic memories in the rat. *Curr. Biol.* 28, 1628–1634. doi: 10.1016/j.cub.2018.04.006
- Patterson, F. G. P., and Gordon, W. (2001). “Twenty-seven years of project Koko and Michael,” in *All Apes Great and Small, Vol. 1: African Apes*, eds B. M. F. Galdikas, N. E. Briggs, L. K. Sheeran and J. Goodall (New York: Kluwer), 165–176.
- Petkov, C. I., and Jarvis, E. D. (2012). Birds, primates and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4:12. doi: 10.3389/fnevo.2012.00012
- Pettitt, P. (2018). Hominin evolutionary thanatology from the mortuary to funerary realm: the palaeoanthropological bridge between chemistry and culture. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 373:20180212. doi: 10.1098/rstb.2018.0212
- Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., and Knight, R. T. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proc. Natl. Acad. Sci. U S A* 113, 11366–11371. doi: 10.1073/pnas.1603312113
- Pilley, J. W., and Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behav. Processes* 86, 184–195. doi: 10.1016/j.beproc.2010.11.007
- Pinker, S. (2010). The cognitive niche: coevolution of intelligence, sociality and language. *Proc. Natl. Acad. Sci. U S A* 107, 8993–8999. doi: 10.1073/pnas.0914630107
- Rauscher, F., Krauss, R. M., and Chen, Y. (1996). Gesture, speech and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* 7, 226–231. doi: 10.1111/j.1467-9280.1996.tb00364.x
- Rugg, M. D., and Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Curr. Opin. Neurobiol.* 23, 255–260. doi: 10.1016/j.conb.2012.11.005
- Russon, A., and Andrews, K. (2001). Orangutan pantomime: elaborating the message. *Biol. Lett.* 7, 627–630. doi: 10.1098/rsbl.2010.0564
- Savage-Rumbaugh, S., Shanker, S. G., and Taylor, T. J. (1998). *Apes, Language and the Human Mind*. New York: Oxford University Press.
- Seyfarth, R. M., and Cheney, D. L. (2018). “The social origins of language,” in *The Social Origins of Language*, eds R. M. Seyfarth and D. L. Cheney (Princeton, NJ: Princeton University Press), 9–33.
- Slocumbe, K. E., and Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Curr. Biol.* 15, 1779–1784. doi: 10.1016/j.cub.2005.08.068
- Suddendorf, T. (2013). Mental time travel: continuities and discontinuities. *Trends Cogn. Sci.* 17, 151–152. doi: 10.1016/j.tics.2013.01.011
- Suddendorf, T., and Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genet. Soc. Gen. Psychol. Monogr.* 123, 133–167.
- Suddendorf, T., and Corballis, M. C. (2007). The evolution of foresight: what is mental time travel and is it unique to humans? *Behav. Brain Sci.* 30, 299–351. doi: 10.1017/s0140525x07001975
- Tanner, J. E., and Perlman, M. (2017). Moving beyond ‘meaning’: gorillas combine gestures into sequences for creative display. *Lang. Commun.* 54, 56–72. doi: 10.1016/j.langcom.2016.10.006
- Tattersall, I. (2012). *Masters of the Planet: The Search for Human Origins*. New York, NY: Palgrave Macmillan.
- Timmermann, A., and Friedrich, T. (2016). Late Pleistocene climate drivers of early human migration. *Nature* 538, 92–95. doi: 10.1038/nature19365
- Tomasello, M. (2008). *The Origins of Human Communication*. Cambridge, MA: MIT Press.

- Tulving, E. (1972). "Episodic and semantic memory," in *Organization of Memory*, eds E. Tulving and W. Donaldson (New York, NY: Academic Press), 382–403.
- Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12. doi: 10.1037/h0080017
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25. doi: 10.1146/annurev.psych.53.100901.135114
- Vaz, A. P., Inati, S. K., Brunel, N., and Zagloul, K. A. (2019). Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. *Science* 363, 975–978. doi: 10.1126/science.aau8956
- Venkataraman, V. V., Kraft, T. S., Dominy, N. J., and Endicott, K. M. (2017). Hunter-gatherer residential mobility and the marginal value of rainforest patches. *Proc. Natl. Acad. Sci. U S A* 114, 3097–3102. doi: 10.1073/pnas.1617542114
- Villa, P., and Roebroeks, W. (2014). Neandertal demise: an archaeological analysis of the modern human superiority complex. *PLoS One* 9:e96424. doi: 10.1371/journal.pone.0096424
- Wagner, S., Nusbaum, H., and Goldin-Meadow, S. (2004). Probing the mental representation of gesture: is handwaving spatial? *J. Mem. Lang.* 50, 395–407. doi: 10.1016/j.jml.2004.01.002
- Warren, D. E., and Duff, M. C. (2019). Fast mappers, slow learners: word learning without hippocampus is slow and sparse irrespective of methodology. *Cogn. Neurosci.* doi: 10.1080/17588928.2019.1593120 [Epub ahead of print].
- Wearing, D. (2005). *Forever Today*. New York: Doubleday.
- Wilson, A. G., Pizzo, M. J., and Crystal, J. D. (2013). Event-based prospective memory in the rat. *Curr. Biol.* 23, 1089–1093. doi: 10.1016/j.cub.2013.04.067
- Zhu, Z., Dennell, R., Huang, W., Wu, Y., Qiu, S., and Ouyang, T. (2018). Hominin occupation of the Chinese Loess Plateau since about 2.1 million years ago. *Nature* 559, 608–611. doi: 10.1038/s41586-018-0299-4

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

Copyright © 2019 Corballis. 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.





# Downstream Behavioral and Electrophysiological Consequences of Word Prediction on Recognition Memory

Ryan J. Hubbard<sup>1,2\*</sup>, Joost Rommers<sup>3</sup>, Cassandra L. Jacobs<sup>4</sup> and Kara D. Federmeier<sup>1,2,5</sup>

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Lin Wang,  
Massachusetts General Hospital,  
Harvard Medical School,  
United States  
Jan Philipp Röer,  
Heinrich Heine University of  
Düsseldorf, Germany

### \*Correspondence:

Ryan J. Hubbard  
rjhubba2@illinois.edu

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 25 June 2019

**Accepted:** 12 August 2019

**Published:** 28 August 2019

### Citation:

Hubbard RJ, Rommers J, Jacobs CL  
and Federmeier KD  
(2019) Downstream Behavioral and  
Electrophysiological Consequences  
of Word Prediction on  
Recognition Memory.  
Front. Hum. Neurosci. 13:291.  
doi: 10.3389/fnhum.2019.00291

<sup>1</sup>Department of Psychology, University of Illinois, Urbana-Champaign, IL, United States, <sup>2</sup>Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign, IL, United States, <sup>3</sup>Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands, <sup>4</sup>Department of Psychology, Center for Mind and Brain, University of California, Davis, Davis, CA, United States, <sup>5</sup>Program in Neuroscience, University of Illinois, Urbana-Champaign, IL, United States

When people process language, they can use context to predict upcoming information, influencing processing and comprehension as seen in both behavioral and neural measures. Although numerous studies have shown immediate facilitative effects of confirmed predictions, the downstream consequences of prediction have been less explored. In the current study, we examined those consequences by probing participants' recognition memory for words after they read sets of sentences. Participants read strongly and weakly constraining sentences with expected or unexpected endings ("I added my name to the list/basket"), and later were tested on their memory for the sentence endings while EEG was recorded. Critically, the memory test contained words that were predictable ("list") but were never read (participants saw "basket"). Behaviorally, participants showed successful discrimination between old and new items, but false alarmed to the expected-item lures more often than to new items, showing that predicted words or concepts can linger, even when predictions are disconfirmed. Although false alarm rates did not differ by constraint, event-related potentials (ERPs) differed between false alarms to strongly and weakly predictable words. Additionally, previously unexpected (compared to previously expected) endings that appeared on the memory test elicited larger N1 and LPC amplitudes, suggesting greater attention and episodic recollection. In contrast, highly predictable sentence endings that had been read elicited reduced LPC amplitudes during the memory test. Thus, prediction can facilitate processing in the moment, but can also lead to false memory and reduced recollection for predictable information.

**Keywords:** language comprehension, prediction, false memory, recognition, ERP

## INTRODUCTION

The process of prediction has been suggested to play a role in many areas of cognition and behavior, with some arguing that one of the core functions of the brain is to use previously learned associations and top-down control to predict future events (Bar, 2007, 2009; Bubic et al., 2010; Clark, 2013). This function of predicting upcoming information may play a particularly important role in language comprehension (Federmeier, 2007; Kuperberg and Jaeger, 2016), as incoming linguistic information must be processed rapidly. Essentially, by using the bottom-up sensory information provided by written and spoken words, combined with previously learned world knowledge, semantic, and syntactic information, the brain can quickly create and continuously update a representation of likely upcoming linguistic information, which facilitates processing when this information is encountered.

As evidence of the impact of predictability on language comprehension, behavioral work has shown that words that are highly predictable and fit into the ongoing sentence context are processed more rapidly than less predictable words (West and Stanovich, 1978; Fischler and Bloom, 1979; Schuberth et al., 1981; Schwanenflugel and LaCount, 1988; Duffy et al., 1989; Simpson et al., 1989; Hess et al., 1995). Similarly, eye-tracking studies have demonstrated that predictable words are anticipated and read more quickly than unpredictable words (Ehrlich and Rayner, 1981; Altmann and Kamide, 1999; Frisson et al., 2005; Kamide, 2008). Research using event-related potentials (ERPs) has identified that the predictability of words affects the amplitude of the N400, a centroparietal negativity peaking around 400 ms that is thought to index access of semantic memory (Kutas and Hillyard, 1984; Federmeier and Kutas, 1999; Wlotko and Federmeier, 2007, 2012; Kutas and Federmeier, 2011; DeLong et al., 2014). Additionally, unexpected but plausible words that disconfirm a prediction elicit a late, frontally-distributed positivity, which has been hypothesized to index a revision process of some kind (Federmeier et al., 2007; Otten and Van Berkum, 2008; DeLong et al., 2011, 2014; Thornhill and Van Petten, 2012).

There is thus substantial evidence that predictability can lead to facilitated processing of expected information when it is encountered. There are also consequences of processing inputs that violate predictions, as indexed by the late frontal positivity. Do these consequences that are evident in ERPs have corresponding behavioral costs? In early work using lexical decision tasks, identification of predictable words was consistently faster than unpredictable words, but prediction violations did not always lead to response slowing when compared to “baseline” conditions, which varied across the literature (Schuberth and Eimas, 1977; Fischler and Bloom, 1979; Schwanenflugel and Shoben, 1985). Other recent work, in which subjects read sentences at their own pace while eye movements were tracked, reported no evidence of slowing or an increase in re-reading for unexpected words (Luke and Christianson, 2016; Frisson et al., 2017). Therefore, across multiple behavioral paradigms of language processing, convincing evidence of

behavioral costs associated with prediction violations has been lacking.

In addition, behavioral and electrophysiological effects of prediction have predominantly been measured at the time of encountering the predicted or unexpected stimulus. Although this has been useful for identifying the immediate effects of prediction, it leaves open what downstream effects confirmed or disconfirmed predictions might have on later cognition. In order to investigate these potential downstream effects, the present study tested participants’ episodic memory for sentence final words of sentences that varied in contextual constraint. The memory test contained words that had been highly predictable, weakly predictable, or unexpected. This allowed for a comparison of the downstream effects of having predictions confirmed or disconfirmed. Critically, the test also included words that were likely to have been predicted but were never actually observed during reading (because the sentence instead had ended with an unexpected word); we will refer to these items as lures.

In addition to behavioral memory measures, the present study recorded EEG to further probe how predictability influences memory processes. Examining ERPs during the memory test allowed us to draw inferences about the neurocognitive processes involved in successfully recognizing, or false alarming to, predictable and unexpected words. Previous studies have identified two major components associated with recognition memory (Rugg and Curran, 2007)—the N400, which has been linked to conceptual fluency or familiarity (Paller and Kutas, 1992; Curran, 2000, 2004; Voss and Federmeier, 2011), with greater familiarity leading to smaller N400s, and the LPC, a left-lateralized posterior component temporally extending from 500 to 800 ms, which is related to recollection or retrieval of more detailed episodic information (Düzel et al., 1997; Rugg et al., 1998; Woodruff et al., 2006; Yu and Rugg, 2010), with greater recollection eliciting more positive LPCs. The amplitudes of these ERP components during the memory test may differ based on the prior predictability of the words or the constraint of the sentences they were presented in, which would provide information about the state of the representations of these items.

Two main issues were of interest: first, we compared memory for predictable words and unexpected words. Here, context-driven prediction could influence the encoding of information into long-term memory by modulating the level of attention given to the predictable or unpredictable information that is being encoded (Craig et al., 1996). Paying more attention to certain stimuli could modulate the depth or level of processing, leading to a more stable and persistent memory representation (Craig and Lockhart, 1972; Craig and Tulving, 1975). In eye-tracking experiments with natural reading, individuals spend less time looking at and exhibit fewer regressions to predictable words (Ehrlich and Rayner, 1981), suggesting they may, in fact, pay less attention to them. Rommers and Federmeier (2018a), investigating ERP repetition effects for previously predictable and unpredictable words, also found that previously predictable words showed reduced downstream repetition effects, suggesting that prediction can lead to an impoverished initial representation. In the case of unpredictable words, some evidence points toward attentional enhancement of encoding: an

item in a list of words that is physically or semantically distinct from the others will be more likely to be recalled (Von Restorff, 1933), unexpected sentence endings draw more attention away from and lead to disruption of serial recall (Röer et al., 2019), and unexpected or error-related events modulate early attention-related ERPs (Wills et al., 2007), suggesting that distinctive, unpredictable events might be more attended to and then more easily remembered. Indeed, studies have reported better recognition memory performance for sentence endings that had been unpredictable (compared with predictable endings), supporting the idea that such words are encoded more strongly (Corley et al., 2007; Federmeier et al., 2007). We further probed the memory processes underlying the recognition of previously predictable and previously unexpected words. In particular, if previously encountered sentence endings increase conceptual priming at test, they should show a reduced N400, whereas if they increase recollection processes, they should elicit an enhanced LPC.

We were also interested in the responses to the lures. If prediction during sentence comprehension leads to pre-activation of information associated with an upcoming word, then participants may show greater false alarms to lures as compared to completely new items. This would constitute a cost of prediction, in that lingering representations can cause false recognition. Alternatively, if the prediction disconfirmation leads to strong revision processes that suppress previously expected information, participants may show fewer false alarms to lures as compared to completely new items. Previous studies have employed an implicit memory paradigm in which participants predict a high cloze ending, are given an unexpected ending, and then must complete a mid-cloze sentence that could potentially be completed by a previous high-cloze or unexpected ending (Hartman and Hasher, 1991; Lorsch et al., 1996; Hasher et al., 1997). These studies have focused mainly on inhibition and control processes; however, they have demonstrated that individuals tend to retain the expected but disconfirmed endings in some form. In terms of explicit memory, classic studies using the Deese-Roediger-McDermott (DRM) paradigm have shown that individuals will recall an unstudied semantic associate (e.g., “sleep”) following study of a list of related words (“dream,” “bed,” “night,” et cetera), suggesting that the representation of the lure was activated and erroneously selected during retrieval (Deese, 1959; Roediger and McDermott, 1995; Steffens and Mecklenbräuer, 2007). In these studies, false alarming is largely driven by semantic similarity of items, and generally occurs immediately following the study. In the current experiment, participants read sentences that were not semantically similar, and were tested after reading several items; thus, a finding of increased false alarming to lures would be a powerful demonstration of prediction’s lasting effects on recognition memory.

In addition to behavioral effects, we were also interested in the processes involved in false recognition, as revealed by electrophysiological responses; however, previous results of when and how false recognition manifests in the ERP have been mixed (Curran et al., 2001; Wolk et al., 2006; Geng et al., 2007; Beato et al., 2012; Chen et al., 2012). A recent ERP

study showed that words that were previously expected, but not presented, elicited a “pseudo-repetition” effect (Rommers and Federmeier, 2018b); namely, these items showed ERP effects similar to repeated words, suggesting they were not fully suppressed. We hypothesized that, if similar processes also influence end-state recognition responses, these predicted but unobserved lures would show higher false alarms than new items. Furthermore, we used the N400 and LPC to help clarify the neurocognitive mechanisms involved in prediction-based false alarms vs. correct rejections of lures, focusing on whether these responses were associated with priming and/or recollection during the recognition test.

## MATERIALS AND METHODS

### Participants

Thirty-three right-handed, native speakers of English with normal or corrected-to-normal vision from the University of Illinois, Urbana-Champaign participated in the experiment and were paid \$10 an hour or received course credit for their participation. No participant had a history of neuropsychological or psychiatric disorders. Procedures were approved by the IRB of the University of Illinois, and all participants signed consent forms prior to participation. Based on previous work using these same materials to examine ERPs during sentence comprehension (Federmeier et al., 2007), the *a priori* number of subjects was set to 32; mid-way through data collection, a participant’s recorded data was noisy, and thus an extra subject was run. Data analysis led to the removal of another subject’s data due to high trial loss, leading to a sample size of 31 participants in the final analyses.

### Materials

The stimuli were comprised of 192 English sentences, a subset of the sentences used in Federmeier et al. (2007). The cloze probabilities of the endings of the sentences were previously determined in a norming study in which the subjects filled in the final word of the sentence frame with the word they “would generally expect to find completing the sentence fragment.” In the current experiment, half of the stimuli (96 sentences) were strongly constraining, while the other half were weakly constraining. A sentence was considered strongly constraining if the cloze probability of the most commonly completed word was 0.68 or higher, and was considered weakly constraining if the cloze probability was 0.42 or lower. Additionally, half of the strongly constraining sentences (48 sentences) ended with the expected word, while the other half ended with an unexpected word; this was also true of the weakly constraining sentences. Unexpected words all had a cloze probability close to 0 (max = 0.088). Thus, participants read 48 strongly constraining sentences with expected endings (SCE; mean cloze = 0.83), 48 with unexpected endings (SCU), 48 weakly constraining sentences with expected endings (WCE; mean cloze = 0.28), and 48 with unexpected endings (WCU). These stimuli were evenly split into eight blocks (six of each condition in each block). **Table 1** provides the lexical properties (word frequency, concreteness, imageability, familiarity, word length) of the sentence ending words. Target words averaged 5–6 letters

**TABLE 1** | Lexical properties of sentence ending words.

Condition	Frequency	Concreteness	Imageability	Familiarity	Word length
SCE	4.16	506.10	527.31	574.00	4.98
SCU	3.06	502.45	520.82	554.67	6.29
WCE	3.96	501.10	518.38	575.94	5.52
WCU	3.22	485.83	520.19	554.027	5.85

Values represent means across items. Frequency values are log transformed and obtained from Kucera and Francis (1967). Concreteness, imageability, and familiarity values obtained from the MRC psycholinguistic database.

**TABLE 2** | Examples of experimental materials.

SC	Tim threw a rock and broke the	E	window	(Match)	window
		U	camera	(Match)	camera
WC	His ring fell into a hole in the			(Lure)	window
		E	sink	(Match)	sink
		U	couch	(Match)	couch
				(Lure)	sink

SC, strong constraint; WC, weak constraint; E, expected; U, unexpected. Match and Lure refer to the items that appear during the memory test.

**TABLE 3** | Lexical properties of test words.

Condition	Frequency	Concreteness	Imageability	Familiarity	Word length
SCE Match	4.21	526.58	539.95	573.63	5.04
SCU Match	2.94	477.00	503.84	546.33	6.21
SC Lure	4.13	507.45	527.86	571.77	5.29
WCE Match	4.04	503.87	525.00	582.44	5.38
WCU Match	2.99	476.22	509.67	552.94	6.08
WC Lure	3.43	509.81	527.18	580.94	5.29
New	3.70	497.77	525.31	560.15	5.77

Values represent means across items. Frequency values are log transformed and obtained from Kucera and Francis (1967). Concreteness, imageability, and familiarity values obtained from the MRC psycholinguistic database.

in length and were fairly concrete, imageable, and familiar; unexpected endings tended to be of lower frequency on average than expected endings but were similar across constraint.

After each block of sentence reading, participants took a memory test. For the recognition memory test, participants were presented with single words, the majority of which were words that had ended the previously read sentences. “Matches” were words that had previously been seen as sentence endings (either expected or unexpected). “Lures” were words that might have been expected (they were the most likely completion of a sentence from the prior block) but were never actually presented (because the sentence had ended with an unexpected word instead). As an example of a “Lure” item, during the encoding phase a participant might read the sentence “I added my name to the basket,” where basket is an unexpected ending, and in the test phase read the word “list,” the expected ending of the sentence. “New” words had never been presented in the block. The test also contained some sentence-medial words, to ensure that participants would be motivated to pay attention to and encode the entire sentence. Half of the test items that were previously sentence ending words were from strongly constraining sentences while the other half were from weakly constraining sentences. **Table 2** provides examples of the different types of test items. There was an equal number of items presented in each of the conditions during each test block (totaling 24 in each condition, along with 48 new items, over the course of the experiment), as

well as an equal number of “old” and “new” items each test, so as not to bias responses. As with the sentence endings, lexical properties of the test items (see **Table 3** for details) were similar across conditions, with some variation in frequency; we aimed to test the impact of the frequency variability in our statistical models.

The memory test constrained the stimuli used and the order of presentation, in that each test item had to be unique, as well as not repeat. For example, participants might read the sentence “he played with the dog,” see the word “dog” during the memory test, and later read the sentence “the dog ate the food.” To avoid participants reading both sentences before being tested, the stimulus list was pseudo-randomized, such that any sentence containing a critical test item in the middle of it was presented only after the item had already been tested. All participants read the same list of stimuli; although the order of presentation of each stimulus within blocks was randomized, the order of presentation of the blocks was not.

## Procedure

Participants were seated in an electrically shielded EEG recording booth approximately 100 cm from a CRT computer monitor. Prior to starting the experiment, we verified that all participants could easily read the presented information from this distance. Additionally, participants were given an explanation of the experimental procedure, as well as a short practice session to familiarize them with the task. Words that appeared in the



practice sentences and test items did not appear as critical test words in the actual experiment.

The experiment was divided into eight study-test blocks, in which participants first studied a set of sentences, and then were tested on their memory for critical words. During the encoding phase of each study-test block, participants were instructed to read the sentences silently and to try to remember what they read, as their memory would be tested. Sentences were presented word by word on the screen, with each word appearing in the center of the screen for 200 ms, followed by a 300 ms interstimulus interval. After the last word of the sentence was presented, a blank screen was presented for 500 ms, followed by a fixation cross for 1,000 ms. Participants were instructed to try not to blink when they were reading the sentence, and to blink and rest their eyes once the fixation cross appeared. Following the encoding phase, participants were given math problems to complete for 30 s. The math problems were simply given as a distractor between the study and test phases—thus, performance on the math section was not analyzed.

After the math section, participants started the test phase of the block. Each trial began with a fixation cross in the center of the screen for 1,000 ms, which was then replaced by a test word. After 1,000 ms, a confidence scale appeared underneath the test word, at which point participants could make their response. Upon making a response, the trial would end and the next trial would begin. The confidence scale consisted of four points—“Sure New,” “Maybe New,” “Maybe Old,” and “Sure Old.” Participants were instructed to respond with “Old” if they thought the test word was a word they had seen during the encoding phase and otherwise to respond “New.” Additionally, they were told to try to use the whole scale of confidence and to use the “Maybe” option if they felt like they were guessing or unsure. Finally, participants were instructed to try not to blink during the initial presentation of the word, but told that once the confidence scale appeared and they could make their response, as well as during the fixation cross, they could blink. The test phase was self-paced, in that participants could take as long as they needed to respond.

## EEG Recording and Processing

EEG data were recorded from 26 Ag/AgCl electrodes embedded into a flexible elastic cap and distributed over the scalp in an equidistant arrangement; see icon in **Figure 2**. Five additional electrodes were attached, including one on each mastoid bone behind the ear, one adjacent to the outer canthus of each eye, used for monitoring of the horizontal electro-oculogram (EOG), and one below the lower eyelid of the left eye, used for monitoring of blinks. Electrode impedances were kept below 5 k $\Omega$ . Signals were amplified by a BrainVision amplifier with a 16-bit A/D converter, an input impedance of 10 M $\Omega$ , a bandpass filter of 0.016–100 Hz, and a sampling rate of 1 kHz. The left mastoid electrode was used as a reference for on-line recording; offline, the average of the left and right mastoid electrodes was used as a reference.

Following data collection, each raw EEG time series was passed through a 0.1–30 Hz Butterworth filter with a 12 dB/oct roll-off. The signal was segmented into epochs from –200 to 1,000 ms relative to the onset of each sentence ending word

during encoding and each test item during the test phase. Following subtraction of the 200 ms prestimulus baseline, and artifact correction (described below), epochs within each bin were averaged together to create an ERP for each subject and bin. Prior to calculating statistics, individual subject ERPs were passed through an additional 20 Hz lowpass filter.

To correct for ocular artifacts, a bipolar VEOG channel was created by subtracting data in the lower eye channel from the most frontocentral channel (MiP<sub>f</sub>), and that channel was then scanned with a sliding window step function to detect blinks. For subjects who had a large number of blinks, the data were run through AMICA (Palmer et al., 2011), an ICA decomposition algorithm that generalizes Infomax and multiple mixtures approaches adaptively. Following decomposition, the correlation between the timecourse of each component and the VEOG channel was calculated in order to find the component(s) containing blinks. Components with a high correlation were removed from trials marked as containing blinks. The remaining components were then recombined to reconstruct the EEG data, which were then scanned with an additional sliding window amplitude threshold (300 ms sliding time window, 50 ms step size, 90  $\mu$ V threshold), and finally manually checked by the experimenter for any additional artifacts. In total, an average of 8% of trials were removed, with a range of 2% to 11% across participants. Artifacts were spread fairly evenly across conditions, resulting in an average of 22 trials in each condition of the memory test.

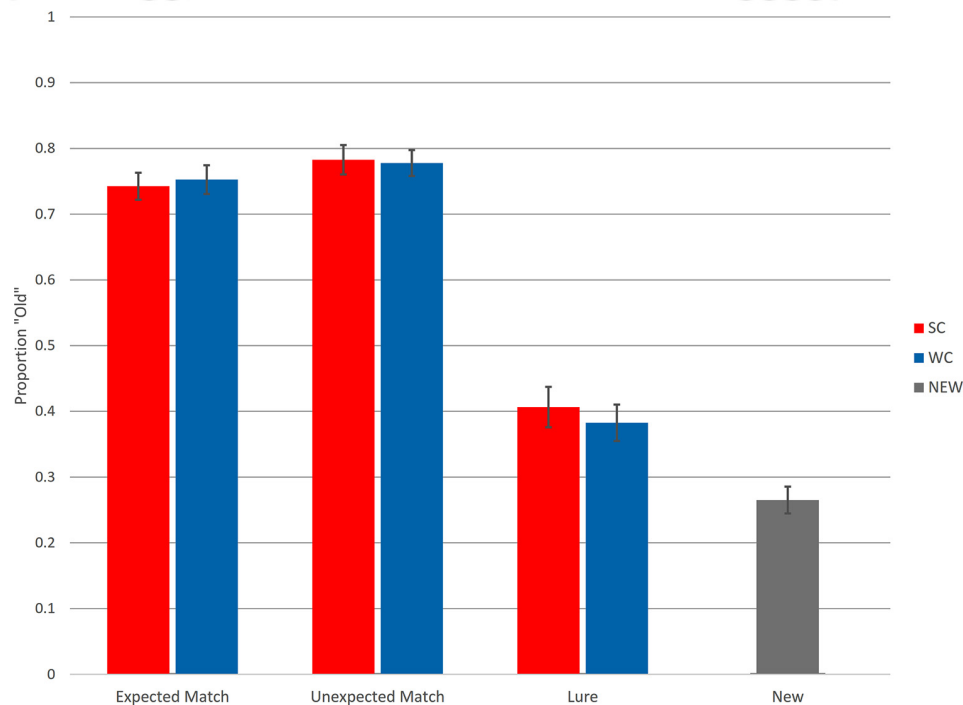
For the ERP analyses, statistical analyses were performed on channel clusters as opposed to single channels to improve the signal to noise ratio. Component-based analyses were done using the signal-averaged across channel clusters and time windows based on prior work: N400 at a central cluster (shown in **Figure 2**; Federmeier et al., 2007), 300–500 ms; frontal positivity at a frontal cluster, 700–1,000 ms (shown in **Figure 2**; DeLong et al., 2014); LPC at a left parietal cluster (shown in **Figure 3**; Woroch and Gonsalves, 2010; Addante et al., 2012), 500–800 ms. For other effects, as described in the results, cluster-based permutations with restricted time windows were used in order to explore the data while retaining statistical power and maintaining Type I error rate (Fields and Kuperberg, 2018). Plotted ERPs were filtered with a 10 Hz lowpass filter for clarity of visualization.

## RESULTS

### Behavior

Proportion “Old” responses is plotted in **Figure 1**. For Matches, “Old” was a correct response, whereas for New items and Lures, “Old” was an incorrect response. Analyses revealed no differences in confidence across experimental conditions and generally low trial numbers for “Maybe” responses; thus, “Maybe” responses were combined with “Sure” responses for behavioral and ERP analyses. Overall, participants successfully discriminated Matches from New items. Collapsing across Match conditions and comparing to New items, the average  $d'$  was





**FIGURE 1 |** Recognition memory accuracy. Proportion “Old” responses are plotted on the Y axis. SC, strong constraint; W, weak constraint. Error bars reflect standard error around the mean.

1.41, with a range of 0.72–2.76. Recognition accuracy between Expected and Unexpected Matches appeared similar, whereas participants false alarmed more to Lures compared to New items.

To assess the pattern statistically, behavioral responses (Old or New) on each trial were submitted to a mixed-effects logistic regression model fit by maximum likelihood using the lme4 package in R (Jaeger, 2008). Random factors included intercepts for items and slopes and intercepts for participants for each fixed effect. Correlations between random factors were not calculated to ease convergence of the models. Wald’s *z*-scores were computed for each coefficient to test for significance.

The first model compared responses to Lures with responses to New items by modeling responses to those items with Condition (Lures, New) as a fixed factor. Recognition accuracy differed between Lures and New items ( $\beta = 0.75$ ,  $z = 3.18$ ,  $p < 0.01$ ), but accuracy did not differ between Strong Constraint Lures and Weak Constraint Lures ( $\beta = 0.11$ ,  $z = 0.42$ ,  $p = 0.68$ ). Thus, participants showed greater false alarms to Lures compared to New items.

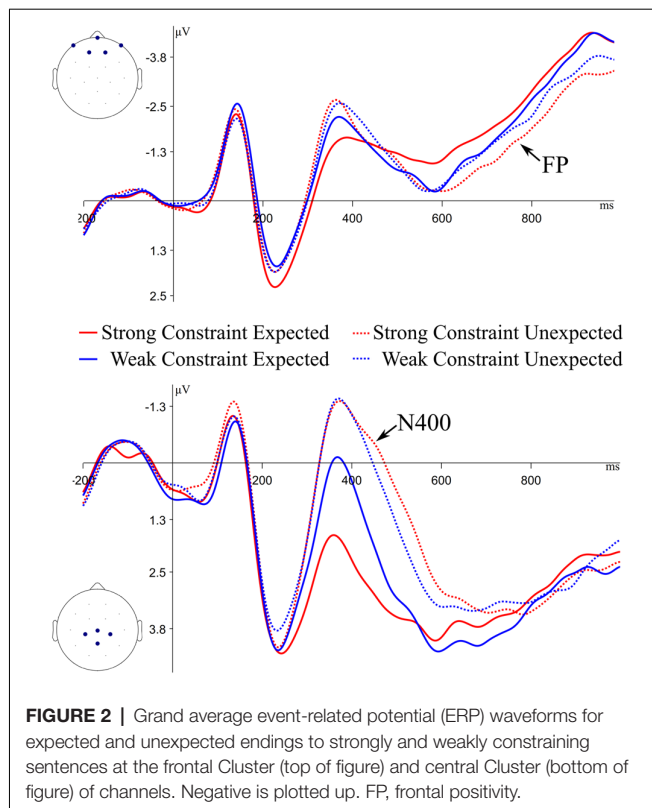
Although we attempted to control the lexical properties of stimuli, it could be the case that a subset of the Lures were more frequent than other Lures or New items, and this could have contributed to the false alarm effect. To assess this, a second model was fit with Condition and log-transformed Word Frequency as fixed effects. Frequency had a significant effect on responses ( $\beta = 0.30$ ,  $z = 4.72$ ,  $p < 0.01$ ), with higher Frequency leading to a greater number of “Old” responses, but recognition

accuracy still differed between Lures and New items ( $\beta = 0.64$ ,  $z = 2.91$ ,  $p < 0.01$ ). Thus, word frequency did not completely explain the false alarm effect that we observed.

The next model assessed responses for Matches by modeling responses with Constraint, Expectedness, and the interaction (Constraint \* Expectedness) as fixed factors. None of the coefficients, Constraint ( $\beta < 0.01$ ,  $z = 0.06$ ,  $p = 0.95$ ), Expectedness ( $\beta = 0.21$ ,  $z = 1.26$ ,  $p = 0.21$ ), or the interaction ( $\beta = 0.13$ ,  $z = 0.27$ ,  $p = 0.79$ ), returned significant *z*-scores. Including word frequency in the model (C\*E\*F) did not change previous results, although word frequency seemed to have a tendency to reduce “Old” responses ( $\beta = 0.10$ ,  $z = -1.94$ ,  $p = 0.05$ ). Thus, behavioral accuracy for Match items did not differ based on constraint or expectedness.

## Sentence Final Word ERPs

ERPs to sentence final words were analyzed to determine if prior effects seen with these materials (e.g., Federmeier et al., 2007) were replicated. Grand average ERPs at the sentence final word at the frontal and central cluster are plotted in **Figure 2**. To assess effects statistically, linear mixed-effects models were used (Baayen et al., 2008), using the lme4 and lmerTest packages in R. Random factors included intercepts for items and slopes and intercepts for participants. As with the behavioral analyses, correlations between random factors were not calculated to ease convergence of the models. The reported *t*-tests used the Satterthwaite approximations to calculate degrees of freedom (Satterthwaite, 1946).

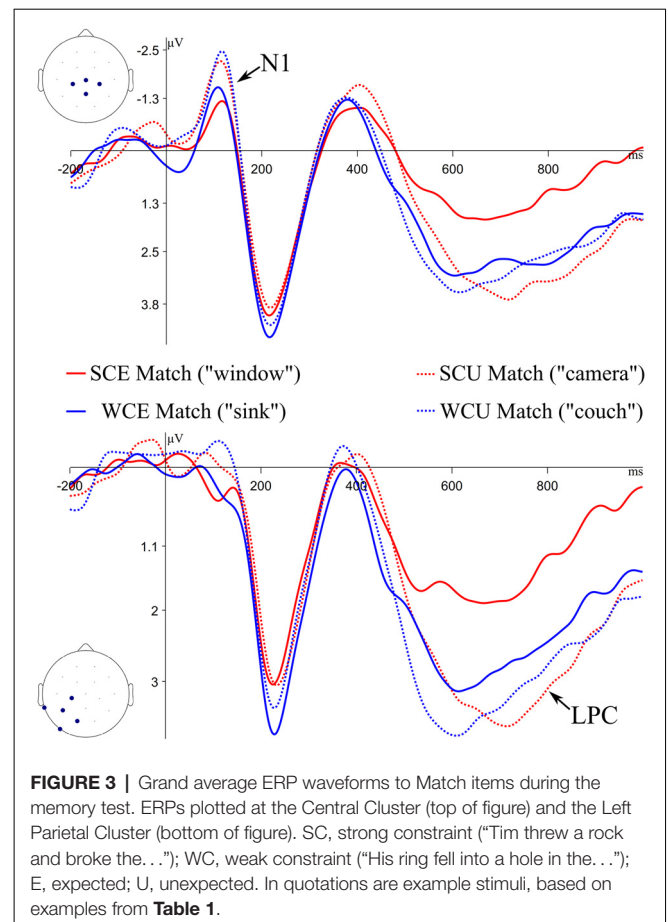


N400 amplitudes were compared between weakly constrained expected (WCE) endings and strongly constrained expected (SCE) endings, as well as between WCE and unexpected (U) endings (collapsed across constraint, as this has repeatedly been shown not to affect N400 responses). There were significant differences in N400 amplitude between WCE and SCE endings ( $\beta = 1.32$ ,  $t = 2.87$ ,  $p < 0.01$ ), as well as between WCE and U endings ( $\beta = 1.70$ ,  $t = 4.42$ ,  $p < 0.01$ ). Thus, the graded N400 effect was replicated in this experiment.

ERPs to sentence final words were also analyzed to determine if the frontal positivity to Strong Constraint Unexpected endings was replicated. The frontal positivity has been operationalized as a difference between Strong Constraint Unexpected (SCU) and Weak Constraint Unexpected (WCU) endings (Federmeier et al., 2007), or a difference between expected (E) endings and SCU endings (DeLong et al., 2014), so both of these differences were tested. There were no significant differences in frontal positivity amplitudes between the SCU and WCU conditions ( $\beta = 0.41$ ,  $t = 0.95$ ,  $p = 0.35$ ); however, SCU endings elicited larger positivities than E endings ( $\beta = 0.84$ ,  $t = 2.01$ ,  $p = 0.05$ ). A follow-up comparison of WCU and E conditions showed no significant differences ( $\beta = 0.42$ ,  $t = -1.04$ ,  $p = 0.31$ ). Thus, the frontal positivity from SCU endings was more positive than other conditions, replicating prior work, but did not differ significantly from the WCU condition.

## Recognition Memory ERPs: Matches

ERPs to correctly recognized test items were analyzed to assess recognition memory processes. The grand average ERPs at



the central cluster to expected and unexpected Matches from strongly and weakly constraining sentences are plotted in Figure 3. ERPs are time-locked to the onset of the test item, and only correct responses are included.

LPC mean amplitudes from 500 to 800 ms at the Left Parietal cluster were submitted to a linear mixed effect model with fixed effects of Expectancy (E vs. U) and Constraint (SC vs. WC). The fixed effect of Expectancy was significant ( $\beta = 1.14$ ,  $t = 2.53$ ,  $p = 0.01$ ), whereas Constraint ( $\beta = 0.51$ ,  $t = -0.97$ ,  $p = 0.34$ ) and the interaction ( $\beta = 0.80$ ,  $t = 0.94$ ,  $p = 0.35$ ) were not. A follow-up comparison of LPCs from SCE Matches and WCE Matches trended toward significance ( $\beta = 0.94$ ,  $t = -1.93$ ,  $p = 0.06$ ). Unexpected Matches generated more positive LPC amplitudes compared to Expected Matches, and SCE Matches generated the smallest LPC amplitudes.

Visual inspection suggested an additional effect on the N1, a component that is part of the visual evoked potential and is sensitive to attention (Mangun and Hillyard, 1991). To assess this effect, we performed a *post hoc* exploratory analysis, using a cluster-based permutation test with a restricted time window based on previous literature to increase statistical power (Maris and Oostenveld, 2007; Groppe et al., 2011; Fields and Kuperberg, 2018). In this test, *t*-tests were calculated at each time-point and channel, and significant *t*-values that were adjacent in space and time were clustered together. Clusters were

characterized by taking the sum of  $t$ -values within the adjacent points. These observed clusters were compared to a permutation distribution, generated by shuffling the condition labels of the data, finding clusters, and summing the  $t$ -values of the clusters 2,500 times. Distributions of the most extreme cluster sums were created for comparison to the observed cluster sums. Reported  $p$ -values represent the percentile ranking of the observed clusters compared to the permutation distribution. Here,  $t$ -tests tested differences between Expected Matches and Unexpected Matches at each channel and time-point within the 50–175 ms window, and a family-wise alpha of 0.05 was used.

The results of this analysis are displayed in **Figure 4**. A significant difference between Expected and Unexpected Matches was found (cluster-wise  $p < 0.05$ ). This difference had a temporal extent from 81 to 153 ms and a central-posterior topography, similar to previously reported posterior visual N1 effects, though somewhat earlier in time (Di Russo et al., 2001; Hopf et al., 2002). Thus, Unexpected Matches elicited more negative N1 potentials compared to Expected Matches<sup>1</sup>. To test for the possibility of pre-stimulus activity leading to the appearance of an N1 effect, an addition permutation test was run on the same contrast in the 0–80 ms time window. No significant clusters were found ( $p = 0.29$ ).

## Recognition Memory ERPs: Lures

Of particular interest for the analysis of ERPs to Lures was if ERPs differed between false alarms and correct rejections, and whether this ERP difference was affected by constraint. However, few studies have investigated ERP differences to false alarms and correct rejections, particularly for previously predicted information. Thus, while we were interested in early vs. late differences, there were not *a priori* predictions about particular ERP components to target in the post-N400 time window. We thus used time-constrained permutation tests, as described for the N1 analyses (Fields and Kuperberg, 2018). ERPs to SC and WC Lures were separated into Correct and Incorrect bins based on the response given (pooled across “Maybe” and “Sure”), and the difference between these ERPs was calculated. These difference waves were submitted to cluster-based permutation tests to test time-points for significant differences from 0, using a family-wise alpha value of 0.05. Separate permutation tests were run for Strong Constraint and Weak Constraint lures, and to increase statistical power and focus on times of interest, separate permutation tests were run for time windows of 300–500 ms (N400) and 500–1,000 ms.

Results of the permutation tests and ERPs are plotted in **Figure 5**. For the Strong Constraint Lure comparison, a significant difference (cluster-wise  $p = 0.04$ ) between false alarms and correct rejections was found in the 300–500 ms time window, while no significant differences were found in the late

window. This difference began from the onset of the analysis window and continued to 488 ms, with a central-posterior topography. For the Weak Constraint Lure comparison, a significant difference (cluster-wise  $p < 0.01$ ) between false alarms and correct rejections was found in the late time window, while no significant differences were found in the earlier window. This cluster showed a broad right-lateralized topography, with a right frontal maxima, and a temporal extent of 594–1,000 ms. These results suggest that mechanisms with different timecourses led to false alarming based on the constraint of the item<sup>2</sup>.

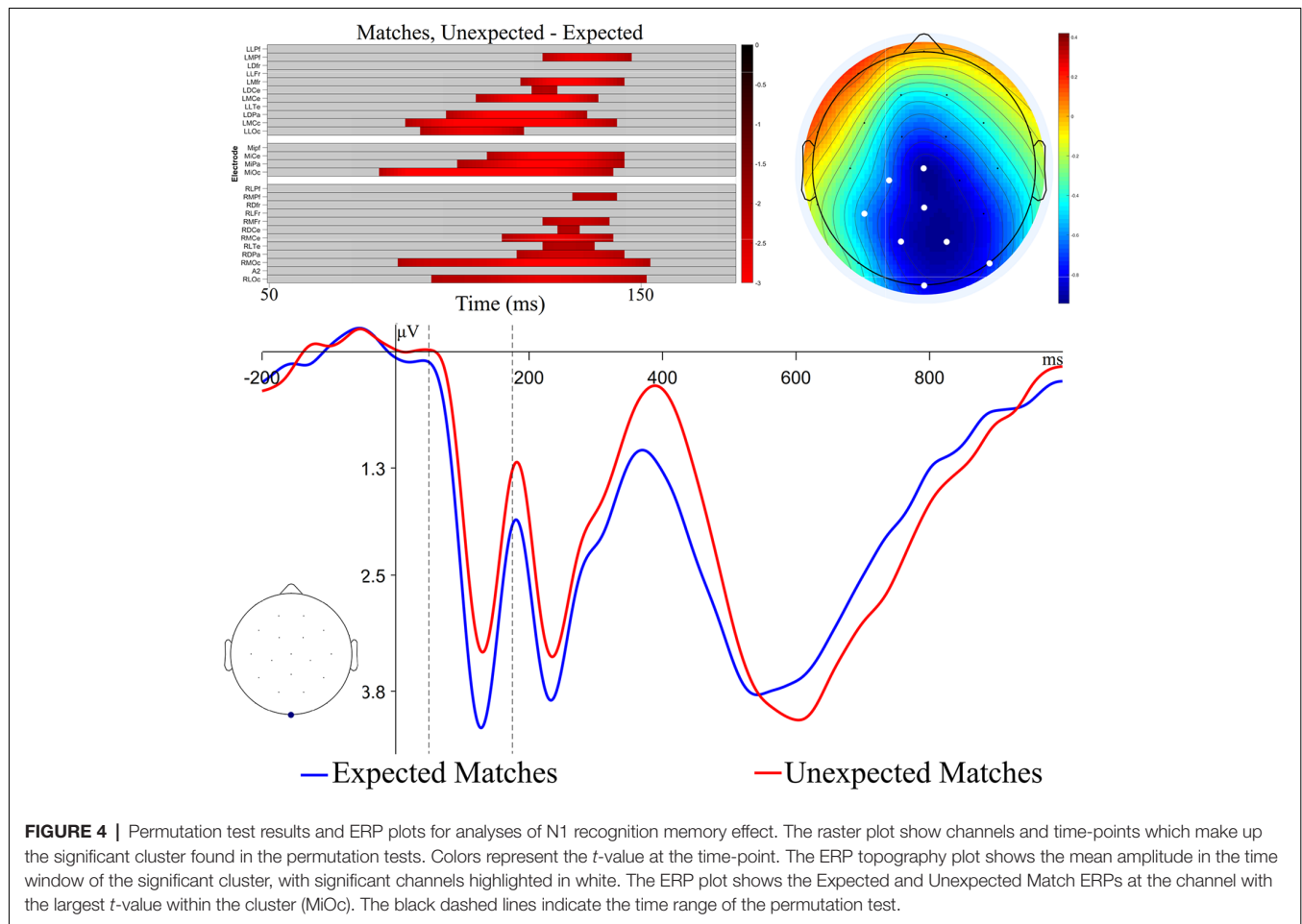
The behavioral effect of interest was the comparison of false alarm rates of Lure items compared to false alarm rates of New items; therefore, we were also interested in how the electrophysiological differences associated with false alarming to Lures compared to those associated with false alarming to New items. **Figure 6** plots correct rejection and false alarm ERPs for Weak Constraint Lures as well as New items; although the ERPs at the same channel as before are plotted, the ERP patterns between these conditions were fairly similar across other channels as well. Permutation tests testing for differences between correct rejections and false alarm ERPs to New items in both the 300–500 ms and 500–1,000 ms windows were not significant (early,  $p = 0.09$ ; late,  $p = 0.11$ ), but numerically, false alarming to Weak Constraint Lures seemed to have engaged similar neurocognitive processes as false alarming to New items.

## DISCUSSION

In this study, participants read strong and weak constraint sentences that ended with either an expected or unexpected-but-plausible word and then were tested on their memory for sentence ending words, new words, and predictable endings that had never been seen (lures). ERP responses during sentence reading replicated previously shown effects. We observed a graded N400 pattern (Federmeier et al., 2007), such that N400s were smallest to expected items in strong constraint sentences, intermediate to expected items in weak constraint sentences, and largest to unexpected items. We also found a post-N400 frontal positivity, larger for unexpected than expected words and numerically largest for unexpected words in strongly constraining sentences (where predictions can be correspondingly stronger). Different from the pattern in Federmeier et al. (2007), we did not observe a significant difference between unexpected items in strongly and weakly constraining contexts, seemingly because there was also some level of frontal positivity for the unexpected items in the weakly constraining sentences. It is possible that the memory task induced different reading strategies than the passive comprehension task in Federmeier et al. (2007). For example, Brothers et al. (2017) reported a larger frontal positivity to

<sup>1</sup>A mixed effect analysis was also run on single trial N1 amplitudes derived from significant cluster timepoints and channels, with fixed effects of expectancy and word frequency, to control for lexical confounds. The effect of expectancy was significant ( $\beta = 0.21$ ,  $t = 3.41$ ,  $p < 0.01$ ), while frequency was not ( $\beta = 0.03$ ,  $t = 1.53$ ,  $p = 0.13$ ). However, since estimates were derived based on cluster analyses, this mixed effect analysis could be considered double-dipping, and further replication of this effect will be necessary.

<sup>2</sup>Mixed effect analyses were also run on single trial SC and WC lure amplitudes derived from significant clusters, with fixed effects of correct/incorrect and word frequency. For both analyses, the fixed effect of correctness was significant (SC:  $\beta = 0.36$ ,  $t = 2.64$ ,  $p = 0.01$ ; WC:  $\beta = 0.87$ ,  $t = 3.61$ ,  $p < 0.01$ ), while word frequency was not significant (SC:  $\beta = 0.09$ ,  $t = 1.45$ ,  $p = 0.16$ ; WC:  $\beta = 0.05$ ,  $t = 0.72$ ,  $p = 0.48$ ). As with the N1 effect, these results could be considered double dipping and replication will be necessary.



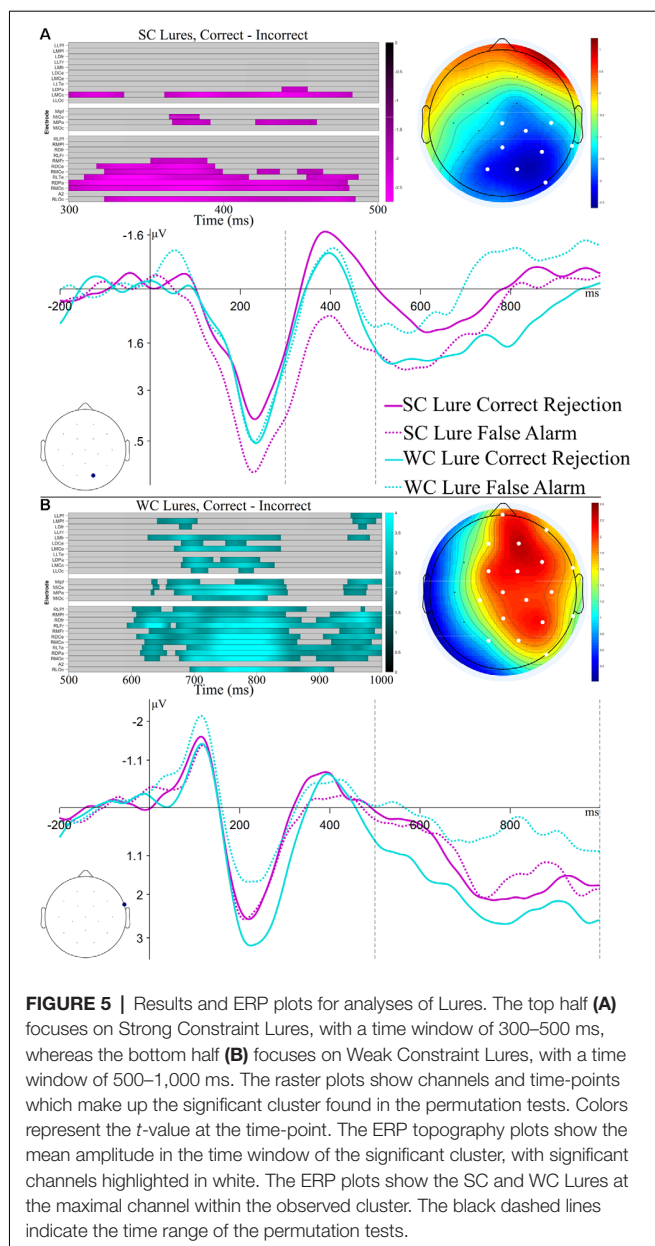
unexpected words when participants were instructed to predict upcoming information compared to when they simply read for comprehension. Anticipating an imminent memory test may have encouraged participants to read more attentively and devote more resources to prediction.

The central question for this study concerned participants' later memory for sentence-ending words they had predicted and/or read. Behaviorally, hit rates were numerically higher for unexpected than for expected matches, though no reliable effect was found. A similar pattern had previously been seen for word recognition at the end of the experiment using these stimuli; higher hit rates were also found for expected words that had completed weakly vs. strongly constraining sentences (Federmeier et al., 2007; see also Corley et al., 2007). The ERPs during the memory test in the present study, however, revealed that LPC responses elicited by unexpected Matches were more positive than those to expected Matches, suggesting greater recollection for unexpected words. Additionally, LPC amplitudes differed between strongly and weakly constrained expected matches, with more positive LPCs for weak constraint matches. This LPC pattern mirrors the behavioral memory performance pattern observed in Federmeier et al. (2007). This pattern may arise because prediction trades off with depth of encoding, such that participants process—and hence encode—predicted words

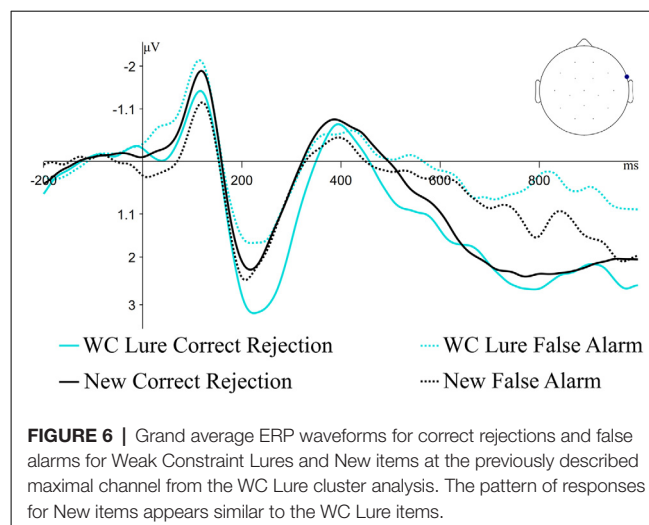
less. In other words, the information needed to verify that an expectation is met may require less attention and less stimulus-driven processing than that needed to encode a stimulus that readers could not predict. A recent ERP repetition study supports this account (Rommers and Federmeier, 2018a). Words that had first been encountered as expected sentence endings of strongly constraining sentences showed reduced ERP repetition effects (when seen again in a weakly constraining sentence) compared to those that had first been seen in weakly constraining sentences. Thus, predictability may have downstream costs: when information is pre-activated, comprehension may take place in a top-down “verification mode” (Van Berkum, 2010), in which readers need only confirm that the stimulus matches with the expectation. This process achieves speedier processing in the moment by sacrificing thorough processing of the bottom-up input, ultimately leading to impoverished representations. Future studies investigating memory for predicted information could assess this further by examining ERPs for misses or incorrect responses, as trial numbers were too low to assess misses here.

Surprisingly, unexpected matches also elicited larger (more negative) N1 amplitudes than did expected matches. N1 amplitude modulations are not routinely reported in electrophysiological studies of recognition memory. Although unexpected sentence endings may have received greater depth of





processing during encoding, ERP studies examining retrieval of words that were deeply or shallowly encoded have not reported modulations of the N1 (Rugg et al., 1998, 2000; Allan et al., 2000). However, N1 modulations have been observed in the context of visual attention and categorization. The N1 is sensitive to the allocation of attentional resources (Mangun and Hillyard, 1991; Hillyard and Anllo-Vento, 1998) and may reflect an early, attention-dependent visual discrimination process that is sensitive to category membership (Vogel and Luck, 2000; Hopf et al., 2002). In one study (Curran et al., 2002), participants were trained in separating abstract blob images into two separate categories (similar or dissimilar to a prototype) and were later given a recognition memory test on the images. The N1 during the recognition test was sensitive to category membership, but not to old/new differences, similar to the current reported



results. Differences in predictability during sentence reading may have led to separable categories during recognition testing; however, given the *post hoc* nature of the analysis of the N1 in the current study, it will be important to replicate the effect in future work, as well as to confirm that the results cannot be explained by other factors (such as lexical variables).

A critical manipulation in the current study was the inclusion of lures—items that were likely to have been predicted during sentence reading but that were never actually presented (because an unexpected word appeared instead). Behaviorally, individuals were significantly more likely to false alarm to Lures than to New items that had not been studied, suggesting increased accessibility or fluency for these items. This pattern is consistent with claims that words are predicted and pre-activated as a sentence unfolds (Federmeier, 2007; Kutas et al., 2011) and further reveals that such predictive pre-activation can have long-lasting effects. Here, several sentences were presented in each block, and each block was followed by interfering math problems, and yet participants still showed increased false alarming to these lures. This finding mirrors previously reported effects from studies on false memory using the DRM paradigm, in which subjects falsely recall—and are more likely to falsely recognize (Gallo, 2010)—critical lures that are semantically similar to studied items. However, a number of differences between the paradigms make the current findings particularly striking. First, in DRM experiments, the lure items are usually closely related to an entire list of words. Here, instead, each predicted sentence ending used as a Lure test item was related to only one sentence in a block, and the sentences were not semantically related to each other. Moreover, different from the DRM paradigm, in the present study predictions were explicitly disconfirmed, *via* the presentation of an unexpected word (which was always semantically unrelated to the predicted ending). Thus, these findings suggest that expected representations are not fully suppressed when a prediction is disconfirmed and that false memories can arise for such disconfirmed information. This presents another cost of prediction during language comprehension: individuals may falsely remember reading or



hearing words that were not actually experienced, simply because they were predicted in the moment, and those predictions linger.

An alternative explanation of the luring effect is that participants could have tried to use the word presented during the test as a cue to perform a retrospective search through memory for a sentence that might have included it. By this account, when a Lure was presented, subjects were able to retrieve a likely sentence frame for that word, and thus more false alarming occurred. Similar to Neely and Keefe's (1989) hybrid prospective-retrospective processing theory, this retroactive search could be performed regardless of any pre-activation of the test item. However, in the case of the Lures in the present study, the associated sentence was completed by an unexpected word. For a retroactive search strategy to work, the unexpected word that originally completed the sentence and its effect on the sentence-level meaning that was extracted would need to be ignored, thus rendering the Lures as ineffective search cues.

Behaviorally, participants did not show a greater rate of false alarms to lures from strongly constraining sentences compared to lures from weakly constraining sentences. However, electrophysiological analyses revealed that different underlying patterns of brain activity were associated with false alarming across constraint. False alarming to strong constraint lures correlated with an earlier, N400-like effect, whereas false alarming to weak constraint lures was associated with a later, right-lateralized effect that was fairly broadly distributed. The N400-like pattern to the lures from the strong constraint sentences is consistent with the idea that false alarms to these items were driven by an increase in conceptual fluency or familiarity (Voss and Federmeier, 2011; Wang et al., 2015). A plausible account of this effect is that when words or concepts are strongly predicted, they linger, such that when the word is encountered again, it is processed more fluently or is more familiar, which behaviorally is associated with a tendency to mark these words as "old" and electrophysiologically is associated with a reduced N400 response. The later right-lateralized effect observed following false alarms to weak constraint lures may be comparable to the right frontal old/new effect in the recognition memory literature, which is thought to index decision making, evaluation, and post-retrieval monitoring processes (Hayama et al., 2008; Cruse and Wilding, 2009; Hayama and Rugg, 2009) and has been related to lure discrimination (Morcom, 2015). Thus, despite a lack of behavioral differences in false alarming based on constraint, it appears different processes may have led to false alarms depending on the prior constraint of the item: a more rapid semantic matching based process for strong constraint lures and a slower, more top-down decision process for weak constraint lures. Future studies could use experimental manipulations to dissociate these effects; for instance, employing speeded recognition decisions would likely increase false alarm

rates for weak constraint lures, but might not affect strong constraint lures.

Overall, these results demonstrate that prediction during language comprehension has important downstream effects on recognition memory. Participants were more likely to false alarm to predictable, but never observed words compared to unexpected and unstudied words, suggesting unobserved predictions are not fully suppressed and remain accessible in memory. Individuals also had enhanced memory for unexpected information, as evidenced by larger LPC amplitudes during recognition testing, along with a larger N1 response. Finally, ERPs revealed sentential constraint-based differences in the neurocognitive mechanisms involved in false alarming to lures, with earlier semantic matching processes contributing to false alarms to strongly predicted information, but later decision-making processes contributing to false alarms to weakly predicted information. Ultimately, prediction during language comprehension does have costs: namely, predicting upcoming words in sentences can produce more rapid processing in the moment, but can lead to impoverished memory of predictable information and false remembering of unobserved predictions.

## DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

### Human Subject Research

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

## AUTHOR CONTRIBUTIONS

RH, JR, CJ, and KF contributed to the conception, design of the study and wrote the manuscript. RH and JR collected data. CJ created code for generating stimulus lists with non-repeating stimuli. RH performed the statistical analysis. All authors contributed to manuscript revision, read and approved the submitted version.

## FUNDING

This work was supported by National Institute on Aging Grant R01-AG026308, as well as a James S. McDonnell Foundation Scholar Award to KF. JR was partially supported by NWO Veni grant 275-89-032.

## REFERENCES

- Addante, R. J., Ranganath, C., and Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: evidence of accurate source recognition without recollection. *Neuroimage* 62, 439–450. doi: 10.1016/j.neuroimage.2012.04.031
- Allan, K., Robb, W. G., and Rugg, M. D. (2000). The effect of encoding manipulations on neural correlates of episodic retrieval. *Neuropsychologia* 38, 1188–1205. doi: 10.1016/S0028-3932(00)00013-0
- Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1

- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289. doi: 10.1016/j.tics.2007.05.005
- Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243. doi: 10.1098/rstb.2008.0310
- 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
- Brothers, T., Swaab, T. Y., and Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *J. Mem. Lang.* 93, 203–216. doi: 10.1016/j.jml.2016.10.002
- Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010.00025
- Chen, H., Voss, J. L., and Guo, C. (2012). Event-related brain potentials that distinguish false memory for events that occurred only seconds in the past. *Behav. Brain Funct.* 8:36. doi: 10.1186/1744-9081-8-36
- Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Corley, M., MacGregor, L. J., and Donaldson, D. I. (2007). It's the way that you, er, say it: hesitations in speech affect language comprehension. *Cognition* 105, 658–668. doi: 10.1016/j.cognition.2006.10.010
- Craik, F. I., Govoni, R., Naveh-Benjamin, M., and Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *J. Exp. Psychol. Gen.* 125, 159–180. doi: 10.1037/0096-3445.125.2.159
- Craik, F. I., 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
- Craik, F. I., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294. doi: 10.1037/0096-3445.104.3.268
- Cruse, D., and Wilding, E. L. (2009). Prefrontal cortex contributions to episodic retrieval monitoring and evaluation. *Neuropsychologia* 47, 2779–2789. doi: 10.1016/j.neuropsychologia.2009.06.003
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Mem. Cognit.* 28, 923–938. doi: 10.3758/bf03209340
- Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia* 42, 1088–1106. doi: 10.1016/j.neuropsychologia.2003.12.011
- 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
- Curran, T., Tanaka, J. W., and Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18. doi: 10.3758/cabn.2.1.1
- 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
- DeLong, K. A., Quante, L., and Kutas, M. (2014). Predictability, plausibility and two late ERP positivities during written sentence comprehension. *Neuropsychologia* 61, 150–162. doi: 10.1016/j.neuropsychologia.2014.06.016
- DeLong, K. A., Urbach, T. P., Gropp, D. M., and Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology* 48, 1203–1207. doi: 10.1111/j.1469-8986.2011.01199.x
- Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., and Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp.* 15, 95–111. doi: 10.1002/hbm.10010
- Duffy, S. A., Henderson, J. M., and Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 791–801. doi: 10.1037/0278-7393.15.5.791
- Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., and Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proc. Natl. Acad. Sci. U S A* 94, 5973–5978. doi: 10.1073/pnas.94.11.5973
- Ehrlich, S. F., and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *J. Verbal Learn. Verbal Behav.* 20, 641–655. doi: 10.1016/s0022-5371(81)90220-6
- Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x
- Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495. doi: 10.1006/jmla.1999.2660
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84. doi: 10.1016/j.brainres.2006.06.101
- Fields, E. C., and Kuperberg, G. R. (2018). Having your cake and eating it too: flexibility and power with mass univariate statistics for ERP data. *PsyArXiv [Preprint]*. doi: 10.31234/osf.io/qfkgc
- Fischler, I., and Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *J. Verbal Learn. Verbal Behav.* 18, 1–20. doi: 10.1016/s0022-5371(79)90534-6
- Frisson, S., Harvey, D. R., and Staub, A. (2017). No prediction error cost in reading: evidence from eye movements. *J. Mem. Lang.* 95, 200–214. doi: 10.1016/j.jml.2017.04.007
- Frisson, S., Rayner, K., and Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 862–877. doi: 10.1037/0278-7393.31.5.862
- 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
- Geng, H., Qi, Y., Li, Y., Fan, S., Wu, Y., and Zhu, Y. (2007). Neurophysiological correlates of memory illusion in both encoding and retrieval phases. *Brain Res.* 1136, 154–168. doi: 10.1016/j.brainres.2006.12.027
- Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x
- Hartman, M., and Hasher, L. (1991). Aging and suppression: memory for previously relevant information. *Psychol. Aging* 6, 587–594. doi: 10.1037/0882-7974.6.4.587
- Hasher, L., Quig, M. B., and May, C. P. (1997). Inhibitory control over no-longer-relevant information: adult age differences. *Mem. Cognit.* 25, 286–295. doi: 10.3758/bf03211284
- Hayama, H. R., Johnson, J. D., and Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: specific or non-specific? *Neuropsychologia* 46, 1211–1223. doi: 10.1016/j.neuropsychologia.2007.11.021
- Hayama, H. R., and Rugg, M. D. (2009). Right dorsolateral prefrontal cortex is engaged during post-retrieval processing of both episodic and semantic information. *Neuropsychologia* 47, 2409–2416. doi: 10.1016/j.neuropsychologia.2009.04.010
- Hess, D. J., Foss, D. J., and Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *J. Exp. Psychol. Gen.* 124, 62–82. doi: 10.1037/0096-3445.124.1.62
- Hillyard, S. A., and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. U S A* 95, 781–787. doi: 10.1073/pnas.95.3.781
- Hopf, J. M., Vogel, E., Woodman, G., Heinze, H. J., and Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *J. Neurophysiol.* 88, 2088–2095. doi: 10.1152/jn.2002.88.4.2088
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang. Linguist. Compass* 2, 647–670. doi: 10.1111/j.1749-818x.2008.00072.x
- Kučera, H., and Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299
- Kutas, M., DeLong, K. A., and Smith, N. J. (2011). “A look around at what lies ahead: prediction and predictability in language processing,” in *Predictions in the Brain: Using Our Past to Generate a Future*, ed. M. Bar (Oxford: Oxford University Press), 190–207.

- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0
- Lorsbach, T. C., Wilson, S., and Reimer, J. F. (1996). Memory for relevant and irrelevant information: evidence for deficient inhibitory processes in language/learning disabled children. *Contemp. Educ. Psychol.* 21, 447–466. doi: 10.1006/ceps.1996.0030
- Luke, S. G., and Christianson, K. (2016). Limits on lexical prediction during reading. *Cogn. Psychol.* 88, 22–60. doi: 10.1016/j.cogpsych.2016.06.002
- Mangun, G. R., and Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 1057–1074. doi: 10.1037/0096-1523.17.4.1057
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Morcom, A. M. (2015). Resisting false recognition: an ERP study of lure discrimination. *Brain Res.* 1624, 336–348. doi: 10.1016/j.brainres.2015.07.049
- Neely, J. H., and Keefe, D. E. (1989). Semantic context effects on visual word processing: a hybrid prospective-retrospective processing theory. *Psychol. Learn. Motiv.* 24, 207–248. doi: 10.1016/s0079-7421(08)60538-1
- Otten, M., and Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes* 45, 464–496. doi: 10.1080/01638530802356463
- Paller, K. A., and Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *J. Cogn. Neurosci.* 4, 375–392. doi: 10.1162/jocn.1992.4.4.375
- Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2011). *AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components*. San Diego, CA: Technical Report, Swartz Center for Computational Neuroscience.
- 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
- Röer, J. P., Bell, R., Körner, U., and Buchner, A. (2019). A semantic mismatch effect on serial recall: evidence for interlexical processing of irrelevant speech. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 515–525. doi: 10.1037/xlm0000596
- Rommers, J., and Federmeier, K. D. (2018a). Predictability's aftermath: downstream consequences of word predictability as revealed by repetition effects. *Cortex* 101, 16–30. doi: 10.1016/j.cortex.2017.12.018
- Rommers, J., and Federmeier, K. D. (2018b). Lingering expectations: a pseudo-repetition effect for words previously expected but not presented. *Neuroimage* 183, 263–272. doi: 10.1016/j.neuroimage.2018.08.023
- Rugg, M. D., Allan, K., and Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *J. Cogn. Neurosci.* 12, 664–678. doi: 10.1162/089892900562291
- Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257. doi: 10.1016/j.tics.2007.04.004
- Rugg, M., Mark, R., Walla, P., Schloerscheidt, A., Birch, C., and Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature* 392, 595–598. doi: 10.1038/33396
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics* 2, 110–114. doi: 10.2307/3002019
- Schubert, R. E., and Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 27–36. doi: 10.1037/0096-1523.3.1.27
- Schubert, R. E., Spoehr, K. T., and Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Mem. Cognit.* 9, 68–77. doi: 10.3758/bf03196952
- Schwanenflugel, P. J., and LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 344–354. doi: 10.1037/0278-7393.14.2.344
- Schwanenflugel, P. J., and Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *J. Mem. Lang.* 24, 232–252. doi: 10.1016/0749-596x(85)90026-9
- Simpson, G. B., Peterson, R. R., Casteel, M. A., and Burgess, C. (1989). Lexical and sentence context effects in word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 88–97. doi: 10.1037/0278-7393.15.1.88
- Steffens, M. C., and Mecklenbräuker, S. (2007). False memories: phenomena, theories and implications. *Z. Psychol. J. Psychol.* 215, 12–24. doi: 10.1027/0044-3409.215.1.12
- Thornhill, D. E., and Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *Int. J. Psychophysiol.* 83, 382–392. doi: 10.1016/j.ijpsycho.2011.12.007
- Van Berkum, J. J. (2010). The brain is a prediction machine that cares about good and bad—any implications for neuropsychology? *Ital. J. Linguist.* 22, 181–208. Available online at: <http://linguistica.sns.it/RdL/2010.html>
- Vogel, E. K., and Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology* 37, 190–203. doi: 10.1017/s0048577200981265
- Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychol. Forsch.* 18, 299–342. doi: 10.1007/bf02409636
- Voss, J. L., and Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology* 48, 532–546. doi: 10.1111/j.1469-8986.2010.01085.x
- Wang, W., Li, B., Gao, C., Xiao, X., and Guo, C. (2015). Electrophysiological correlates associated with contributions of perceptual and conceptual fluency to familiarity. *Front. Hum. Neurosci.* 9:321. doi: 10.3389/fnhum.2015.00321
- West, R. F., and Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Dev.* 49, 717–727. doi: 10.1111/j.1467-8624.1978.tb02373.x
- Wills, A. J., Lavric, A., Croft, G. S., and Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: evidence from event-related potentials and eye tracking. *J. Cogn. Neurosci.* 19, 843–854. doi: 10.1162/jocn.2007.19.5.843
- Wlotko, E. W., and Federmeier, K. D. (2007). Finding the right word: hemispheric asymmetries in the use of sentence context information. *Neuropsychologia* 45, 3001–3014. doi: 10.1016/j.neuropsychologia.2007.05.013
- Wlotko, E. W., and Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology* 49, 770–785. doi: 10.1111/j.1469-8986.2012.01366.x
- Wolk, D. A., Schacter, D. L., Lygizos, M., Sen, N. M., Holcomb, P. J., Daffner, K. R., et al. (2006). ERP correlates of recognition memory: effects of retention interval and false alarms. *Brain Res.* 1096, 148–162. doi: 10.1016/j.brainres.2006.04.050
- Woodruff, C. C., Hayama, H. R., and Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Res.* 1100, 125–135. doi: 10.1016/j.brainres.2006.05.019
- Woroch, B., and Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Res.* 1317, 180–191. doi: 10.1016/j.brainres.2009.12.074
- Yu, S. S., and Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Res.* 1320, 74–84. doi: 10.1016/j.brainres.2009.12.071

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

Copyright © 2019 Hubbard, Rommers, Jacobs and Federmeier. 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.



# How In-Group Bias Influences Source Memory for Words Learned From In-Group and Out-Group Speakers

Sara Iacozza<sup>1,2\*</sup>, Antje S. Meyer<sup>1,3</sup> and Shiri Lev-Ari<sup>4</sup>

<sup>1</sup> Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, <sup>2</sup> International Max Planck Research School for Language Sciences, Nijmegen, Netherlands, <sup>3</sup> Radboud University Nijmegen, Nijmegen, Netherlands, <sup>4</sup> Department of Psychology, Royal Holloway, University of London, Egham, United Kingdom

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Rupa Gordon,  
Augustana College, United States  
Efthymia C. Kapnoula,  
The University of Iowa, United States  
Si On Yoon,  
University of Illinois  
at Urbana-Champaign, United States

### \*Correspondence:

Sara Iacozza  
sara.iacozza@mpi.nl

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 07 June 2019

**Accepted:** 21 August 2019

**Published:** 12 September 2019

### Citation:

Iacozza S, Meyer AS and  
Lev-Ari S (2019) How In-Group Bias  
Influences Source Memory for Words  
Learned From In-Group  
and Out-Group Speakers.  
Front. Hum. Neurosci. 13:308.  
doi: 10.3389/fnhum.2019.00308

Individuals rapidly extract information about others' social identity, including whether or not they belong to their in-group. Group membership status has been shown to affect how attentively people encode information conveyed by those others. These findings are highly relevant for the field of psycholinguistics where there exists an open debate on how words are represented in the mental lexicon and how abstract or context-specific these representations are. Here, we used a novel word learning paradigm to test our proposal that the group membership status of speakers also affects how speaker-specific representations of novel words are. Participants learned new words from speakers who either attended their own university (in-group speakers) or did not (out-group speakers) and performed a task to measure their individual in-group bias. Then, their source memory of the new words was tested in a recognition test to probe the speaker-specific content of the novel lexical representations and assess how it related to individual in-group biases. We found that speaker group membership and participants' in-group bias affected participants' decision biases. The stronger the in-group bias, the more cautious participants were in their decisions. This was particularly applied to in-group related decisions. These findings indicate that social biases can influence recognition threshold. Taking a broader scope, defining how information is represented is a topic of great overlap between the fields of memory and psycholinguistics. Nevertheless, researchers from these fields tend to stay within the theoretical and methodological borders of their own field, missing the chance to deepen their understanding of phenomena that are of common interest. Here, we show how methodologies developed in the memory field can be implemented in language research to shed light on an important theoretical issue that relates to the composition of lexical representations.

**Keywords:** in-group bias, novel word learning, source memory, decision bias, lexical representations



## INTRODUCTION

Previous findings have shown that people utilize any cue they have available (e.g., gender, social class) to establish whether or not others are members of their own in-group (e.g., Bargh et al., 2012). Group membership can affect how people process and remember information related to those others, with in-group information receiving more attention and being better remembered than out-group information (Hugenberg et al., 2010; Greenstein et al., 2016). While advantages for in-group members have been reported to affect a wide range of cognitive phenomena (see Xiao et al., 2016; Molenberghs and Louis, 2018 for reviews), they have not been directly tested in the context of language processing and language learning, yet.

Such effects are relevant for models of language processing because they have consequences for an ongoing debate on how words are represented in the mental lexicon. One aspect of this broad issue is how well listeners maintain information that is not strictly linguistic but that relates to the context, such as the social identity of the speaker producing a word. In the memory literature, this type of information is referred to as source memory and it is a topic that has been extensively studied. By using memory tests developed to probe source memory, researchers in the field of psycholinguistics can gain a better understanding of how speaker-related information is encoded in the representations of words and whether the encoding of such information is modulated by social factors, such as the group membership status of the speakers.

The aim of the current study is to investigate the proposal that in-group biases permeate language processing as well, and that they affect the level of detail of speaker-related information that is encoded when learning new words. We propose that representations of words learned from in-group members are more likely to contain highly specific speaker-related information, as compared to representations of words learned from out-group members, and that such differences are in turn influenced by how strongly each learner prefers their in-group members over out-group members.

Before turning to the current study, we review the relevant literature. We start by reporting evidence that shows that the social identity of the speaker affects how listeners process language. We then describe existing exemplar-based theories of language processing that provide a theoretical framework for understanding effects of speaker identity on language processing. We then point to a potential limitation of these models, namely, their tendency to assume that the speech of all speakers is treated equally. We propose that existing models should integrate parameters that allow different degrees of encoding specificity and assigning different weight to linguistic input depending on speaker group membership status. Specifically, we propose that linguistic information provided by in-group speakers is encoded in more detail than information from out-group speakers. We motivate our proposal with evidence from non-linguistic studies in social psychology that report group membership effects on memory and information processing.

Previous research indicates that when interacting with others, information about their social identity is rapidly extracted

(see Bargh et al., 2012, for a review) and can influence people's attitudes and preferences toward those others (e.g., Greenwald and Banaji, 1995; Jones and Fazio, 2010; Kinzler et al., 2011). There exists diverse evidence showing that others' social identity can influence how listeners process language. For instance, it has been shown that, when a speaker's social identity is made available via the speaker's voice, listeners take the identity into consideration and have particular expectations about what will likely be said. If these expectations are not met, such as when the desire of looking like Britney Spears is reported in a man's voice, language processing becomes harder (Van Berkum et al., 2008, see also Walker and Hay, 2011; Martin et al., 2016). Similarly, speaker social identity can affect how listeners perceive speech sounds (e.g., Johnson et al., 1999; Niedzielski, 1999; Hay et al., 2006a,b). For example, changing listeners' expectations of a speaker's place of residence affected their responses in a diphthong identification task. Participants reported hearing what they believed to be more representative of the supposed speaker's linguistic community, independent of the actual linguistic input, which was identical across the two conditions (Niedzielski, 1999). This suggests that information about the speaker affects speech perception.

In short, this body of evidence shows that information related to the speaker's identity is extracted along with the linguistic input and can influence the processing of the latter. Existing exemplar-based models of speech processing argue that the reason that social information is used in language processing is because it is encoded along with linguistic input. These models state that linguistic experiences are encoded as rich episodic memories (i.e., exemplars) (e.g., Hay et al., 2006a; Goldinger, 2007; Nielsen, 2011; see Drager and Kirtley, 2016, for a review) that contain information which is both language-specific (e.g., includes phonetic, lexical, and syntactic details) and context-specific (e.g., includes pragmatics, speakers' characteristics) (e.g., Drager and Kirtley, 2016).

Recently, in a new model by Münster and Knoeferle (2018) the contributions of encoding speakers and listeners' characteristics during on-line language processing were formally defined. Grounded on a large body of empirical evidence, the model posits that comprehending language in context, by, for instance, extracting both speaker-specific and language-specific input in tandem, may speed up and/or ease comprehension. For example, consider a scenario in which the utterance "Every evening I drink some wine before I go to sleep" is produced by an adult speaker. Based on the age of the speaker, listeners can build up probabilistic inferences about what is more likely to follow the verb *drink* (e.g., in the case of an adult, the word *wine* is more probable than the word *milk*). By pre-activating lexical items that are more probable, listeners can easily make sense of the new piece of information, i.e., the word *wine*, speeding up comprehension (see Münster and Knoeferle, 2018 for details).

Crucially, Sumner et al. (2014) proposed that the social context might not only be encoded with the linguistic input but might modulate the strength of its encoding. In support of this account, Sumner et al. (2014) showed that idealized phonetic variants are encoded with greater weight than common, therefore more frequent, phonetic variants. According to their model, phonetic variants with higher prestige (i.e., idealized



ones) receive an advantage in representation and processing as compared to variants characterized by lower prestige. Extending their theory to more general linguistic processes, one could hypothesize that people would encode linguistic variations more strongly if they are associated with contexts and speakers that have a special status.

Here, we propose that learning new words from speakers that are ascribed a special status might lead to lexical representations that are richer in contextual information (e.g., speaker-related information), as compared to representations of words learned from speakers without a special status. An example of speakers that are ascribed a special status is the case of in-group members. Indeed, there is evidence suggesting that group membership influences input processing and learning. For instance, memory is usually better for in-group faces than out-group faces (e.g., Van Bavel et al., 2008; Hugenberg et al., 2010) and for information delivered by in-group than by out-group members (e.g., Frable and Bem, 1985; Wilder, 1990). Furthermore, people learn better and process more quickly new associations between previously neutral stimuli (e.g., geometrical shapes) and in-group membership (e.g., the logo of their favorite football club) than associations involving out-group membership (Moradi et al., 2015; Enock et al., 2018).

One way in which in-group biases may work is via the recruitment of additional cognitive resources (Meissner et al., 2005; Van Bavel and Cunningham, 2012). Such additional resources have been suggested to lead to in-group representations that are characterized by a higher level of detail than out-group representations. For example, when processing in-group related information, people were shown to encode the source of information in more detail than when the information was related to out-group members. This resulted in them being better in a source memory task when identifying in-group sources than out-group sources (e.g., Greenstein et al., 2016), suggesting that being exposed to in-group membership boosts the encoding of individual-specific information (see Hugenberg et al., 2010, for a similar account).

No study has tested whether lexical representations for the same words can depend on the identity of the speaker that tends to use them. If this is the case, this will have implications for language learning, language processing, and linguistic representations. It would extend current theories that examine the role of input and its distribution in language acquisition and representation by showing that the same distribution can have different effects depending on who are the speakers that provide different tokens in the input. As a first step, the current study was designed to investigate which social information learners encode when they learn new words from speakers who either belonged to the learners' social group (i.e., in-group members) or did not do so (i.e., out-group members).

## The Current Study

We hypothesize that listeners encode the social identity of the speakers from whom they learn novel words and that the social identity influences how detailed speaker-specific information is encoded. To test these predictions, we carried out the current study in which we examined participants' source memory for

words learned from speakers from different social groups. In the Main Experiment, participants were exposed to a learning context in which they learned new words from speakers who supposedly shared their university affiliation (i.e., in-group speakers) and from speakers with a different affiliation (i.e., out-group speakers). In the Control Experiment, participants learned from two groups of speakers who supposedly attended two foreign universities. Since in the Control Experiment the group membership was not manipulated, because both universities were unrelated to the participants, we could check that the patterns hypothesized to be found in the Main experiment were indeed a reflection of the social saliency ascribed to speakers' group membership and not simply a consequence of the contrastive nature of our manipulation (i.e., teaching competing labels spoken by different groups of speakers).

During the word learning task, all participants in both experiments learned novel labels for uncommon gadgets. Crucially, target gadgets received two competing but equally fitting labels, one from a speaker of each affiliation (e.g., *citrus-peeler* vs. *citrus-schiller*, in English lemon peeler vs. lemon stripper). Afterward, source memory for these words was tested in a recognition memory test. Participants were shown one speaker and one label at a time and asked if the speaker had produced the label in the previous phase (i.e., forced choice: yes/no). Lastly, we collected participants' implicit in-group bias (see "Materials and Methods" section for details).

In the Main experiment, we predicted that participants would spontaneously monitor the speakers' group membership status. Consequently, when asked to recognize the source of the new words, we expected participants to remember speaker social group but to struggle remembering the exact speaker that produced each word. Therefore, they should be more likely to misattribute words to incorrect speakers within the same affiliation than between different affiliations, i.e., there should be source memory confusion. Following our hypothesis about different levels of detail depending on social salience and group membership of the speakers, we predicted that words learned from in-group speakers would contain a higher level of detail about who produced them, compared to words learned from out-group speakers. This would result in in-group linguistic representations that are more speaker-specific and less prone to source memory confusion than out-group representations. Crucially, this in-group advantage should be stronger for participants exhibiting stronger in-group bias. This pattern is expected to result in a significant interaction involving speaker group membership and individual in-group bias. In the Control Experiment, we expected no differences between the two speaker affiliations. This would show that differential processing of information learned from different groups is specific to cases where group membership is socially salient.

## MATERIALS AND METHODS

### Participants

One-hundred-twenty-four native Dutch speakers (age range: 18–26 years) participated in the study after providing their

informed consent, as approved by the Ethics committee of the Social Sciences department of the Radboud University Nijmegen (project code: ECSW2014-1003-196). All participants were students or recent graduates of Radboud University Nijmegen. All participants were female, as were the speakers from whom they learned the labels. This was done to avoid that an additional social dimension (i.e., gender) of in-group status could interact with the one we manipulated (i.e., academic affiliation). Participants were randomly assigned to either the Main Experiment ( $n = 62$ ) or the Control Experiment ( $n = 62$ ).

## Materials

### Materials for the Word Learning Task

#### Speakers

Eight fictitious speakers were created by pairing female faces selected from the Chicago Face Database (Ma et al., 2015) with the voices of native Dutch female speakers recorded in our laboratory. Prior to the experiment, voices were matched for perceived typicality and attractiveness (paired  $d$ -tests,  $p$ s > 0.05) via a norming on-line survey in which twenty different participants participated. Each speaker was a unique combination of one face and one voice, consistent across participants. Speakers' academic affiliation was randomized across participants and indicated by the logo of the supposed affiliation displayed underneath the photo.

#### Affiliation logos

For the Main Experiment, original-color pictures of the logos of the Radboud University Nijmegen (i.e., in-group affiliation) and the ROC Nijmegen (i.e., out-group affiliation) were used. For the Control Experiment, original-color pictures of the logos of Pisa and Florence universities were used.

#### Objects and labels

Twenty-four images of unfamiliar gadgets (e.g., lemon peeler) and their corresponding labels were selected via a norming study (see **Supplementary Appendix 1** for details). Half of the gadgets, hereinafter referred to as target gadgets, were presented with two competing labels, which were equated for goodness-of-fit and frequency. The other 12 gadgets were presented with a single label and served as fillers. All labels were produced by each speaker and audio-recorded.

### Materials for the Individual In-Group Bias Task

#### Affiliation logos

The same logos used in the word learning task were used here.

#### Geometrical shapes

Black shapes for triangle, square, and circle were used.

## Procedure

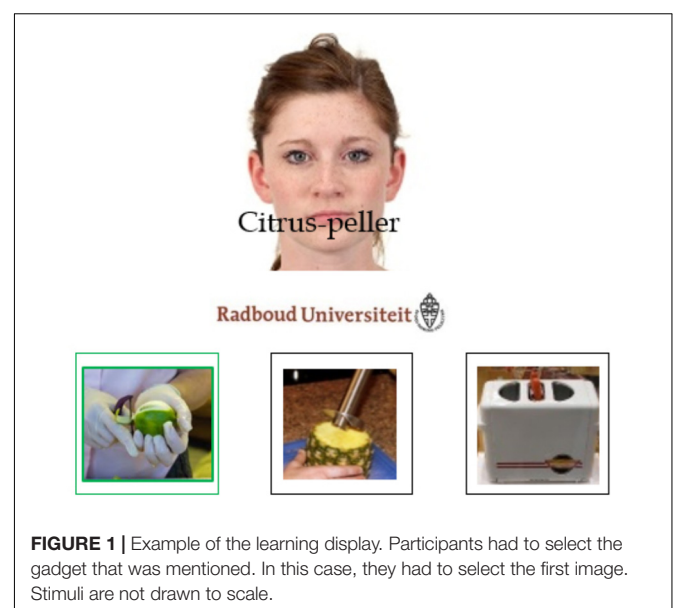
### Word Learning Task

The word learning task consisted of an exposure phase and a test phase. The exposure phase was presented as a communication task in which participants were instructed to pay attention to all the stimuli presented (i.e., faces, gadgets and labels) and select gadgets based on what the speakers said, with no explicit reference to the academic logos. Participants saw 24

gadgets, each named by speakers of both groups. Half were target gadgets, for which the two groups of speakers provided competing, but equally fitting, labels, whereas the other half were fillers, for which unique labels were provided. Fillers were included to minimize participants' awareness of the nature of the experimental manipulation (i.e., the contrastive nature of the labels). Note that not all speakers referred to all the gadgets. In fact, each gadget was only labeled by two of the eight speakers (one per group of speakers). Speaker group affiliation, speaker-label pairing, and label-group affiliation pairings were fully randomized per participant. On each trial, a photo of a speaker, together with the corresponding affiliation logo, was displayed (800 ms). Then, while the photos of speaker and logo were still on screen, the audio-recording related to the gadget label was played. Simultaneously, the written form of the label was superimposed upon the speaker's mouth (1500 ms). Next, three gadgets appeared on the screen and participants selected the one that fit the audio and the written label (see **Figure 1** for an example of the learning display<sup>1</sup>). If the response was wrong, the audio was repeated. Two exposure blocks were administered with half of the gadgets (i.e., six fillers and six targets) introduced in the first block, and the other half introduced in the second block. The gadgets were randomly allocated in the first or second exposure block per participant. Three exposure rounds were administered per block so that each display was repeated three times, once per round, in a randomized trial order.

After each exposure block, participants performed a surprise source-memory recognition test on the gadgets introduced in the preceding exposure block only. In each trial, they saw a photo of a speaker with their affiliation logo and a written label (see **Figure 2**). Participants indicated whether the speaker had produced the label in the previous exposure phase via key press

<sup>1</sup>Due to copyright issues, none of the pictures of the gadgets in the example corresponds to actual stimuli, but they provide a good approximation of the type of stimuli we used.



**FIGURE 1** | Example of the learning display. Participants had to select the gadget that was mentioned. In this case, they had to select the first image. Stimuli are not drawn to scale.



(i.e., forced choice: yes/no). Decisions were self-paced. Across the two memory test blocks, there were 288 trials in which all possible speaker-label pairings were shown. Of those 288 trials, 96 were filler-related trials (subsequently excluded from the analyses) and 192 were trials in which target gadgets were shown. Of the 192 target-gadget trials, 24 were matching trials (i.e., the speaker had indeed produced the label) and 168 mismatching trials (i.e., the label had not been used by the speaker). Of those mismatching trials, 72 were within-affiliation mismatching trials (showing a label along with a wrong speaker from the same affiliation as the correct one), 72 were between-affiliation mismatching trials (showing a label a speaker from the wrong affiliation). The remaining 24 trials showed a speaker with a label that competed with the one she used (e.g., the speaker that had used “citrus-schiller” was displayed with “citrus-peller”). They were only included to make all possible speaker-label combinations available, but they were not analyzed. Note that in all mismatching trials, the correct answer was that the pairing was incorrect because the speaker depicted in the photo had not used the displayed label in the exposure task.

### Implicit In-Group Bias Task

Participants' individual in-group bias was measured in a perceptual matching task (Moradi et al., 2015), which has been shown to provide results that are reliable within individuals and across different test sessions (Stolte et al., 2017). Three geometric shapes (circle, square, triangle) were randomly paired with logos of three academic affiliations. For the Main experiment, the

logos depicted the in-group university – the Radboud University Nijmegen, and two out-group affiliations – the ROC Nijmegen and Tilburg University. To keep the two experiments comparable, participants in the Control experiment performed the task with logos of the Italian universities that appeared in the word learning task (Pisa and Florence) and a third Italian university, Bologna. Each association was initially presented ten times. Then, participants performed a practice block of 24 trials, followed by two blocks of 120 experimental trials each. In both practice and test trials, a fixation cross (500 ms) preceded a blank screen (between 1000 and 2000 ms) and the simultaneous presentation of logo and shape (600 ms), following the timings utilized in Moradi et al. (2015). Participants had 1500 ms to judge the accuracy of the pairing. Feedback was given only during practice. In-group bias in this task is usually indexed by faster and more accurate responses for stimuli that are newly associated with in-group membership compared to stimuli associated with out-group membership (e.g., Moradi et al., 2015).

## RESULTS

All analyses were performed with mixed-effects modeling as implemented in the lme4 package (version 1.1-15; Bates et al., 2014) in R (R Core Team, 2016) and the models' random structures were determined following the procedure suggested by Bates et al. (2015).

Before turning to the main analyses from the source memory test, we performed a sanity check to confirm that, at the group level, participants in the Main experiment showed the expected in-group bias in the perceptual matching task used to extract individual in-group bias measures.

### Group-Level In-Group Bias

#### Analyses Over RTs

Prior to analyses, trials with incorrect responses or with RTs faster than 200 ms or slower than 2100 ms were excluded. For these sanity-check analyses, we selected only matching trials (i.e., in which the logo of the university was displayed with the associated geometrical shape) which referred to the in-group university and the out-group university used in the study (i.e., the ROC Nijmegen). We then performed an outlier removal procedure by removing trials with RTs 2.5 SDs or higher from the mean per condition, per participant. The resulting dataset was analyzed using linear mixed-effect model in which log(10)-transformed RTs were predicted by the fixed effect for Group Membership (In-group vs. Out-group, reference level: In-group). We added per-participant random intercept and by-participant random slope for Group Membership. Results confirmed the usual patterns for this task: participants were faster at recognizing in-group-related associations than out-group-related associations (in-group: mean = 709 ms, SD = 212 vs. out-group: mean = 754 ms, SD = 199;  $\beta = -0.01$ , SE = 0.003,  $t = -5.03$ ,  $p < 0.0001$ ).

#### Analyses Over Accuracy

As with the RT analysis, the analysis included only matching trials (i.e., trials in which the logo of the university was displayed with

the associated geometrical shape) which referred to the in-group university and the out-group university used in the study (i.e., the ROC Nijmegen). Accuracy was analyzed using a logistic mixed-effect model with a fixed effect for Group Membership (In-group vs. Out-group, reference level: In-group). We added per-participant random intercept and by-participant random slope for Group Membership. Results confirmed the usual patterns for this task: participants were better at recognizing in-group-related associations than out-group-related associations (in-group: mean = 94.70%, SD = 22.4 vs. out-group: mean = 92.88%, SD = 25.72;  $\beta = 0.4$ , SE = 0.1,  $t = 3.17$ ,  $p < 0.01$ ).

The analyses confirmed that in the Main experiment, at the group-level, participants showed a strong in-group bias for their own university. Successively, we extracted individual measures of in-group bias by calculating a per-participant measure of effect size, namely Cohen's  $d$ , from both accuracy and RTs over in-group versus out-group matching trials. The measure calculated over RTs was not a significant predictor in any of the models we ran; thus, we will focus on the measure derived from accuracy.

Next, the results from the Main and Control experiments are presented separately because the in-group vs. out-group contrast only applies to the former experiment. The data from each experiment was analyzed following the outlined steps: (1) planned analyses on matching and mismatching trials, separately; and (2) *post hoc* analyses over  $d$ -prime and response bias values.

## Main Experiment

After each exposure round in the word learning task, participants were tested with a recognition memory test. In this test, they were presented with matching or mismatching speaker-label pairings and had to decide via key press if the label had or had not been produced by the speaker. We carried out analyses over matching and mismatching trials separately. We predicted that participants would show more accurate source memory of in-group labels, as compared to out-group labels, and that such advantage would be modulated by participants' own in-group biases.

### Matching Trials

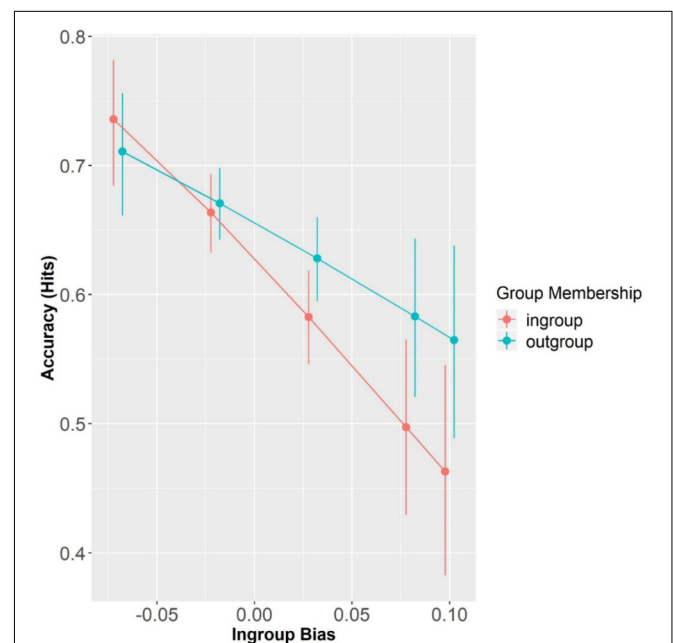
To test whether source memory was better for in-group than for out-group words, we ran a logistic mixed effects model with accuracy as the dependent measure and fixed effects for Group Membership (In-group vs. Out-group, reference level: In-group), In-group Bias (centered continuous predictor), and their interaction. Block (Block1 vs. Block2, reference level: Block1) was included as covariate to control for potential confounds.<sup>2</sup> We added per-participant and per-items random intercepts and a by-participant slope for Group Membership.

Overall, participants' accuracy in the matching trials was 63.08% (SD = 48.28) and above chance level, as confirmed by a one-sample  $t$ -test (i.e., 50%) ( $t = 10.41$ ,  $p < 0.001$ ). Results

showed that neither Group Membership ( $\beta = 0.10$ , SE = 0.13,  $z = 0.75$ ,  $p = 0.45$ ) nor its interaction with In-group Bias significantly predicted accuracy ( $\beta = 3.13$ , SE = 3.23,  $z = 0.97$ ,  $p = 0.33$ ). Participants' accuracy did not differ between Block1 and Block2 ( $\beta = 0.02$ , SE = 0.11,  $z = 0.19$ ,  $p = 0.34$ ). However, participants' In-group Bias significantly predicted accuracy, but only at the reference level, i.e., in-group membership ( $\beta = -6.90$ , SE = 3.17,  $z = -2.18$ ,  $p < 0.05$ ). By re-leveling Group Membership with Out-group as the reference level, we saw that accuracy for out-group speaker-label pairs was not modulated by the individual measure of In-group Bias ( $\beta = -3.76$ , SE = 2.93,  $z = -1.29$ ,  $p = 0.20$ ) (see **Figure 3**). This means that the more in-group biased participants were, the less accurate they were at recognizing speaker-label pairs, in particular when the speaker-label pairs were of their in-group.

### Mismatching Trials

To test whether speaker group membership influenced the level of detail for speaker-specific information encoded with the new words, we analyzed accuracy on mismatching trials. By looking at participants' performance on within-affiliation mismatching trials, where labels were paired with incorrect speakers but belonging to the same affiliation as the correct source, we were able to test whether the source-related information for novel words was speaker-specific (participants should have rejected the wrong source) or group-specific (participants would have incorrectly accepted the wrong source). We hypothesized that people would encode more speaker-specific information with in-group labels than with out-group labels. We therefore predicted greater confusion among out-group speakers than among in-group speakers in the within-affiliation mismatching



**FIGURE 3** | Accuracy (Hits) as a function of Group Membership and In-group Bias (centered). Error bars represent standard errors.

<sup>2</sup>To ensure that the patterns of results were comparable across both testing blocks, we also ran a mixed-effect model where response accuracy was modeled by Group Membership (In-group vs. Out-group, reference level: In-group), In-group Bias (centered continuous predictor), Block (Block1 vs. Block2, reference level: Block1) and their interactions. We added per-participant and per-items random intercepts and a by-participant slope for Group Membership. Results from this analysis showed that neither the main effect of Block ( $p = 0.37$ ) nor its interactions with the other variables ( $ps > 0.16$ ) significantly predicted response accuracy.



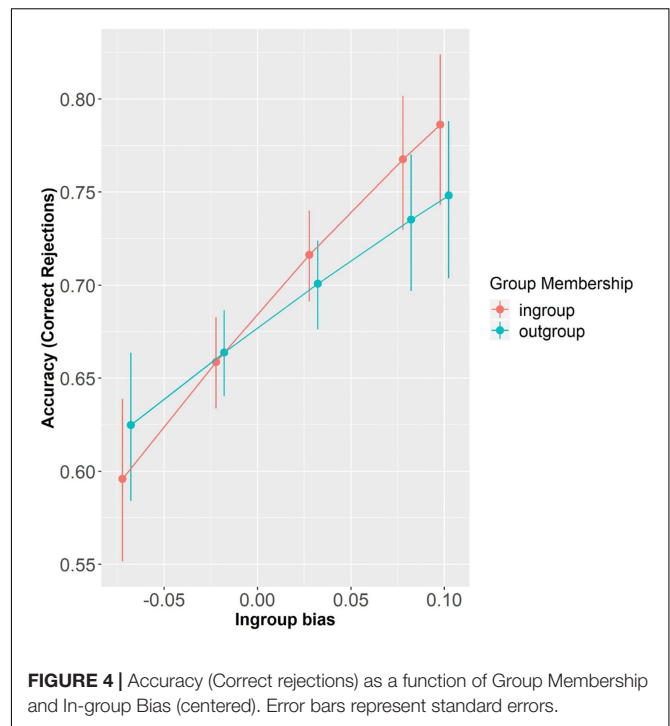
trials. We also predicted that this difference in accuracy would depend on individual In-group Bias, such that the greater In-group Bias participants exhibited, the greater difference they should show between in-group vs. out-group trials. Conversely, in between-affiliation mismatches (i.e., where an in-group label was shown with out-group members, and vice versa) no differences were expected.

To test these hypotheses, we ran a logistic mixed model analysis with fixed effects for Mismatch Type (Within- vs. Between-affiliation, reference level: Within), Group Membership (In-group vs. Out-group, reference level: In-group), In-group Bias (centered continuous measure), and their interaction terms. We added Block as covariate, per-participant and per-item random intercepts and by-participant slopes for Group Membership and Mismatch Type.

Overall, participants' accuracy on mismatching trials was 65.79% (SD = 47.45) and above chance level (i.e., 50%), as confirmed by a one-sample *t*-test ( $t = 31.31$ ,  $p < 0.001$ ). As expected, participants were more accurate for between-affiliation mismatches than for within-affiliation mismatches ( $\beta = 0.53$ , SE = 0.14,  $z = 3.10$ ,  $p < 0.0001$ ; mean = 70.35%, SD = 45.68 and mean = 61.22%, SD = 48.73, respectively). This shows that participants encoded speakers' affiliations. Due to a practice effect, they were also more accurate in Block2 than in Block1 ( $\beta = 0.79$ , SE = 0.05,  $z = 15.95$ ,  $p < 0.0001$ ; mean = 73.61%, SD = 44.08 and mean = 58.09%, SD = 49.35, respectively). Participants' performance was also significantly predicted by In-group Bias at the reference levels ( $\beta = 7.98$ , SE = 3.03,  $z = 2.64$ ,  $p < 0.01$ ) and by a marginally significant interaction of In-group Bias with Group Membership ( $\beta = -3.52$ , SE = 1.96,  $z = -1.80$ ,  $p = 0.07$ ), which suggests that participants with different strengths of In-group Bias were differently affected by speaker Group Membership. Specifically, simple effect analyses revealed that the larger the In-group Bias, the better participants were at correctly rejecting pairings involving the in-group membership ( $\beta = 7.98$ , SE = 3.03,  $z = 2.64$ ,  $p < 0.01$ ). On the other hand, participants' In-group Bias did not predict their performance with pairings involving the out-group membership ( $\beta = 4.46$ , SE = 2.78,  $z = 1.6$ ,  $p = 0.11$ ) (see **Figure 4**).

Furthermore, neither the two-way interaction between In-group bias and Mismatch Type ( $\beta = -5.21$ , SE = 3.48,  $z = -1.5$ ,  $p = 0.13$ ), nor the three-way interaction between Mismatch Type, Group Membership, and In-group Bias reached significance ( $\beta = 3.10$ , SE = 2.54,  $z = 1.22$ ,  $p = 0.22$ ). Therefore, participants' performance in both between- and within-affiliation mismatches was comparably affected by the Group Membership  $\times$  In-group bias interaction.

In short, results from the matching trials revealed a *negative* relationship between In-group Bias and response accuracy, especially for in-group pairings. This pattern suggests that participants with stronger in-group bias were more likely to produce misses with in-group speaker-label pairs. On the other hand, results from the mismatching trials revealed a *positive* relationship between In-group Bias and accuracy, meaning that those strongly biased participants also produced fewer false alarms when in-group pairings were involved. These seemingly contradictory results can be reconciled by



stepping away from simple accuracy analyses and by relying on signal detection theory measurements which capture detection sensitivity (namely, *d*-prime) and response bias (namely, *C*).

## D-Prime and C Values

Analyses over *d*-prime and *C* measures allow us to test whether participants' sensitivity and response bias during decision making processes differed for in-group vs. out-group related decisions. We calculated two *d*-prime values and two *C* values per participant for in-group and out-group trials separately. In order to generate values that reflected participants' decisions to purely in-group or out-group trials, *d*-prime and *C* values were calculated from participants' performance in matching trials (i.e., hit rates) and within-affiliation mismatching trials (i.e., false-alarm rates).<sup>3</sup> Between-affiliation mismatches were not considered for these analyses because they were created by having an element (either label or speaker) from each group and were therefore not purely in-group or out-group related. We ran two linear mixed-effect models with either *d*-prime or *C* values as the dependent variable and Group Membership (In-group vs. Out-group, reference level: In-group), In-group Bias and their interaction as fixed effects. The models included per-participant random intercepts.

The model that explored the relationship between individual *d*-prime and the independent variables showed no significant main effects or interactions ( $ps > 0.57$ ), suggesting that participants' sensitivity was not modulated by speaker Group

<sup>3</sup>To calculate *C* and *d*-prime values, we firstly followed Macmillan and Creelman (2004) and converted 0 values in False Alarms to  $1/2N$  and 1 values in Hit rates to  $1-1/2N$ . Next, we subtracted the *z*-scored False Alarms rate from the *z*-scored Hit rate. *C* values, were calculated using the following formula:  $(-0.5) \times (z\text{-scored (Hit\_Rate)} + z\text{-scored (False\_Alarms\_rate)})$ .



Membership or their own In-group Bias, nor the interaction between them (see **Figure 6A**).

On the other hand, the model exploring C values showed a significant main effect of In-group Bias ( $\beta = 8.60$ ,  $SE = 2.55$ ,  $t = 3.38$ ,  $p < 0.001$ ) so that the more in-group biased, the more conservative participants were in their decision (i.e., having a bias for “no” responses). Importantly, there was a significant interaction between In-group bias and Group Membership ( $\beta = -4.26$ ,  $SE = 1.97$ ,  $t = -2.16$ ,  $p < 0.05$ ), showing that participants with different In-group Bias strength were differently affected by speaker Group Membership. Simple effect analyses revealed that while In-group Bias strongly modulated participants’ response bias with in-group labels ( $\beta = 8.60$ ,  $SE = 2.55$ ,  $t = 3.38$ ,  $p < 0.001$ ), this was only marginally so with out-group labels ( $\beta = 4.34$ ,  $SE = 2.55$ ,  $t = 1.71$ ,  $p = 0.09$ ). These findings show that participants differed in their response bias as a function of Group Membership and In-group Bias, so the more in-group biased they were, the more conservative they were in their in-group related decisions, as compared to out-group related decision (see **Figure 5**). In other words, they were more careful in attributing in-group words to any in-group speaker.

## Control Experiment

We hypothesized that the tendency to monitor speaker social identity was dependent on whether the affiliations were perceived as socially salient, or relevant. To test this, we ran a control experiment in which participants learned new words from Dutch native students attending two Italian universities, as part of an exchange program. In this experiment, group membership was not manipulated. Participants still learned from two groups of speakers, like in the Main Experiment, but here the speakers’

affiliations were supposed to be socially neutral because the speakers belonged to two foreign universities. Therefore, no differences were expected between the two groups. To control for potential visual dissimilarities between the logos used, participants performed the same perceptual matching task as in the Main experiment, responding to pairings involving the logos of the Italian universities. Similar to what we did in the Main experiment, we calculated an individual measure that in this case can be seen as an index of Visual Bias. This individual measure was entered in the statistical analyses.

## Matching Trials

We ran a logistic mixed effects model with accuracy as the dependent measure and fixed effects for Affiliation (University1 vs. University2, reference level: University1), Visual Bias (centered), and their interaction. Block was included as covariate to control for potential confounds. We added per-participant and per-items random intercepts and by-participant slope for Affiliation.

Overall, participants’ accuracy in the matching trials was 57.52% ( $SD = 49.48$ ) and above chance level, as confirmed by a one-sample  $t$ -test (i.e., 50%) ( $t = 5.84$ ,  $p < 0.0001$ ). Neither Affiliation, nor Visual Bias or their interaction significantly predicted accuracy ( $ps > 0.27$ ). Participants’ accuracy was better in Block2 than in Block1 ( $\beta = 0.28$ ,  $SE = 0.11$ ,  $z = 2.51$ ,  $p < 0.05$ ).

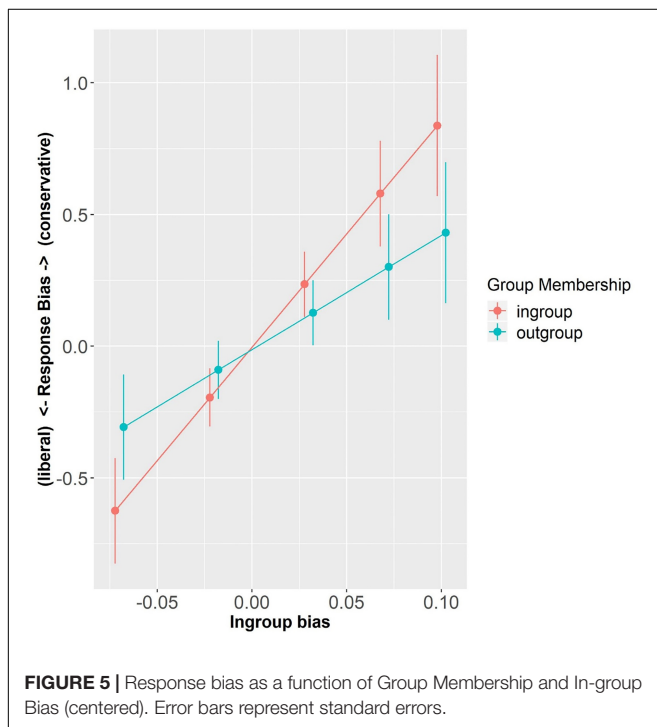
## Mismatching Trials

We ran a logistic mixed model analysis with fixed effects for Mismatch Type (Within- vs. Between-affiliation, reference level: Within), Affiliation (University1 vs. University2, reference level: University1), Visual Bias (centered continuous measure), and their interaction terms. We added Block as covariate, per-participant and per-item random intercepts and by-participant slopes for Affiliation and Mismatch Type.

Overall, participants’ accuracy on mismatching trials was 69.37% ( $SD = 46.10$ ) and above chance level (i.e., 50%), as confirmed by a one-sample  $t$ -test ( $t = 39.53$ ,  $p < 0.0001$ ). Generally, participants were more accurate in the between-affiliation mismatches than in the within-affiliation mismatches ( $\beta = 0.20$ ,  $SE = 0.07$ ,  $z = 2.64$ ,  $p < 0.01$ ; mean = 70.55%,  $SD = 45.59$  and mean = 68.18%,  $SD = 46.58$ , respectively), indicating that even the irrelevant social affiliations were encoded to some degree. Participants were also more accurate in Block2 than in Block1 ( $\beta = 0.92$ ,  $SE = 0.05$ ,  $z = 18.00$ ,  $p < 0.0001$ ; mean = 78.14%,  $SD = 41.33$  and mean = 60.73%,  $SD = 48.84$ , respectively). None of the other main effects or interactions resulted significant ( $ps > 0.16$ ), showing that, unlike the modulating effect of in-group bias in the Main Experiment, participants’ memory for speaker-label pairings was not modulated by Visual Bias.

## D-Prime and C Values

To be consistent, we also performed analyses over d-prime and C values, as we did in the Main experiment. Crucially, we did not expect any differences between the two academic affiliations. We calculated two d-prime values and two C values per participant for the two affiliations separately. We ran two linear mixed-effect models with either d-prime or C values as the dependent



variable and Affiliation University1 vs. University2, reference level: University1), Visual Bias (centered continuous measure), and their interaction terms. The models included per-participant random intercepts.

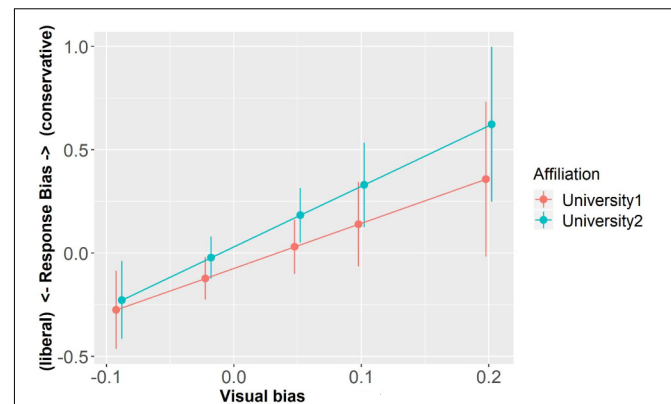
The model that explored the relationship between individual *d*-prime and the independent variables showed no significant main effects or interactions ( $p > 0.67$ ), suggesting that participants' sensitivity was not modulated by speaker Affiliation or their Visual Bias, nor the interaction between them (see **Figure 6B**).

Similarly, the model exploring *C* values showed no significant main effect of Visual Bias or interaction ( $p > 0.24$ ). There was a marginal effect of Affiliation ( $\beta = 0.11$ ,  $SE = 0.06$ ,  $t = 1.90$ ,  $p = 0.06$ ) with decisions made about University2 being numerically more conservative than decisions involving University1 (see **Figure 7**).

## DISCUSSION

We used a novel word learning paradigm to test whether learners of new words monitored speakers' social identity, such as their group and individual identity. Furthermore, we asked whether group membership status of the speakers and individual in-group biases of the learners affected the level of detail of speaker-specific information encoded in the novel lexical representations. We additionally performed a control experiment and ensured that the patterns found in the Main experiment were indeed a reflection of the social saliency ascribed to speakers' group membership and not simply a consequence of the contrastive nature of our manipulation (i.e., teaching competing labels spoken by different groups of speakers).

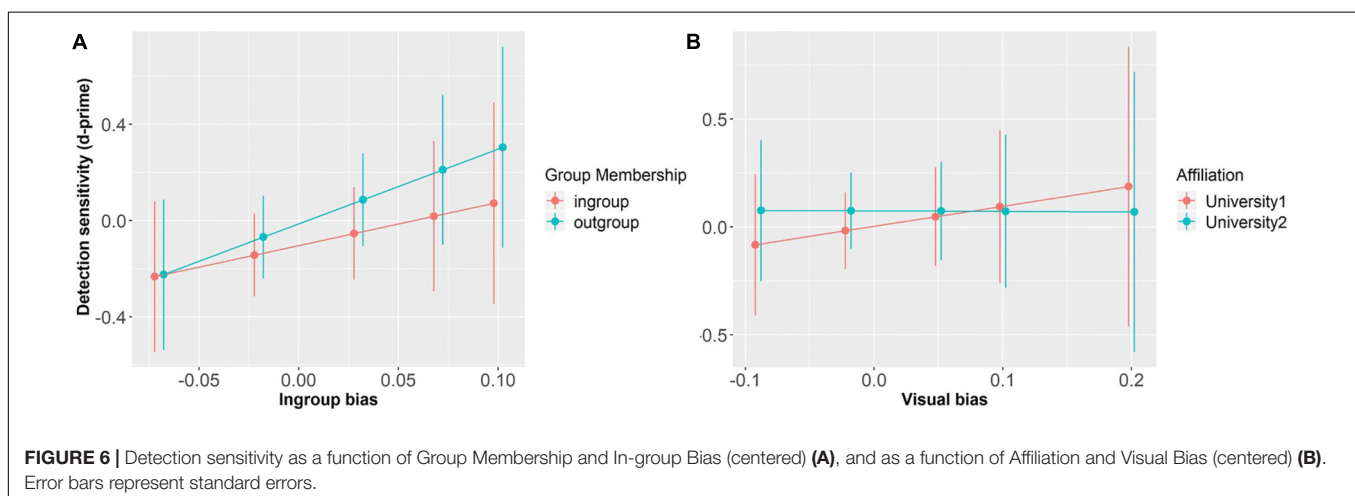
In the test phase of the word learning task, participants' source memory for the new words was tested in an alternative forced-choice task (i.e., yes/no) where they decided whether displayed speaker-label pairs matched or mismatched what they learned in the exposure phase. This task offered a proxy for investigating the level of detail of speaker-specific information in the novel representations. Results confirmed our prediction regarding the general tendency to encode in parallel both linguistic content and



**FIGURE 7 |** Response bias as a function of Affiliation and Visual Bias (centered). Error bars represent standard errors.

speakers' social identity (i.e., speakers' affiliation). This tendency was reflected in the fact that participants made more within-affiliation errors than between-affiliation errors, i.e., source memory confusion. This finding provides further support for models of word learning where linguistic units are encoded together with speaker-related information (exemplar models e.g., Hay et al., 2006a; Goldinger, 2007; Nielsen, 2011; see Drager and Kirtley, 2016, for a review).

Concerning our hypotheses about the effects of Group Membership and In-group Bias, the results revealed a more complex pattern than we had predicted. We had predicted that participants would encode in-group labels with a higher level of detail of speaker-specific information, as compared to out-group labels. This phenomenon was expected to be reflected in (a) a higher proportion of hit rates for matching in-group speaker-label pairs and (b) a higher proportion of correct rejections for within-affiliation in-group speaker-label pairs. Both effects were predicted to be positively modulated by the individual In-group Bias, so that the stronger the bias, the stronger the effects. We found that indeed participants with stronger in-group bias were better at correctly rejecting wrong in-group pairings (i.e., in the



**FIGURE 6 |** Detection sensitivity as a function of Group Membership and In-group Bias (centered) (A), and as a function of Affiliation and Visual Bias (centered) (B). Error bars represent standard errors.

mismatching trials). However, when looking at the matches, the results revealed that those participants with stronger in-group bias were also more likely to miss matching in-group speaker-label pairs.

These seemingly contradictory results are hard to reconcile when relying only on accuracy (i.e., correct/incorrect). For this reason, we relied on signal detection theory measurements, such as *d*-prime and *C* values, to gain a deeper understanding of the phenomenon. These measures capture both hit rates and false-alarm rates for conceptually similar items and allow us to test whether participants' detection ability and/or response bias differed for in-group vs. out-group speaker-label pairs. Results showed that participants' detection sensitivity was not modulated by our social manipulations such that they were equally sensitive to in-group and out-group speaker-label pairings. On the other hand, the model exploring *C* values showed that the more in-group biased, the more conservative participants were in their decision (i.e., having a bias for "no" responses), and this was particularly applied to in-group related decisions. That is, participants' in-group bias and speakers' group membership influenced *how liberally* decisions were made, so that participants with stronger in-group bias were more careful in attributing in-group labels to any speaker. This pattern explains why participants' in-group bias negatively predicted hit rates and positively predicted correct rejection rates: the stronger the in-group bias, the more likely participants responded "no" to in-group speaker-label pairs.

How do our findings reconcile with the initial predictions and with previous literature? While previous studies showed that source memory was more accurate for information related to in-group membership, compared to information related to out-group membership (e.g., Hugenberg et al., 2010; Greenstein et al., 2016), in the current study we showed that the scenario can be more complex. Participants with a stronger bias were more accurate at correctly rejecting mismatches involving in-group labels, but they were also more likely to miss in-group matches. Looking closely at these patterns, we could deduce, and confirm with our analyses, that it was participants' response bias that was mainly affected by our social manipulation of group membership, and by participants' in-group bias. Participants with stronger in-group bias were in fact more cautious when attributing in-group labels to any speakers.

Our results resemble previous findings by Castano et al. (2002), who investigated if high vs. low in-group identifiers differed in their decision preferences when they had to categorize ambiguous faces as either in-group (i.e., Northern Italians) or out-group (i.e., Southern Italians) members. They found that participants that strongly identified with their in-group membership were less likely to classify a target face as in-group member, as compared to participants with a lower in-group identification score (see Yzerbyt et al., 1995; Blascovich et al., 1997; for similar results). The authors claimed that such a pattern was supportive of the *In-group overexclusion hypothesis* (Leyens and Yzerbyt, 1992), which states that when people are in doubt about classifying targets as either in-group or out-group, they tend to exclude them from their in-group. Such a hypothesis seems to apply to our dataset as well where participants with

stronger in-group bias were more conservative when attributing in-group labels to speakers.

We consider why it is that learners' in-group bias and speakers' group membership status might lead to differences in response preferences, but not in detection sensitivity, as we had predicted. In other words, what might it mean that an individual with strong in-group bias is selectively more conservative when making a decision that involves her in-group membership? Originally, we had predicted group membership and in-group biases to play a role during the *encoding* of novel words, leading to in-group representations with more highly detailed speaker-specific information, as compared to out-group representations. The lack of modulation on the detection sensitivity measure by these social variables suggests that in-group and out-group labels did not differ in how they were encoded. Instead, we found a significant Group Membership  $\times$  In-group bias effect on response bias, so that the stronger the in-group bias, the more conservative participants' responses were in relation to in-group labels, but not in relation to out-group labels.

We believe that these differences in decision bias might reflect asymmetries during retrieval processes for in-group related episodic events, as compared to out-group related events. Previous research has shown that response bias acts during memory retrieval processes (Windmann et al., 2002) and depends on criterion setting functions of the prefrontal cortex (Schacter et al., 1998; Swick and Knight, 1999; Miller et al., 2001). During recognition decision-making processes, this brain region is considered to be involved in initiating, monitoring and controlling item-retrieval from memory to maintain a description of the information being sought and actively inhibit memory traces that do not match this description (Buckner, 1996; Fletcher et al., 1998; Wagner et al., 1998; Henson et al., 1999; Tomita et al., 1999). Therefore, Windmann et al. (2002) suggest that differences in response bias, especially when independent of the accuracy of the memory, can be explained by the fact that decision makers differ in what they prioritize in the task (i.e., the detection of matches or mismatches).

In light of this evidence, our findings might reflect differences in recognition threshold for in-group vs. out-group memory traces. During the decision processes, the inhibitory system of those participants who were more in-group biased was activated to a larger extent to avoid creating false positives and attributing in-group information to any source. Attributing in-group labels to incorrect speakers might have been perceived as more hurtful than missing the detection of correct in-group speaker-label pairs, as the in-group overexclusion hypothesis states. If this was indeed the case, these findings would validate the claim that in-group membership information recruits the control system to a larger degree than out-group membership does, as has been previously suggested (Meissner et al., 2005; Van Bavel and Cunningham, 2012). Furthermore, such a response bias could contribute to the effect known as out-group homogeneity in face recognition and categorization tasks (Castano et al., 2002), where new out-group faces produce more false alarms than new in-group faces do, supporting the claim that out-group members are perceived as more homogeneous.

Of course, it is important to replicate the present novel findings using different groups of speakers and different tasks, as to ensure that these effects and biases do not reflect poor recognition and/or high cognitive load in general. While the analyses revealed that participants' accuracy was above chance level, it was still relatively low. Note that participants learned about the affiliations of the speakers during the word learning task, by seeing the faces of the speakers together with the logos of the supposed academic affiliations. This means that during the source memory test, they were potentially retrieving from their memory multiple pieces of information (e.g., speaker's affiliation, label's source). On that point, it is worth mentioning that even though in-group trials included a logo that might be more familiar than the out-group logo, as it is participants' own university logo, participants did not exhibit superior memory for in-group items. Future studies should test whether our finding replicates when the source memory task is simplified, for instance, by participants learning the group membership status of speakers in an earlier experimental session, and in a more natural way (e.g., by listening to speakers referring to their university lives).

Similarly, to gain a deeper understanding of how speakers' group membership and individual in-group biases influence language learning, it would be important to test whether source memory (i.e., the speaker) and item memory (i.e., the word) are equally affected by these social factors. While in this study we investigated the encoding of context-related information in the representations of novel words, and tested if its specificity was modulated by group membership and individual in-group biases, further research should test whether these factors influence the linguistic component of the representations, too. According to our general hypotheses, labels learned from in-group speakers would be easier to remember than words learned from out-group speakers.

If these patterns are substantiated, they will have far-reaching implications for theories of language learning and processing, as well as theories concerning prejudice and stereotyping. For instance, the results suggest that interlocutors' group membership status and listeners' individual biases may influence how likely newly acquired information is to be generalized to other interlocutors. In particular, for in-group speakers, listeners with a strong in-group bias appear to be more cautious when attributing in-group related information to other speakers, preventing over-generalization, whereas speakers with low in-group bias may be more liberal in their generalizations. One may

wonder whether this greater caution relates to social stereotypes as well. It is well known that people tend to homogenize out-group members whereas they are aware of the heterogeneity of their own in-group. It would be interesting to examine to what degree such findings relate to the findings from this study about individuals' greater cautiousness in attributing information to in-group compared with out-group members.

Further research should explore more how social characteristics that are ascribed to both speakers and contexts during language processing, and information processing more generally, influence encoding and storage, and how these, in turn, affect decision processes during memory retrieval. Such research would shed further light on the intersection between memory and processing, including language processing, and, importantly, how this intersection is influenced by the social properties of the input.

## DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the Social Sciences department of the Radboud University Nijmegen (project code: ECSW2014-1003-196). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SI organized the database, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the conception and design of the study, wrote sections of the manuscript, and contributed to the manuscript revision, read, 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/fnhum.2019.00308/full#supplementary-material>

## REFERENCES

- Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., and Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends Cogn. Sci.* 16, 593–605. doi: 10.1016/j.tics.2012.10.002
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv*
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv*.
- Blascovich, J., Wyer, N. A., Swart, L. A., and Kibler, J. L. (1997). Racism and racial categorization. *J. Pers. Soc. Psychol.* 72:1364. doi: 10.1037//0022-3514.72.6.1364
- Buckner, R. L. (1996). Beyond HERA: contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychon. Bull. Rev.* 3, 149–158. doi: 10.3758/BF03212413
- Castano, E., Yzerbyt, V., Bourguignon, D., and Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/out-group categorization. *J. Exp. Soc. Psychol.* 38, 315–322. doi: 10.1006/jesp.2001.1512
- R Core Team (2016). *R: A Language and Environment for Statistical Computing [Computer Software Manual]*. Vienna: R Core Team
- Drager, K., and Kirtley, M. J. (2016). "Awareness, salience, and stereotypes in exemplar-based models of speech production and perception," in *Awareness*



- and Control in Sociolinguistic Research, ed. M. Babel, (Cambridge: Cambridge University Press), 1–24. doi: 10.1017/cbo9781139680448.003
- Enock, F., Sui, J., Hewstone, M., and Humphreys, G. W. (2018). Self and team prioritisation effects in perceptual matching: evidence for a shared representation. *Acta Psychol.* 182, 107–118. doi: 10.1016/j.actpsy.2017.11.011
- Fletcher, P. C., Shallice, T., Frith, C. D., Frackowiak, R. S., and Dolan, R. J. (1998). The functional roles of prefrontal cortex in episodic memory. II. Retrieval. *Brain* 121, 1249–1256. doi: 10.1093/brain/121.7.1249
- Frable, D. E., and Bem, S. L. (1985). If you are gender schematic, all members of the opposite sex look alike. *J. Pers. Soc. Psychol.* 49:459. doi: 10.1037//0022-3514.49.2.459
- Goldinger, S. D. (2007). “A complementary-systems approach to abstract and episodic speech perception,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, (Homburg: Saarbrücken Univ), 49–54.
- Greenstein, M., Franklin, N., and Klug, J. (2016). In-group versus out-group source memory. *Exp. Psychol.* 63, 150–158. doi: 10.1027/1618-3169/a000322
- Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27. doi: 10.1037/0033-295X.102.1.4
- Hay, J., Nolan, A., and Drager, K. (2006a). From fush to feesh: exemplar priming in speech perception. *Linguist. Rev.* 23, 351–379.
- Hay, J., Warren, P., and Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *J. Phon.* 34, 458–484. doi: 10.1016/j.wocn.2005.10.001
- Henson, R. N. A., Shallice, T., and Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain* 122, 1367–1381. doi: 10.1093/brain/122.7.1367
- Hugenberg, K., Young, S. G., Bernstein, M. J., and Sacco, D. F. (2010). The categorization-individuation model: an integrative account of the other-race recognition deficit. *Psychol. Rev.* 117:1168. doi: 10.1037/a0020463
- Johnson, K., Strand, E. A., and D’Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *J. Phon.* 27, 359–384. doi: 10.1006/jpho.1999.0100
- Jones, C. R., and Fazio, R. H. (2010). Person categorization and automatic racial stereotyping effects on weapon identification. *Pers. Soc. Psychol. Bull.* 36, 1073–1085. doi: 10.1177/0146167210375817
- Kinzler, K. D., Corrivéau, K. H., and Harris, P. L. (2011). Children’s selective trust in native-accented speakers. *Dev. Sci.* 14, 106–111. doi: 10.1111/j.1467-7687.2010.00965.x
- Leyens, J. P., and Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: impact of valence and confirmation on stereotypical information search. *Eur. J. Soc. Psychol.* 22, 549–569. doi: 10.1002/ejsp.2420220604
- Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5
- Macmillan, N. A., and Creelman, C. D. (2004). *Detection Theory: A User’s Guide*. Hove: Psychology Press.
- Martin, C. D., Garcia, X., Potter, D., Melinger, A., and Costa, A. (2016). Holiday or vacation? the processing of variation in vocabulary across dialects. *Lang. Cogn. Neurosci.* 31, 375–390. doi: 10.1080/23273798.2015.1100750
- Meissner, C. A., Brigham, J. C., and Butz, D. A. (2005). Memory for own-and other-race faces: a dual-process approach. *Appl. Cogn. Psychol.* 19, 545–567. doi: 10.1002/acp.1097
- Miller, M. B., Handy, T. C., Cutler, J., Inati, S., and Wolford, G. L. (2001). Brain activations associated with shifts in response criterion on a recognition test. *Can. J. Exp. Psychol.* 55:162. doi: 10.1037/h0087363
- Molenberghs, P., and Louis, W. R. (2018). Insights from fMRI studies into ingroup bias. *Front. Psychol.* 9:1868. doi: 10.3389/fpsyg.2018.01868
- Moradi, Z., Sui, J., Hewstone, M., and Humphreys, G. W. (2015). In-group modulation of perceptual matching. *Psychon. Bull. Rev.* 22, 1255–1277. doi: 10.3758/s13423-014-0798-8
- Münster, K., and Knoeferle, P. (2018). Extending situated language comprehension (Accounts) with speaker and comprehender characteristics: toward socially situated interpretation. *Front. Psychol.* 8:2267. doi: 10.3389/fpsyg.2017.02267
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927X99018001005
- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *J. Phon.* 39, 132–142. doi: 10.1016/j.wocn.2010.12.007
- Schacter, D. L., Norman, K. A., and Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annu. Rev. Psychol.* 49, 289–318. doi: 10.1146/annurev.psych.49.1.289
- Stolte, M., Humphreys, G., Yankouskaya, A., and Sui, J. (2017). Dissociating biases towards the self and positive emotion. *Q. J. Exp. Psychol.* 70, 1011–1022. doi: 10.1080/17470218.2015.1101477
- Sumner, M., Kim, S. K., King, E., and McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Front. Psychol.* 4:1015. doi: 10.3389/fpsyg.2013.01015
- Swick, D., and Knight, R. T. (1999). Contributions of prefrontal cortex to recognition memory: electrophysiological and behavioral evidence. *Neuropsychology* 13:155. doi: 10.1037//0894-4105.13.2.155
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., and Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401:699. doi: 10.1038/44372
- Van Bavel, J. J., and Cunningham, W. A. (2012). A social identity approach to person memory: group membership, collective identification, and social role shape attention and memory. *Pers. Soc. Psychol. Bull.* 38, 1566–1578. doi: 10.1177/0146167212455829
- Van Bavel, J. J., Packer, D. J., and Cunningham, W. A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychol. Sci.* 19, 1131–1139. doi: 10.1111/j.1467-9280.2008.02214.x
- Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054
- Wagner, A. D., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. (1998). Prefrontal cortex and recognition memory. Functional-MRI evidence for context-dependent retrieval processes. *Brain* 121, 1985–2002. doi: 10.1093/brain/121.10.1985
- Walker, A., and Hay, J. (2011). Congruence between ‘word age’ and ‘voice age’ facilitates lexical access. *Lab. Phonol.* 2, 219–237. doi: 10.1515/labphon.2011.007
- Wilder, D. A. (1990). Some determinants of the persuasive power of in-groups and out-groups: organization of information and attribution of independence. *J. of Personal. Soc. Psychol.* 59:1202. doi: 10.1037//0022-3514.59.6.1202
- Windmann, S., Urbach, T. P., and Kutas, M. (2002). Cognitive and neural mechanisms of decision biases in recognition memory. *Cereb. Cortex* 12, 808–817. doi: 10.1093/cercor/12.8.808
- Xiao, Y. J., Coppin, G., and Van Bavel, J. J. (2016). Perceiving the world through group-colored glasses: a Perceptual model of intergroup relations. *Psychol. Inq.* 27, 255–274. doi: 10.1080/1047840x.2016.1199221
- Yzerbyt, V. Y., Leyens, J. P., and Bellour, F. (1995). The ingroup overexclusion effect: identity concerns in decisions about group membership. *Eur. J. Soc. Psychol.* 25, 1–16. doi: 10.1002/ejsp.2420250102

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

Copyright © 2019 Iacozza, Meyer and Lev-Ari. 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.





# Searching for Semantic Knowledge: A Vector Space Semantic Analysis of the Feature Generation Task

Rebecca A. Cutler<sup>1\*</sup>, Melissa C. Duff<sup>2</sup> and Sean M. Polyn<sup>1</sup>

<sup>1</sup> Department of Psychology, Vanderbilt University, Nashville, TN, United States, <sup>2</sup> Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, United States

## OPEN ACCESS

### Edited by:

Praveen Pilly,  
HRL Laboratories, United States

### Reviewed by:

Marc N. Coutanche,  
University of Pittsburgh, United States  
Matthew Grilli,  
University of Arizona, United States

### \*Correspondence:

Rebecca A. Cutler  
rebecca.a.cutler@vanderbilt.edu

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 01 July 2019

**Accepted:** 17 September 2019

**Published:** 04 October 2019

### Citation:

Cutler RA, Duff MC and Polyn SM  
(2019) Searching for Semantic  
Knowledge: A Vector Space Semantic  
Analysis of the Feature Generation  
Task. *Front. Hum. Neurosci.* 13:341.  
doi: 10.3389/fnhum.2019.00341

A recent neuropsychological study found that amnesic patients with hippocampal damage (HP) and severe declarative memory impairment produce markedly fewer responses than healthy comparison (CO) participants in a semantic feature generation task (Klooster and Duff, 2015), consistent with the idea that hippocampal damage is associated with semantic cognitive deficits. Participants were presented with a target word and asked to produce as many features of that word as possible (e.g., for target word “book,” “read words on a page”). Here, we use the response sequences collected by Klooster and Duff (2015) to develop a vector space model of semantic search. We use this model to characterize the dynamics of semantic feature generation and consider the role of the hippocampus in this search process. Both HP and CO groups tended to initiate the search process with features close in semantic space to the target word, with a gradual decline in similarity to the target word over the first several responses. Adjacent features in the response sequence showed stronger similarity to each other than to non-adjacent features, suggesting that the search process follows a local trajectory in semantic space. Overall, HP patients generated features that were closer in semantic space to the representation of the target word, as compared to the features generated by the CO group, which ranged more widely in semantic space. These results are consistent with a model in which a compound retrieval cue (containing a representation of the target word and a representation of the previous response) is used to probe semantic memory. The model suggests that the HP group’s search process is restricted from ranging as far in semantic space from the target word, relative to the CO group. These results place strong constraints on the structure of models of semantic memory search, and on the role of hippocampus in probing semantic memory.

**Keywords:** hippocampus, semantic search, amnesia, relational memory, vector space model

## 1. INTRODUCTION

The most dramatic effects of hippocampal and medial temporal lobe damage are in the domain of episodic and autobiographical memory. Patients with bilateral damage to the hippocampus typically have dense anterograde amnesia, resulting in an inability to form new memories of their ongoing experience (Milner et al., 1998). This amnesic condition is consistent with the dominant view of hippocampal function: That hippocampus constructs a summary representation of the widespread cortical activity representing the details of an experienced event, and rapid synaptic

plasticity binds this hippocampal representation to these widespread cortical patterns (Mishkin et al., 1983; McClelland et al., 1995; Eichenbaum, 2000). As such, hippocampus is proposed to be critically involved in binding the representations of event details to the spatiotemporal context in which they occurred, which is a defining characteristic of episodic memory (Tulving, 1972; Eichenbaum et al., 2007, 2012).

The nature of hippocampal involvement in semantic memory processes is less well settled. By one view, the hippocampus is involved in the acquisition (and possibly curation) of semantic memories through a consolidation process. Hippocampally dependent memory traces corresponding to episodic experiences are periodically reactivated, allowing cortical structures to slowly learn statistically reliable semantic characteristics of the world and the things in it (McClelland et al., 1995; Norman and O'Reilly, 2003; Eichenbaum, 2004). This view is consistent with work showing that after adult-onset hippocampal injury, the acquisition of new semantic knowledge is impaired (Gabrieli et al., 1988; Bayley and Squire, 2002; Manns et al., 2003; O'Kane et al., 2004; Sharon et al., 2011; Warren and Duff, 2014). However, it may be the case that cortical structures can form semantic memories without a functioning hippocampus. Despite dense episodic amnesia, patients with developmental hippocampal damage can still acquire new semantic knowledge (Vargha-Khadem et al., 1997). However, semantic learning in these patients seems to be slower and less flexible than in healthy individuals (Elward and Vargha-Khadem, 2018). It is possible that consolidation is better thought of as a gradual process, without a clear point at which hippocampus stops being involved (Winocur et al., 2010).

Putting aside the question of acquisition, a wide range of neuropsychological studies have shown that patients with hippocampal damage have minimal impairment in their ability to use their existing semantic knowledge. These patients perform at normal or near-normal levels on tests of their vocabulary breadth, their ability to define words and name objects, and even their ability to retrieve long-known associative pairings, such as the names of famous faces (Reed and Squire, 1998; Verfaellie et al., 2000; Schmolck et al., 2002; Westmacott and Moscovitch, 2002). In contrast, patients with damage to lateral temporal cortex, and especially anterior temporal cortex, are impaired at these semantic tasks, suggesting an anatomical dissociation of function (Irish et al., 2016). As such, the dominant view is that utilization of existing semantic knowledge does not involve hippocampus, but rather involves other cortical regions such as anterior temporal lobe (Ralph et al., 2017).

A number of recent studies have challenged this view, by demonstrating that patients with hippocampal or medial temporal lobe damage are impaired on certain tasks involving the utilization of existing semantic memory. When recounting well-known fairy tales and bible stories, these patients produce fewer details (Verfaellie et al., 2014). When producing event narratives, they use words rated lower on imageability scales (Hilverman et al., 2017), and generate fewer words in free association when cues were highly imageable and low frequency (Sheldon et al., 2013). In general, patients with medial temporal lobe damage show a retrograde impairment in retrieving information from

personal semantic memory, including memories ranging back to early childhood (Grilli and Verfaellie, 2014). These findings are bolstered by periodic reports from the neuroimaging and neuropsychological literature of hippocampal involvement in semantic tasks (Henke et al., 1999; Sheldon and Moscovitch, 2012; Race et al., 2013). Furthermore, the response properties of hippocampal cells suggest that semantic information is embedded in hippocampal neural representations. For example, a substantial proportion of cells in human hippocampus show category-specific responses (Kreiman et al., 2000), and individual cells can show invariant responses to particular concepts, e.g., by responding to a particular celebrity across different images as well as to the celebrity's name presented in text (Quiroga et al., 2005; Quiroga, 2012).

A recent study by Klooster and Duff (2015) provides further evidence for hippocampal involvement in semantic memory processes. They used tasks that were developed for psycholinguistic and language-learning research, and are designed to characterize vocabulary depth and semantic richness. The Word Associates Test is an evaluative task in which a participant has to identify synonyms and collocates of a target word (collocates are words that tend to occur together in text or speech, such as *innate* and *ability*, or *maiden* and *voyage*) (Read, 1993, 1998). They also used two generative tasks. One of these was a feature generation task in which a target item was presented, and the person was asked to name as many features or characteristics of the item as they could in a 2 min interval. For example, if the target item was *book*, a participant might respond with the feature "you read words on a page." The second was a senses task in which participants were presented with a target word and given 1 min to list senses of the word (e.g., the word *bank* can mean a financial institution, or the bank of a river). Patients were impaired on all three of these tasks relative to a set of healthy comparison participants. The most marked deficit was in the feature generation task: whereas the healthy comparison group produced upwards of 20 features on average for a given target word, the amnesic patients produced roughly half as many.

These results raise the possibility that hippocampal damage gives rise to a semantic memory deficit that is masked by patients' normal-range performance on tasks that probe semantic knowledge at a surface level. In other words, hippocampus may play an important role in semantic processing that goes beyond supporting the initial acquisition of semantic memory through replay of episodic experiences. We propose that the growing body of work establishing the role of hippocampus in relational processing may provide insight into its contribution to semantic processing. Relational processing is engaged whenever multiple arbitrary components of an experience need to be associated to one another, creating a relational representation (Rubin et al., 2014). A number of studies suggest that hippocampally dependent relational processing is engaged in a variety of cognitive domains that extend well beyond episodic memory (Cohen et al., 1997, 1999; Davachi, 2006; Olsen et al., 2012; Olson and Newcombe, 2013).

Episodic memories are inherently relational, in that an event consists of a constellation of item and contextual details that must be bound together to form a new memory trace. For similar

reasons, spatial navigation involves relational processing, as the construction of representations of place and location involve processing the relations between many environmental features (Burgess et al., 2002). As such, the relational memory view of hippocampal function provides a natural explanation for why hippocampal damage is associated with both episodic and spatial memory deficits (Konkel et al., 2008; Konkel and Cohen, 2009). This account also explains behavioral deficits accompanying hippocampal damage in perceptual tasks and short-term memory tasks where the stimuli are comprised of multiple configural features (Hannula et al., 2006; Olson et al., 2006; Warren et al., 2011).

A developing branch of the relational memory literature has examined spatial reconstruction tasks, in which participants try to reconstruct a multi-item display after a short delay to evaluate spatial-relational memory. This task seems to be particularly hippocampally-dependent as participants with lesions in this area have difficulty correctly recalling the spatial relations of studied items (Smith and Milner, 1981; Jeneson et al., 2010; Watson et al., 2013; Horecka et al., 2018). A subset of multi-item spatial encoding tasks have found evidence for a role of the hippocampus in actively guiding visual search, with hippocampal activation corresponding to enhanced subsequent memory (Voss et al., 2011a,b; Lucas et al., 2018). We consider the idea that internal semantic feature search may in some ways parallel navigation and visuospatial exploration, given that the hippocampus seems to facilitate information-gathering and sampling in both processes. We propose that semantic deficits due to hippocampal damage are related to previously observed relational processing deficits.

### 1.1. A Computational Analysis of the Feature Generation Task

In the current study, we examine the data originally collected by Klooster and Duff (2015) to test whether participants' impaired performance on the feature generation task can be understood in terms of a relational semantic deficit. To do this, we use a computational model of semantic representational structure to characterize the memory search processes engaged by the task. This allows us to examine the semantic relations between generated features and the target word, and the relations of the set of generated features to one another. We find substantial differences in the nature of semantic search between the two groups, which we interpret in terms of current theories of semantic and episodic memory search. While Klooster and Duff (2015) also characterized semantic deficits in two other tasks, feature generation performance was the most amenable to semantic analysis: its generative nature allowed us to examine the dynamics of search, and participants overall produced about five times as many responses in this task relative to the senses task.

A number of algorithms have been developed to construct semantic representational codes from either large text corpora (Lund and Burgess, 1996; Landauer and Dumais, 1997; Jones and Mewhort, 2007) or from behavioral responses in a free association task (Steyvers et al., 2004). These are often referred to as vector space models of semantics, as each representational code in the system is a vector of numbers. While the numerical

features that comprise the representations in these vector space models are rarely directly interpretable, they do provide a reference point for each word, such that words with similar features are situated near to one another in the vector space. This computational approach allows us to consider a sequence of responses in the feature generation task as a trajectory through an abstract semantic representational space. This trajectory can be characterized in terms of the semantic distance between the target word and the individual features generated by participants, and the distance of the generated features to one another.

Hills et al. (2012) used a similar approach to characterize performance on a semantic fluency task, in which participants are asked to provide examples from the semantic category "animal." In their model, the vector representation of the most recent response was used as a retrieval cue to determine the next response. The likelihood of recalling a particular word in a search of the category semantic space was proportional to its representational similarity to the most recent response. This framework naturally explains the semantic clustering seen in semantic fluency tasks: The initial response tends to be a highly frequent exemplar of the category (Henley, 1969; Newcombe, 1969), and the continual updating of the retrieval cue causes contiguous responses to be semantically similar to one another (Bousfield and Sedgewick, 1944; Federmeier et al., 2002; Voorspoels et al., 2014).

Our proposed model is similar to the Hills et al. model in that feature responses are based on a blended representation of target word and previous recall information. Critically, all of the HP patients in the (Klooster and Duff, 2015) study were impaired in feature generation. However, as a group they did not show a deficit in a measure of semantic fluency (the Controlled Oral Word Association test), although there seems to be more variability in their performance at the individual level. It is therefore important to understand how semantic feature generation differs from semantic fluency, and how the task demands might reveal the nature of semantic deficits in hippocampal amnesia. In both cases, the participant's knowledge is probed in a constrained way. With semantic fluency, responses are constrained to come from a particular taxonomic category (Gruenewald and Lockhead, 1980). In semantic feature generation, responses are constrained to be in reference to a target word, and are meant to describe properties or characteristics of the referent item. However, these tasks seem to require access to different kinds of conceptual representations. Firstly, the feature generation task cues semantic search with a more specific target than semantic fluency (e.g., "dolphin" vs. "animal," respectively). Secondly, the task demands of feature generation requires retrieval of richer multi-word conceptual representations, whereas semantic fluency requires participants to name exemplars. Lastly, in semantic fluency each response is related to the others by the shared features that comprise category membership. In comparison, in semantic feature generation the adjacently retrieved features are potentially semantically unrelated outside the context of the target word (e.g., "gray in color" and "intelligent animal" for "dolphin"). This type of feature generation seems to require relational memory to access semantically disparate concepts that are related only

given the context of the target word. Our semantic analyses suggest that this task distinction is important and can potentially unearth semantic memory deficits that are otherwise masked in surface-level tasks. In the discussion we consider the critical role of the hippocampus in relational memory, and the key differences between the tasks mentioned above and semantic tasks which are not associated with an impairment in patients with hippocampal amnesia. We will return to the question of how these results inform our understanding of hippocampal engagement in semantic memory search.

## 2. METHODS

### 2.1. Participants

Participants were five patients with bilateral hippocampal damage (HP) exhibiting declarative memory impairments. Fifteen healthy participants (CO) were matched to the patient group on sex, age, and education (three matched participants to each patient). Each patient with hippocampal damage had stable, non-progressive lesions. The etiology of three patients was anoxia/hypoxia—resulting in bilateral hippocampal damage. Two patients had herpes simplex encephalitis, resulting in broader bilateral medial temporal lobe damage, including hippocampus, amygdala, and surrounding cortices. For more details see Klooster and Duff (2015).

### 2.2. Experimental Procedure

In the feature generation task, participants were presented with a target word (e.g., “bed”) and given 2 min to verbally list as many features of that word as possible. Thirty-five target words were sampled from established feature production norms (McRae et al., 2005). Instructions and examples were given to the participant at the start of the task, and were left in front of the participants and repeated by the experimenter regularly. On each trial, the experimenter read the target word aloud, and prompted the participant to begin to report features. If, during this recall period, the participant stopped responding, the target word was repeated by the experimenter, and the participant was encouraged to keep trying to generate features. Responses were video recorded for later transcription and analysis.

### 2.3. Preprocessing Response Sequences

Each participant’s verbal responses were transcribed and two judges coded these responses into a sequence of features. See Klooster and Duff (2015) for details regarding this coding procedure. **Table 1** provides representative examples of features. For the current study, we developed a coding scheme to include all content words and exclude function words from the semantic analysis. Excluded grammatical groups were: personal pronouns, possessive pronouns, auxiliary verbs (be, do, and have), coordinating conjunctions and articles.

### 2.4. Semantic Vector Space Models

In this study we used semantic representations constructed with the Global Vectors (GloVe) algorithm (Pennington et al., 2014), which has excellent coverage of the English language due to the large text corpus used to construct the vector representations.

**TABLE 1 |** Examples of features generated by healthy comparison participants (CO) and patients with hippocampal damage (HP) for the target words “book” and “grapefruit.”

Group	Target word	Feature (word similarity below)	Cosine
CO	Book	Can be <b>bound</b> in either <b>leather</b> or <b>paper</b> 0.2328      0.1809      0.4542	0.2911
HP	Book	<b>Something</b> you <b>read</b> 0.3614      0.5758	0.4686
CO	Grapefruit	<b>Often</b> <b>produced</b> in <b>Florida</b> −0.0674      0.0560      0.1708	0.0531
HP	Grapefruit	<b>Skin</b> <b>like</b> <b>oranges</b> 0.1805      0.0666      0.6017	0.2829

*Bolded words in each feature indicate which words were included in the analysis after the exclusion of non-content words. Underneath each bolded word is the cosine similarity score between that word and the target word. These values were averaged to create the overall cosine similarity score for the feature, which is given in the rightmost column.*

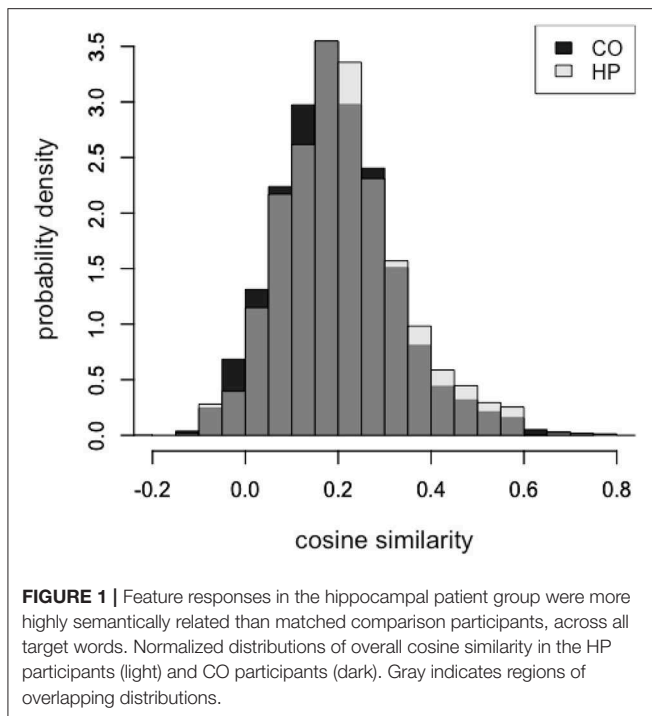
GloVe follows in a long tradition of computational models attempting to quantify the meaning of words by assigning each word a point in a high-dimensional vector space, often containing up to 300 dimensions (Deerwester et al., 1990; Lund and Burgess, 1996; Landauer and Dumais, 1997; Steyvers et al., 2004). These techniques tend to use linear algebraic algorithms (such as singular value decomposition) to construct vector representations given statistics characterizing the co-occurrence of words in a large text corpus.

These semantic vector space models formalize longstanding ideas from linguistics and philosophy regarding how best to characterize the meaning of words. In the linguistics literature, the *Distributional Hypothesis* refers to the notion that words that co-occur across similar contexts tend to have similar or related meanings (Harris, 1954). Linguist J. R. Firth famously summarized the context-dependent nature of meaning with the phrase “You shall know a word by the company it keeps” (Firth, 1957, p. 11). The assignment of vector representations to words and phrases resonates with ideas developed by Wittgenstein (1953), whereby words can be loosely grouped by a combination of shared features. Given these vector representations, the semantic similarity of two words can be quantified using standard distance measures such as Euclidean distance or the cosine angle between two vectors (Kwantes, 2005). In the current work we used the GloVe model to construct a single representation for multi-word features by taking the average of the semantic similarity of each of the feature words to the target word. For feature-to-feature analysis we calculated the pairwise similarity between each word in the two features; the average of these similarity scores was used to represent the similarity of the features to one another.

### 2.5. Bayesian Analysis of Feature Responses

We implemented all Bayesian analyses in R. Initial analysis using a frequentist framework indicated that residuals were





not normally distributed, motivating the use of a Bayesian framework. Further, the nature of cosine values is such that they are log-normally distributed, and a Bayesian framework gives us more flexibility to estimate this. To examine group differences in cosine similarity scores we fit a Bayesian linear mixed effects model using the Stan and brms packages in R (Bürkner, 2017; Stan Development Team, 2017). The binary group predictor (CO vs. HP) was modeled as a fixed intercept and slope. Subject ( $s$ ) and target word ( $w$ ) predictors were modeled as random effects with varying intercepts and normally distributed priors:  $s \sim \text{Normal}(0, \sigma_s)$ ,  $w \sim \text{Normal}(0, \sigma_w)$ . Prior distributions on the variance parameters were uniform:  $\sigma_e, \sigma_s, \sigma_w \sim \text{Uniform}(0, \infty)$ . We estimated the response distribution of cosine similarity scores as log-normal, as preliminary examination of the data showed that this distribution was better described with a log-normal distribution as compared to a normal distribution. We estimated model parameters using Markov-Chain Monte-Carlo (MCMC) methods, using the No U-turn Sampler (NUTS) provided with Stan. For all Stan-based model fits, we ran 4 chains each with 4,000 iterations to ensure chains effectively converged. Chain convergence was confirmed by the  $\hat{r}$  statistic which in all cases approached 1 (indicating maximal convergence).

A second Bayesian linear mixed effects model was designed to characterize the cosine similarity of feature responses to one another (within a given response sequence). Similarity was calculated between features with lag one to four. Lag is defined as the positional difference in the response sequence, with adjacent features assigned a lag of one, features separated by one intervening feature assigned a lag of two, and so forth. Group, subject, and word predictors were modeled as described above, except that the response distribution of cosine similarity scores

was modeled as normally distributed. Prior distributions on the variance parameters, MCMC details, and model comparison details were the same as above.

In order to examine changes in cosine similarity across the response sequence, and changes as a function of lag within a given response sequence, we created a set of Bayesian multilevel models. The data was best modeled by power functions, which take the form  $f(x) = a(x^b)$ , where  $a$  is a scaling factor, and  $x$  is a variable base raised to a constant power,  $b$ . The  $b$  coefficient represents the growth or decay in cosine similarity scores as a function of  $x$ , which represents either response position or positional distance between generated features in different analyses. In our two-level hierarchical models, we estimate the group and subject-level effects of feature responses. At the group level, we estimated the  $a$  parameter for both groups with the prior  $\mu \sim \text{Normal}(0.2, 0.5)$ ,  $\sigma \sim \text{Cauchy}(0, 5)$ , and the  $b$  parameter with the prior  $\mu \sim \text{Normal}(0, 0.5)$ ,  $\sigma \sim \text{Cauchy}(0, 5)$ . We estimated model parameters using MCMC in Stan as described above. All chains converged effectively. As above, these models were compared to a null model without a group-level predictor.

### 3. RESULTS

#### 3.1. Group-Level Shift in Target-Feature Relatedness

Overall, the feature responses made by patients with hippocampal damage tended to be closer in semantic space to the target word (HP:  $\mu = 0.21$ ,  $SD = 0.13$ ) when compared to healthy comparison participants (CO:  $\mu = 0.19$ ,  $SD = 0.13$ ). This positive shift in the cosine distribution of HP responses can be seen in **Figure 1**. We used a Bayesian mixed effects regression framework to investigate the effect of group (HP vs. CO) on cosine similarity of feature response to target word. The model had a fixed effect of group and accounted for variance associated with individual subjects and target word stimuli. The posterior distribution of the MCMC chains for the group coefficient did not include zero ( $\mu = 0.02$ ,  $SD = 0.003$ , 95% CI = [0.002, 0.033]), consistent with a substantial and reliable group difference in cosine similarity. We considered the possibility that the increased cosine similarity of features to target word was driven by an individual patient, therefore we carried out a “leave-one-out” by patient analysis. We iteratively ran the model described above five times, each time excluding one patient’s data. For each iteration, the resultant posterior distribution of the group parameter did not include zero, suggesting that no single patient was skewing the group result.

#### 3.2. Feature-to-Target Relatedness Across Response Positions

Firstly, we were interested in how participants initiate the search for features of a given concept in semantic space. We examined the cosine similarity between the target word and the first five feature responses. Participants in the HP group generally make fewer responses than the CO participants, but routinely make more than 5 responses. As such, restriction to the first 5 responses puts the two groups on relatively even footing in terms of the

number of responses in each response position bin. In both HP and CO groups we found that the cosine similarity to the target word decreased across the first five responses (see **Figure 2**). While the first feature response of both groups was a similar distance from the target word, a group difference emerges over the course of the first five responses, with the CO responses ranging farther in semantic space on average relative to the HP responses.

To characterize these observed trends, we constructed a Bayesian hierarchical model, fitting power function curves to this sequence of response positions. Best-fit curves are presented in **Figure 2B**. The best-fit curves for the two groups have similar starting points (as reflected by the best-fitting  $a$  parameters, HP: 0.2536, CO: 0.2510), but the patient group shows a slower rate of decay in semantic similarity to the target word as the search progresses (as reflected by the best-fitting  $b$  parameters, HP: -0.1243, CO: -0.2018).

In order to characterize the statistical reliability of this difference in decay rates, we examined the MCMC-derived posterior distributions for each of the power function curve parameters. Intuitively, these posterior distributions contain the set of plausible parameter values for each of the groups. As we are interested in determining whether the shift in parameter values is reliable, we constructed a *difference distribution* for each of the parameters: Each sample in the posterior distribution specifies four numbers, the mean  $a$  parameter for the HP and CO groups ( $\bar{a}_{HP}, \bar{a}_{CO}$ ), and the mean  $b$  parameter for the HP and CO groups ( $\bar{b}_{HP}, \bar{b}_{CO}$ ). The difference distributions were constructed by calculating  $\bar{a}_{HP} - \bar{a}_{CO}$  and  $\bar{b}_{HP} - \bar{b}_{CO}$  for each sample in the posterior distribution.

For the  $a$  parameter, the mean of the difference distribution was near-zero (0.0027), with points tending to be evenly distributed around zero (in 57% of posterior samples  $\bar{a}_{HP} > \bar{a}_{CO}$ ). This suggests that both groups initiate semantic search in a similar way. For the  $b$  parameter, the mean of the difference distribution was more substantially positive (0.0762, consistent with a shallower decay for HP), with 83% of the difference distribution falling above zero. In other words, the semantic relatedness of the generated features to the target word decayed more slowly for the HP group, consistent with the idea that the CO group is able to range further from the target word in semantic space. The group difference in the  $b$  parameter is consistent with the group difference established in the first analysis, but is less reliable statistically. This is likely due in part to the restriction of this analysis to the first five response positions, and also to the presence of fairly substantial individual differences, as can be seen in **Figure 2B**.

Considering the first two analyses together, it seems reasonable to infer that the group-level difference characterized in the first analysis is not present in the initial responses. This is consistent with a model in which the difference in semantic relatedness emerges over the course of the response sequence. A follow-up analysis showed that the mean target-to-feature semantic relatedness for the later response positions excluded from this analysis (through to the termination of the response sequences) is similar to the asymptotic values approached by the two power curves estimated here. These results are generally

consistent with a model in which the hippocampus facilitates the retrieval of semantically distant features of the target word. We return to this point in the discussion.

### 3.3. Feature-to-Feature Semantic Relatedness

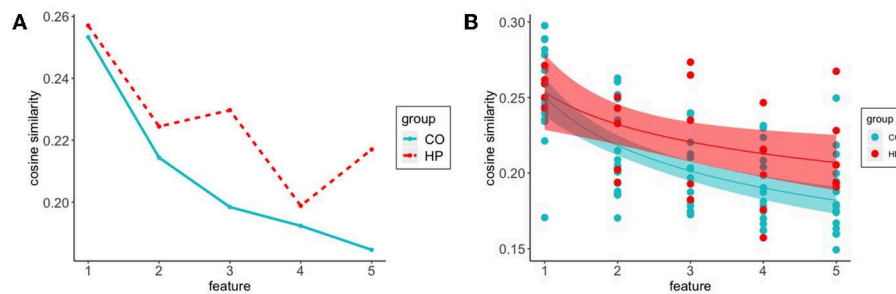
The previous analyses examined the semantic similarity of the words comprising each feature to the target word specific to that trial. In order to better characterize the dynamic nature of semantic memory search, we calculated the semantic relatedness of the reported features to one another, without regard to the semantic identity of the target word. This allowed us to examine how feature-to-feature similarity changed as a function of the relative position of the two features in the response sequence. **Figure 3** shows that as the positional lag between two responses on a given trial increases, there is a substantial decline in cosine similarity. In other words, as two responses become farther apart in the response sequence, they become less semantically related to one another.

We first used a Bayesian linear mixed effects model to estimate average feature-to-feature semantic relatedness, without considering transition lag. As before, the model had a fixed effect of group and accounted for variance associated with individual subjects and target words. The posterior distribution of the group coefficient was centered around zero ( $\mu = 0.0068$ ,  $SD = 0.0119$ ), consistent with the idea that there is no HP vs. CO group difference in feature-to-feature similarity.

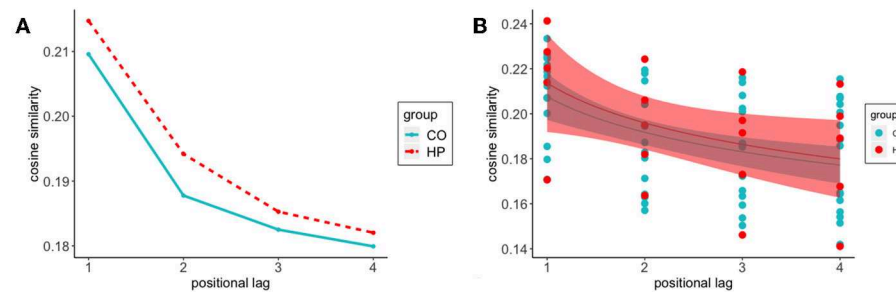
Once more, we constructed a Bayesian hierarchical model, fitting power functions to these curves to determine whether a group difference exists in the non-linear shift of semantic relatedness as feature responses become separated in the response sequence. Best-fit curves are presented in **Figure 3B**, with parameter estimates for each group in the caption. The best-fit curves for the two groups have similar starting points (the  $a$  parameter, HP: 0.2131, CO: 0.2077), and similar decay in semantic similarity as the positional distance between responses increases (the  $b$  parameter, HP: -0.1256, CO: -0.1184). As above, we calculated difference distributions,  $\bar{a}_{HP} - \bar{a}_{CO}$  and  $\bar{b}_{HP} - \bar{b}_{CO}$ , for each sample in the posterior distribution. For the  $a$  parameter, the mean of the difference distribution was near-zero (0.0059), with 65% of the samples in the CO group less than the HP group. For the  $b$  parameter, the mean of the  $\bar{b}_{HP} - \bar{b}_{CO}$  distribution was also near zero (0.0079) with 58% of the HP samples falling below the CO samples. This analysis suggests that the process governing transitions between generated features behaves similarly for the two groups.

## 4. DISCUSSION

We used a vector space model of semantic meaning to investigate differences in how patients with hippocampal damage (HP) and healthy demographically matched comparison participants (CO) performed on a feature generation task. Our results are consistent with the idea that the hippocampus is important for relational semantic memory. We constructed semantic representations of the multi-word features produced by both



**FIGURE 2 |** Cosine similarity of target word to feature responses decreased across the first five responses in both groups. CO participants show a steeper decline compared to a more gradual outward trajectory from target word space in HP participants (A) Average cosine similarity of features to target word for responses 1:5 for CO (-) and HP (-) groups (B) Power function fit to HP ( $a = 0.2536$ ,  $b = -0.1243$ ) and CO ( $a = 0.2510$ ,  $b = -0.2018$ ) initial five responses. Shadows represent 95% confidence intervals.



**FIGURE 3 |** Feature-to-feature cosine similarity decreases as the positional distance between them increases in the response sequence to a given target word. In general, HP patient's responses are more semantically related to one another as compared to CO participants (A) Average cosine similarity between feature positions at lag 1:4 for HP (-) and CO (-) groups (B) Power function fit to HP ( $a = 0.2131$ ,  $b = -0.1256$ ) and CO ( $a = 0.2077$ ,  $b = -0.1184$ ) average cosine similarity of response positions from lag 1:4. Shadows represent 95% confidence intervals.

groups, and examined the representational similarity of these features to the target word representation, and to other features reported in the same trial. We found that there was a group difference in the overall similarity of features to the target word, such that HP patients tended to generate features that were more semantically related to the target word, relative to the CO group. Furthermore, while both groups initiated search at a similar semantic distance from the target word, a difference emerged across the subsequent responses. Feature-to-target semantic similarity generally declined across responses within a trial for both groups, but the decline was reliably steeper for the CO group, consistent with their tendency to produce responses that were on average more distant from the target word in semantic space. We also found evidence for local transitions in a structured semantic space. On a given trial, adjacent features in the response sequence were most similar to one another, with similarity declining steadily as the lag between responses increased. In the following sections we discuss the motivation for using semantic vector representations to model this task, and how they inform our understanding of a relational semantic memory deficit with hippocampal damage.

#### 4.1. Search Process in Semantic and Episodic Memory

We consider how semantic vector space representations could work with mechanisms commonly used to model search processes in episodic and semantic memory. In order to develop a model of the semantic feature generation task, we begin by comparing it to other memory search tasks that have been modeled computationally. In many theories, memory search is guided by the construction and utilization of a retrieval cue: a mental representation that targets and reactivates task-relevant memories. For example, in the semantic fluency model developed by Hills et al. (2012), a retrieval cue containing the most recently reported response was used to target local conceptual representations from the category “animal.” The representational similarity between the retrieval cue and each of the not-yet-recalled animals was used to simulate a decision competition in which the likelihood of a given animal winning the competition was proportional to its semantic similarity to the retrieval cue. The continual updating of the retrieval cue causes contiguous responses to be semantically similar to one another. Smith et al. (2013) used a semantic vector space model to examine the search process in a Remote Associates Test in which the participant

must produce a target word that is semantically related to three presented cue words. Participants were encouraged to vocalize guesses as they attempted to determine the target word, and their model suggested a similar semantic dependence of a given response on the previous responses in the sequence. It is worth noting that hippocampal patients in the Klooster and Duff (2015) study were impaired on a similar Word Associates Test, in which remote semantic associates to a target word had to be identified.

In retrieved-context models of free recall, a retrieval cue comprised of context information is used to target episodic representations of words from a recently studied list (Howard and Kahana, 2002a; Sederberg et al., 2008). In many experiments, the temporal structure of the studied items dominates clustering during the recall period: items studied in nearby list positions tend to be recalled in adjacent output positions. In these models, recalling an item reinstates the context associated with that item at encoding which increases support for its neighbors at retrieval (Kahana, 1996; Kahana et al., 2008; Healey et al., 2019). There is a simultaneous influence of semantic relatedness on the order of recall responses (Howard and Kahana, 2002b; Polyn et al., 2009). As in the semantic fluency task, semantically related study items tend to be produced as contiguous responses in the recall sequence (Romney et al., 1993; Polyn et al., 2011). In a number of free recall models, these semantic organization effects arise from the dynamics of an ever-changing retrieval cue which integrates the representation of the just recalled item (Sirotnin et al., 2005; Polyn et al., 2009; Socher et al., 2009; Morton and Polyn, 2016). Here we consider how similar mechanisms could be used to develop a model of the semantic feature generation task.

## 4.2. Toward a Mechanistic Model of the Feature Generation Task

The current results provide constraints that can be taken into account in future modeling work. With regard to the functioning of the healthy cognitive system, we envision an executive system guiding task performance through the construction of a retrieval cue that probes a semantic memory space. This semantic memory space is populated with representations of known objects as well as representations of their features and characteristics. The retrieval cue activates a particular location in this semantic space, which activates nearby conceptual representations in proportion to their proximity to the activated location. This proximity-based activation is similar to the dynamics of a spreading activation model (Collins and Loftus, 1975; Anderson, 1983). These representations then compete to be retrieved, with their relative activity determining the support for each representation. The cosine similarity scores used in our analyses reflect the proximity of these representations to one another, and as such, can be thought of as approximating the level of support for each representation in this retrieval competition. The winning representation is fully activated, allowing that feature to be verbally reported. The retrieved feature representation can then be used to modify the retrieval cue, and semantic search continues.

The observed behavioral phenomena are consistent with this model. We propose that semantic space is probed and guided

by a compound retrieval cue, containing a representation of the target word as well as a representation of the most recent feature response. The first features retrieved tend to be close in semantic space to the target word, suggesting that the initial search is guided by a retrieval cue that simply contains a representation of the target word. Subsequent responses range further from the target word in the semantic space, and neighboring responses tend to be more similar to one another than to other responses. One way for the system to support retrieval of more distant features in the semantic space is to integrate information related to already retrieved features into the retrieval cue itself, creating a compound cue of target and recent feature information. This compound cue would allow the system to target more distant parts of the semantic space, as features proximal to the already retrieved features would now receive additional support in the retrieval competition. By retaining target word information in the retrieval cue, the system can ensure that retrieved features remain relevant to the current target word. However, as the number of feature responses increases, the target word representation may become progressively less influential in the retrieval cue, allowing the system to range further from its point of initiation (as shown in Figure 4).

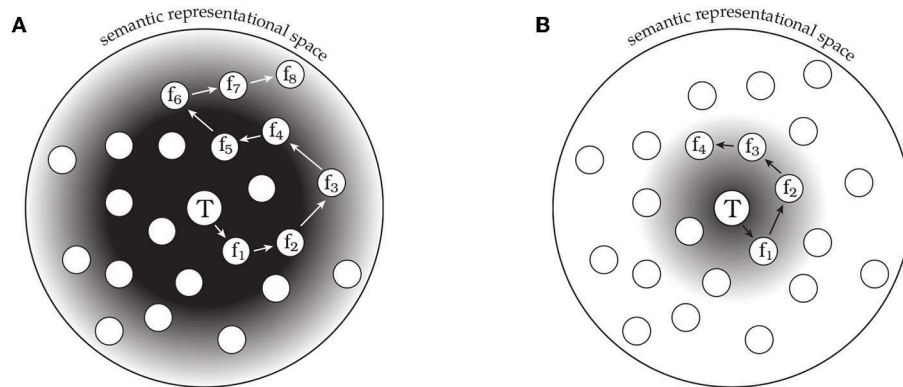
These results also provide constraints regarding the specific contribution of the hippocampus to semantic memory search, although there remain a number of open questions that we discuss in the following sections. Specifically, with hippocampal damage, feature responses have a restricted range in semantic space. However, at the same time the semantic relatedness of successively reported features to one another is unaffected. This raises the possibility that the executive machinery guiding search is unaffected, as it is still able to incorporate information about the previous response into the retrieval cue guiding search. Furthermore, patients are able to reliably stay “on task,” in that they consistently generate valid features of the target word. Indeed, Klooster and Duff (2015) found no significant group difference in the number of unrelated responses ( $p > 0.27$ ) or the number of factually incorrect responses ( $p > 0.62$ ) produced between the CO and HP groups. The deficit seems more specific to the patients’ ability to access distant semantic features of the target word.

## 4.3. Hippocampal Damage and Semantic Memory

As reviewed in the introduction, the hippocampus has been clearly implicated in both relational processing and episodic memory. However, its role in semantic memory is less well characterized. We propose that people with hippocampal damage have difficulty using semantic knowledge in a flexible, relational manner. As mentioned above, neuropsychological studies have found that patients show minimal impairment in basic tasks probing semantic knowledge, but it is possible that these tasks mask a more subtle deficit in relational processing.

A number of studies indicate that the hippocampus contributes to successful relational memory – that is, the formation of long-term memories comprised of multiple elements bound together (Cohen et al., 1999; Konkel and





**FIGURE 4 |** A schematic model of semantic memory search differences in healthy comparison participants (CO, **A**) and patients with hippocampal damage (HP, **B**). The central circle labeled T corresponds to the conceptual representation of the target word, and the surrounding circles correspond to potentially reportable features. The gradient represents the strength of support for features based on the target concept. Arrows indicate the sequence of feature responses ( $f_n$ ) made by the hypothetical participant. The tendency to make transitions to nearby features suggests a search process that is influenced by the most recent response. **(A)** CO participants make a longer sequence of responses that ranges farther in semantic space. **(B)** HP participants produce a shorter response sequence that does not range as far from the target concept.

Cohen, 2009). As we discuss below, relational processes can be independent of long-term memory and can refer to any cognitive mechanism involving relational representations. In the feature generation task the HP group shows an impairment in retrieving rich semantic representations: fewer features are produced, and the produced features do not range as far in semantic space as those produced by the CO group. We discuss possible explanations for this observed deficit. First, we consider that the relational search process, facilitated by a compound retrieval cue, is impaired. Second, we consider whether the deficit could arise directly from an impaired ability to retrieve episodic memories. Third, we explore the possibility that the HP group impairment is due to a general degradation of the semantic space used to represent features.

#### 4.3.1. Relational Binding and the Hippocampus

The relational-binding theory of memory posits that the hippocampus plays a critical role in assembling and relating the disparate details of an experience to form a coherent, holistic representation (Cohen and Eichenbaum, 1993; Ryan et al., 2000; Davachi and Wagner, 2002; Barense et al., 2007; Staresina and Davachi, 2009; Olson and Newcombe, 2013). As such, hippocampal damage affects performance on a variety of tasks outside of the domain of episodic memory. Here, we consider the relevance of this theory to the semantic deficit observed in the feature generation task. By this theory, semantic memories may be generally intact in HP patients. The impairment would arise from an inability to hold multiple or diverse semantic features in mind simultaneously to probe semantic memory effectively.

A number of studies have shown that patients with hippocampal damage have impaired memory for configural information at very short delays (Hannula et al., 2006; Warren et al., 2015), and even when all relevant information remains onscreen (Warren et al., 2011, 2012). Warren et al. (2011) reported an impairment in amnesic patients performing visual

search for a target among complex stimuli which resemble the target to varying degrees. In order to perform this visual search task, one likely has to construct and maintain a complex internal representation of the target stimulus. This internal representation could then be used to determine whether a given lure stimulus matches the target. They found that while comparison participants fixated on the target less often as the trials went on, patients fixated on it at a constant rate across trials, suggestive of hippocampal involvement in maintaining the complex representation of the target item. More recently, Lucas et al. (2018) found that patients with hippocampal amnesia were more likely to engage in random, less structured saccade patterns when studying a spatial array. This randomness was predictive of less accurate spatial reconstruction, consistent with the idea that the hippocampus helps construct and maintain a configural representation of the spatial environment.

Further evidence for a hippocampal role in exploratory viewing comes from neuroimaging studies looking at the fMRI BOLD response in the hippocampus as participants controlled which item they studied in a spatial array (Voss et al., 2011a,b). A key finding was that “spontaneously revisiting” an item (i.e., looking backward at a recently viewed item) produced a subsequent memory benefit for that item and was associated with increased hippocampal connectivity. Interestingly, patients with hippocampal amnesia rarely engaged in this revisiting behavior, suggesting a causal role of the hippocampus in strategic learning of the spatial array (Voss et al., 2017).

The ability to construct or maintain a complex configural representation may be generally important for tasks involving cognitive search (Pachur et al., 2012). A recent paper by Solomon et al. (2019) considers this possibility in their examination of intracranial electroencephalographic activity during episodic memory search. They found correlations between hippocampal theta oscillations, and distances between studied items in both temporal (list position) and semantic (word meaning) spaces.

As hippocampal activity has already been implicated in coding of spatial environments (O'Keefe and Nadel, 1978), these results raise the possibility that hippocampus has a domain-general role in the formation, maintenance, and utilization of cognitive maps of any kind of information.

The visual search results described above are also consistent with the possibility that hippocampus supports search by allowing one to periodically “refresh” a target representation through episodic retrieval. By this account, the deficit in the feature generation task could arise from a difficulty in holding the target word in mind; if the representation of this word is disrupted, an HP patient would be unable to refresh it and continue the search. However, Klooster and Duff (2015) also found that patients were impaired in another semantic task, the Word Associates Test (WAT). In the WAT, all relevant materials are presented simultaneously and remain in view throughout the trial, obviating the need to rely on, or refresh, a representation held in memory. As such, we propose that the critical commonality between the semantic feature generation task and the WAT is the need to hold multiple disparate semantic features in mind simultaneously as part of a retrieval cue, in order to more effectively probe semantic memory. If the HP patient group is impaired in their ability to construct and maintain this retrieval cue, their ability to probe semantic space will be limited, regardless of whether the semantic space itself is degraded.

Relevant to this point, two other impairments related to hippocampal damage bear mentioning. First, individuals with hippocampal amnesia have difficulty forming a coherent mental image of a familiar scene during an imagination task. Fragmented images can be generated, but patients are impaired in relating these to one another to create a holistic representation (Hassabis et al., 2007). Second, individuals with hippocampal amnesia are impaired at constructing semantic narratives that are not autobiographically relevant (e.g., a fairy tale), producing fragmented narratives without clear temporal structure (Rosenbaum et al., 2009). These findings are consistent with a framework in which cognitive deficits in patients with hippocampal damage are not necessarily due to a deficit in the ability to retrieve experiences from memory *per se*, and more so due to a difficulty in assembling and relating disparate details to form a coherent, holistic representation (Kwan et al., 2013).

#### 4.3.2. Alternative Possibilities Regarding the Feature Generation Deficit

Two hypotheses regarding the functional consequences of hippocampal damage are worth considering. First, the possibility that the observed semantic deficits arise from an inability to retrieve autobiographical episodic memories during task performance (Ryan et al., 2008; Greenberg et al., 2009; Greenberg and Verfaellie, 2010), and second, the possibility that semantic knowledge is generally degraded by the absence of a hippocampally mediated consolidation process.

Under the first hypothesis, participants would draw upon autobiographical memories of interacting with a target item in order to generate semantic features. Indeed, in the data

collected by Klooster and Duff (2015) participants sometimes retrieve episodic memories in order to generate semantic features (e.g., for the target word “key,” “*I’ve got a padlock that your key sticks in and it actually screws the padlock shut*”). However, Klooster and Duff (2015) found no reliable differences in the frequency with which each group used personal anecdotes in their responses. Furthermore, the same amnesic patients showed semantic impairments in the Word Associates Test (WAT). As described above, the WAT tests the depth of one’s vocabulary knowledge, asking participants to decide which of several simultaneously presented words are related to a target word (either by meaning or collocation). It is unclear how drawing upon one’s autobiographical experience would help in this task. Furthermore, work characterizing a class of memories termed personal semantics suggests that in some cases the distinction between semantic and episodic memories may not be clear cut (Renoult et al., 2012). Some types of personal semantic memories are thought to be hippocampally dependent (e.g., memories for repeated or regularly recurring events), supporting the idea that there is not necessarily a rigid dichotomy between the episodic and semantic memory systems.

Under the second hypothesis, the periodic replay of episodic memories interleaves reactivation of older semantic memories and newly acquired information, limiting interference between older and newer memories, and generally curating one’s semantic memories. Without a hippocampus, it is possible that the semantic knowledge store is not sufficiently maintained, causing the representations to degrade over time (McClelland et al., 1995; O’Reilly and Rudy, 2001; O’Reilly and Norman, 2002). This could make it more difficult to retrieve information from semantic memory. In terms of the vector space model, degradation of the semantic representations (e.g., by adding noise to them) would tend to make related concepts become more distant from one another. This could explain why patients with hippocampal damage retrieve fewer features and preferentially retrieve features that are close in semantic space, as the more distant concepts may have become so distant as to be inaccessible. We believe this possibility deserves further consideration. The development of a more refined computational model of semantic search may prove informative. Such a model could examine whether the data are more consistent with a model in which hippocampus supports the search process itself (by allowing the discovery of more distant semantic relations) as opposed to a model in which hippocampus is not involved in the search process, but curates the knowledge being searched over.

## 5. CONCLUSIONS

Vector space models of semantic representational structure are valuable tools for the characterization of performance on semantic memory tests. Here, we showed that patients with hippocampal amnesia have difficulty generating features that are semantically distant from a target word. However, the semantic relatedness of produced features

to one another was unaffected. These results are broadly consistent with relational theories, in which hippocampus facilitates exploration of any cognitive representational space. We hope that these results will prove informative for future efforts to develop mechanistically explicit models of semantic memory search.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) of Vanderbilt University (160658). The patients/participants provided their written informed consent to participate in this study.

## REFERENCES

- Anderson, J. R. (1983). A spreading activation theory of memory. *J. Verb. Learn. Verb. Behav.* 22, 261–295.
- Barense, M. D., Gaffan, D., and Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia* 45, 2963–2974. doi: 10.1016/j.neuropsychologia.2007.05.023
- Bayley, P. J., and Squire, L. R. (2002). Medial temporal lobe amnesia: gradual acquisition of factual information by nondeclarative memory. *J. Neurosci.* 22, 5741–5748. doi: 10.1523/JNEUROSCI.22-13-05741.2002
- Bousfield, W. A., and Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *J. Gen. Psychol.* 30, 149–165.
- Burgess, N., Maguire, E. A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron* 35, 625–641. doi: 10.1016/S0896-6273(02)00830-9
- Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–608.
- Cohen, N. J., and Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MA: MIT.
- Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., and Nash, C. (1999). Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. *Hippocampus* 9, 83–98.
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82:407–428.
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* 16, 693–700. doi: 10.1016/j.conb.2006.10.012
- Davachi, L., and Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *J. Neurophysiol.* 88, 982–990. doi: 10.1152/jn.2002.88.2.982
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *J. Am. Soc. Inform. Sci.* 41:391–407.
- Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.* 1, 41–50. doi: 10.1038/35036213
- Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* 44, 109–120. doi: 10.1016/j.neuron.2004.08.028

## AUTHOR CONTRIBUTIONS

MD contributed the data. RC and SP designed the analyses and wrote the manuscript. RC performed statistical analysis. All authors contributed to the research questions, theoretical development, and manuscript revision.

## FUNDING

This work was supported by the National Institute of Health (R01 DC011755: MD); and the National Science Foundation (1756417: SP).

## ACKNOWLEDGMENTS

We thank Nathaniel Klooster and other members of the Communication and Memory Lab for collecting and transcribing the raw data. We also thank Computational Memory Lab members Emily Levine, Alice Li, Lauren Beal, and Blake Andreou for processing of the raw data.

- Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., and Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neurosci. Biobehav. Rev.* 36, 1597–1608. doi: 10.1016/j.neubiorev.2011.07.006
- Eichenbaum, H., Yonelinas, A., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Ann. Rev. Neurosci.* 30, 123–152. doi: 10.1146/annurev.neuro.30.051606.094328
- Elward, R. L., and Vargha-Khadem, F. (2018). Semantic memory in developmental amnesia. *Neurosci. Lett.* 680, 23–30. doi: 10.1016/j.neulet.2018.04.040
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., and Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an erp study. *Psychophysiology* 39, 133–146. doi: 10.1111/1469-8986.3920133
- Firth, J. R. (1957). “A synopsis of linguistic theory, 1930–1955,” in *Studies in Linguistic Analysis*, ed F. R. Palmer (London: Longmans), 168–205.
- Gabrieli, J. D., Cohen, N. J., and Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain Cogn.* 7, 157–177.
- Greenberg, D. L., Keane, M. M., Ryan, L., and Verfaellie, M. (2009). Impaired category fluency in medial temporal lobe amnesia: the role of episodic memory. *J. Neurosci.* 29, 10900–10908. doi: 10.1523/JNEUROSCI.1202-09.2009
- Greenberg, D. L., and Verfaellie, M. (2010). Effects of fixed-and varied-context repetition on associative recognition in amnesia. *J. Int. Neuropsychol. Soc.* 16, 596–602. doi: 10.1017/S1355617710000287
- Grilli, M. D., and Verfaellie, M. (2014). Personal semantic memory: insights from neuropsychological research on amnesia. *Neuropsychologia* 61, 56–64. doi: 10.1016/j.neuropsychologia.2014.06.012
- Gruenewald, P. J., and Lockhead, G. R. (1980). The free recall of category examples. *J. Exp. Psychol. Hum. Learn. Memory* 6:225. doi: 10.1037//0278-7393.6.3.225
- Hannula, D. E., Tranel, D., and Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26, 8352–8359. doi: 10.1523/JNEUROSCI.5222-05.2006
- Harris, Z. S. (1954). Distributional structure. *Word* 10, 146–162.
- Hassabis, D., Kumaran, D., Vann, S. D., and Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1726–1731. doi: 10.1073/pnas.0610561104
- Healey, M. K., Long, N. M., and Kahana, M. J. (2019). Contiguity in episodic memory. *Psychon. Bull. Rev.* 26, 699–720. doi: 10.3758/s13423-018-1537-3

- Henke, K., Weber, B., Kneifel, S., Wieser, H. G., and Buck, A. (1999). Human hippocampus associates information in memory. *Proc. Natl. Acad. Sci. U.S.A.* 96, 5884–5889.
- Henley, N. M. (1969). A psychological study of the semantics of animal terms. *J. Verb. Learn. Behav.* 8, 176–184.
- Hills, T. T., Jones, M. N., and Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychol. Rev.* 119, 431–440. doi: 10.1037/a0027373
- Hilverman, C., Cook, S. W., and Duff, M. C. (2017). The influence of the hippocampus and declarative memory on word use: patients with amnesia use less imageable words. *Neuropsychologia* 106, 179–186. doi: 10.1016/j.neuropsychologia.2017.09.028
- Horecka, K. M., Dulas, M. R., Schwarb, H., Lucas, H. D., Duff, M., and Cohen, N. J. (2018). Reconstructing relational information. *Hippocampus* 28, 164–177. doi: 10.1002/hipo.22819
- Howard, M. W., and Kahana, M. J. (2002a). A distributed representation of temporal context. *J. Math. Psychol.* 46, 269–299. doi: 10.1006/jmps.2001.1388
- Howard, M. W., and Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *J. Memory Lang.* 46, 85–98. doi: 10.1006/jmla.2001.2798
- Irish, M., Bunk, S., Tu, S., Kamminga, J., Hodges, J. R., Hornberger, M., and Piguet, O. (2016). Preservation of episodic memory in semantic dementia: the importance of regions beyond the medial temporal lobes. *Neuropsychologia* 81, 50–60. doi: 10.1016/j.neuropsychologia.2015.12.005
- Jeneson, A., Mauldin, K. N., and Squire, L. R. (2010). Intact working memory for relational information after medial temporal lobe damage. *J. Neurosci.* 30, 13624–13629. doi: 10.1523/JNEUROSCI.2895-10.2010
- Jones, M. N., and Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* 114, 1–37. doi: 10.1037/0033-295X.114.1.1
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory Cogn.* 24, 103–109.
- Kahana, M. J., Howard, M. W., and Polyn, S. M. (2008). “Associative retrieval processes in episodic memory,” in *Cognitive Psychology of Memory. Vol. 2 of Learning and Memory: A Comprehensive Reference, 4 vols. (J. Byrne, Ed.)*, ed H. L. Roediger III (Oxford: Elsevier), 467–490.
- Klooster, N. B., and Duff, M. C. (2015). Remote semantic memory is impoverished in hippocampal amnesia. *Neuropsychologia* 79, 42–52. doi: 10.1016/j.neuropsychologia.2015.10.017
- Konkel, A., and Cohen, N. J. (2009). Relational memory and the hippocampus: representations and methods. *Front. Neurosci.* 3:23. doi: 10.3389/neuro.01.023.2009
- Konkel, A., Warren, D. E., Duff, M. C., Tranel, D., and Cohen, N. J. (2008). Hippocampal amnesia impairs all manner of relational memory. *Front. Hum. Neurosci.* 2:15. doi: 10.3389/neuro.09.015.2008
- Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci.* 3:946. doi: 10.1038/78868
- Kwan, D., Craver, C. F., Green, L., Myerson, J., and Rosenbaum, R. S. (2013). Dissociations in future thinking following hippocampal damage: evidence from discounting and time perspective in episodic amnesia. *J. Exp. Psychol. Gen.* 142:1355–1369. doi: 10.1037/a0034001
- Kwantes, P. J. (2005). Using context to build semantics. *Psychon. Bull. Rev.* 12, 703–710. doi: 10.3758/BF03196761
- Landauer, T. K., and Dumais, S. T. (1997). Solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240.
- Lucas, H. D., Duff, M. C., and Cohen, N. J. (2018). The hippocampus promotes effective saccadic information gathering in humans. *J. Cog. Neurosci.* 31, 186–201. doi: 10.1162/jocn-a-01336
- Lund, K., and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instru. Comput.* 28, 203–208.
- Manns, J. R., Hopkins, R. O., and Squire, L. R. (2003). Semantic memory and the human hippocampus. *Neuron* 38, 127–133. doi: 10.1016/S0896-6273(03)00146-6
- 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.
- McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559. doi: 10.3758/BF03192726
- Milner, B., Squire, L. R., and Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron* 20, 445–468.
- Mishkin, M., Ungerleider, L., and Macko, K. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417.
- Morton, N. W., and Polyn, S. M. (2016). A predictive framework for evaluating models of semantic organization in free recall. *J. Memory Lang.* 86, 119–140. doi: 10.1016/j.jml.2015.10.002
- Newcombe, F. (1969). Missile wounds of the brain: a study of psychological deficits. *J. Neurol. Neurosurg. Psychiatry* 33: 551.
- Norman, K. A., and O’Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* 110:611–646. doi: 10.1037/0033-295X.110.4.611
- O’Kane, G., Kensinger, E. A., and Corkin, S. (2004). Evidence for semantic learning in profound amnesia: an investigation with patient hm. *Hippocampus* 14, 417–425. doi: 10.1002/hipo.20005
- O’Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.
- Olsen, R. K., Moses, S. N., Riggs, L., and Ryan, J. D. (2012). The hippocampus supports multiple cognitive processes through relational binding and comparison. *Front. Hum. Neurosci.* 6:146. doi: 10.3389/fnhum.2012.00146
- Olson, I. R., and Newcombe, N. S. (2013). Binding together the elements of episodes: Relational memory and the developmental trajectory of the hippocampus. *Wiley Handbook Dev. Child. Memory* 1, 285–308. doi: 10.1002/9781118597705.ch13
- Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., and Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.* 26, 4596–4601. doi: 10.1523/JNEUROSCI.1923-05.2006
- O’Reilly, R. C., and Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends Cogn. Sci.* 6, 505–510. doi: 10.1016/S1364-6613(02)02005-3
- O’Reilly, R. C., and Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev.* 108:311–345. doi: 10.1037/0033-295x.108.2.311
- Pachur, T., Raaijmakers, J. G., Davelaar, E. J., Daw, N. D., Dougherty, M. R., Hommel, B., et al. (2012). “Chapter 15: Unpacking cognitive search: mechanisms and processes,” in *Cognitive Search: Evolution, Algorithms, and the Brain*, eds P. M. Todd, T. T. Hills, and T. W. Robbins (Cambridge, MA: The MIT Press), 237–253.
- Pennington, J., Socher, R., and Manning, C. (2014). “Glove: global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Doha: Association for Computational Linguistics), 1532–1543. doi: 10.3115/v1/D14-1162
- Polyn, S. M., Erlichman, G., and Kahana, M. J. (2011). Semantic cuing and the scale-insensitivity of recency and contiguity. *J. Exp. Psychol. Learn. Memory Cogn.* 27, 766–775. doi: 10.1037/a0022475
- Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* 116, 129–156. doi: 10.1037/a0014420
- Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* 13, 587–597. doi: 10.1038/nrn3251
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107. doi: 10.1038/nature03687
- Race, E., Keane, M. M., and Verfaellie, M. (2013). Losing sight of the future: impaired semantic prospection following medial temporal lobe lesions. *Hippocampus* 23, 268–277. doi: 10.1002/hipo.22084
- Ralph, M. A. L., 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
- Read, J. (1993). The development of a new measure of l2 vocabulary knowledge. *Lang. Test.* 10, 355–371.
- Read, J. (1998). “Validating a test to measure depth of vocabulary knowledge,” in *Validation in Language Assessment*, ed A. Kunnan (Mahwah, NJ: Lawrence Erlbaum), 41–60.



- Reed, J. M., and Squire, L. R. (1998). Retrograde amnesia for facts and events: findings from four new cases. *J. Neurosci.* 18, 3943–3954.
- Renoult, L., Davidson, P. S., Palombo, D. J., Moscovitch, M., and Levine, B. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends Cogn. Sci.* 16, 550–558. doi: 10.1016/j.tics.2012.09.003
- Romney, A. K., Brewer, D. D., and Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychol. Sci.* 4, 28–34.
- Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., and Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: evidence from personal, fictional, and semantic narratives in kc. *Neuropsychologia* 47, 2181–2187. doi: 10.1016/j.neuropsychologia.2008.11.028
- Rubin, R. D., Watson, P. D., Duff, M. C., and Cohen, N. J. (2014). The role of the hippocampus in flexible cognition and social behavior. *Front. Hum. Neurosci.* 8:742. doi: 10.3389/fnhum.2014.00742
- Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychol. Sci.* 11, 454–461. doi: 10.1111/1467-9280.00288
- Ryan, L., Cox, C., Hayes, S., and Nadel, L. (2008). Hippocampal activation during episodic and semantic memory retrieval: comparing category production and category cued recall. *Neuropsychologia* 46, 2109–2121. doi: 10.1016/j.neuropsychologia.2008.02.030
- Schmolck, H., Kensinger, E. A., Corkin, S., and Squire, L. R. (2002). Semantic knowledge in patient hm and other patients with bilateral medial and lateral temporal lobe lesions. *Hippocampus* 12, 520–533. doi: 10.1002/hipo.10039
- Sederberg, P. B., Howard, M. W., and Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychol. Rev.* 115, 893–912. doi: 10.1037/a0013396
- Sharon, T., Moscovitch, M., and Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1146–1151. doi: 10.1073/pnas.1005238108
- Sheldon, S., and Moscovitch, M. (2012). The nature and time-course of medial temporal lobe contributions to semantic retrieval: an fmri study on verbal fluency. *Hippocampus* 22, 1451–1466. doi: 10.1002/hipo.20985
- Sheldon, S., Romero, K., and Moscovitch, M. (2013). Medial temporal lobe amnesia impairs performance on a free association task. *Hippocampus* 23, 405–412. doi: 10.1002/hipo.22099
- Sirotni, Y. B., Kimball, D. R., and Kahana, M. J. (2005). Going beyond a single list: modeling the effects of prior experience on episodic free recall. *Psychon. Bull. Rev.* 12, 787–805. doi: 10.3758/BF03196773
- Smith, K. A., Huber, D. E., and Vul, E. (2013). Multiply-constrained semantic search in the remote associates test. *Cognition* 128, 64–75. doi: 10.1016/j.cognition.2013.03.001
- Smith, M. L., and Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia* 19, 781–793.
- Socher, R., Gershman, S., Sederberg, P., Norman, K., Perotte, A. J., and Blei, D. M. (2009). “A Bayesian analysis of dynamics in free recall,” in *Advances in Neural Information Processing Systems 22*, eds Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Vancouver, BC: Curran Associates, Inc.), 1714–1722. Available online at: <http://papers.nips.cc/paper/3720-a-bayesian-analysis-of-dynamics-in-free-recall.pdf>
- Solomon, E. A., Lega, B. C., Sperling, M. R., and Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *bioRxiv* 611681.
- Stan Development Team (2017). *RStan: The R Interface to Stan*. R package version 2.17.2. Stan Development Team.
- Staresina, B. P., and Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron* 63, 267–276. doi: 10.1016/j.neuron.2009.06.024
- Steyvers, M., Shiffrin, R. M., and Nelson, D. L. (2004). “Word association spaces for predicting semantic similarity effects in episodic memory,” in *Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*, ed A. F. Healy (Washington, DC: American Psychological Association), 237–249.
- Tulving, E. (1972). “Episodic and semantic memory,” in *Organization of Memory*, eds E. Tulving, W. Donaldson (New York, NY: Academic Press), 381–403.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connely, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277, 376–380.
- Verfaellie, M., Bousquet, K., and Keane, M. M. (2014). Medial temporal and neocortical contributions to remote memory for semantic narratives: evidence from amnesia. *Neuropsychologia* 61, 105–112. doi: 10.1016/j.neuropsychologia.2014.06.018
- Verfaellie, M., Koseff, P., and Alexander, M. (2000). Acquisition of novel semantic information in amnesia: effects of lesion location. *Neuropsychologia* 38, 484–492. doi: 10.1016/S0028-3932(99)00089-5
- Voorspoels, W., Storms, G., Longenecker, J., Verheyen, S., Weinberger, D. R., and Elvevåg, B. (2014). Deriving semantic structure from category fluency: clustering techniques and their pitfalls. *Cortex* 55, 130–147. doi: 10.1016/j.cortex.2013.09.006
- Voss, J. L., Bridge, D. J., Cohen, N. J., and Walker, J. A. (2017). A closer look at the hippocampus and memory. *Trends Cogn. Sci.* 21, 577–588. doi: 10.1016/j.tics.2017.05.008
- Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., and Cohen, N. J. (2011a). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nat. Neurosci.* 14, 115–120. doi: 10.1038/nn.2693
- Voss, J. L., Warren, D. E., Gonsalves, B. D., Federmeier, K. D., Tranel, D., and Cohen, N. J. (2011b). Spontaneous revisitation during visual exploration as a link among strategic behavior, learning, and the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 108, E402–E409. doi: 10.1073/pnas.1100225108
- Warren, D. E., and Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus* 24, 920–933. doi: 10.1002/hipo.22279
- Warren, D. E., Duff, M. C., Cohen, N. J., and Tranel, D. (2015). Hippocampus contributes to the maintenance but not the quality of visual information over time. *Learn. Memory* 22, 6–10. doi: 10.1101/lm.037127.114
- Warren, D. E., Duff, M. C., Jensen, U., Tranel, D., and Cohen, N. J. (2012). Hiding in plain view: lesions of the medial temporal lobe impair online representation. *Hippocampus* 22, 1577–1588. doi: 10.1002/hipo.21000
- Warren, D. E., Duff, M. C., Tranel, D., and Cohen, N. J. (2011). Observing degradation of visual representations over short intervals when medial temporal lobe is damaged. *J. Cogn. Neurosci.* 23, 3862–3873. doi: 10.1162/jocn-a-00089
- Watson, P. D., Voss, J. L., Warren, D. E., Tranel, D., and Cohen, N. J. (2013). Spatial reconstruction by patients with hippocampal damage is dominated by relational memory errors. *Hippocampus* 23, 570–580. doi: 10.1002/hipo.22115
- Westmacott, R., and Moscovitch, M. (2002). Temporally-graded retrograde memory loss for famous names and vocabulary terms in amnesia and semantic dementia: further evidence for opposite gradients using implicit memory tasks. *Cogn. Neuropsychol.* 19, 135–163. doi: 10.1080/02643290143000123
- Winocur, G., Moscovitch, M., and Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia* 48, 2339–2356. doi: 10.1016/j.neuropsychologia.2010.04.016
- Wittgenstein, L. (1953). *Philosophical Investigations*. Transl. by G. E. M. Anscombe. Oxford: Wiley-Blackwell.

**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 © 2019 Cutler, Duff and Polyn. 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.



# Adult Age Differences in the Use of Conceptual Combination as an Associative Encoding Strategy

Heather D. Lucas<sup>1\*</sup>, Resh S. Gupta<sup>2</sup>, Ryan J. Hubbard<sup>3</sup> and Kara D. Federmeier<sup>3</sup>

<sup>1</sup> Department of Psychology, Louisiana State University, Baton Rouge, LA, United States, <sup>2</sup> Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN, United States, <sup>3</sup> Department of Psychology and Beckman Institute, University of Illinois Urbana-Champaign, Urbana, IL, United States

## OPEN ACCESS

### Edited by:

Vitória Plai,  
Radboud University Nijmegen,  
Netherlands

### Reviewed by:

Rocio Lopez Zunini,  
University of Ottawa, Canada  
Regine Bader,  
Saarland University, Germany

### \*Correspondence:

Heather D. Lucas  
hlucas2@lsu.edu

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 21 June 2019

**Accepted:** 17 September 2019

**Published:** 10 October 2019

### Citation:

Lucas HD, Gupta RS,  
Hubbard RJ and Federmeier KD  
(2019) Adult Age Differences  
in the Use of Conceptual Combination  
as an Associative Encoding Strategy.  
*Front. Hum. Neurosci.* 13:339.  
doi: 10.3389/fnhum.2019.00339

It is well-established that aging impairs memory for associations more than it does memory for single items. Aging also impacts processes involved in online language comprehension, including the ability to form integrated, message-level representations. These changes in comprehension processes could impact older adults' associative memory performance, perhaps by reducing or altering the effectiveness of encoding strategies that encourage semantic integration. The present study examined age differences in the use of a strategy termed conceptual combination, which involves integrating two words (e.g., "winter" and "salad") into a single concept ("a salad for winter"). We recorded ERPs while participants studied unrelated noun pairs using a strategy that either did or did not encourage conceptual combination. We also varied the concreteness of the first noun in each pair in order to measure compositional concreteness effects, or ERP differences at the second noun due to the concreteness of the first noun. At the first nouns, older adults showed word-level concreteness effects that were similar to those of younger adults. However, compositional concreteness effects were diminished in older adults, consistent with reduced semantic integration. Older adults' associative memory performance was better for word pairs studied during the conceptual combination task versus the non-combinatory encoding task; however, the magnitude of the age-related associative memory deficit did not differ between tasks. Finally, analyses of both memory accuracy and trial-by-trial ratings of perceived combination success suggested that older adults had disproportionate difficulty applying the conceptual combination strategy to word pairs that began with abstract nouns. Overall, these results indicate that changes to integrative language processing that occur with age are not independent of – and may sometimes exacerbate – age-related memory decline.

**Keywords:** conceptual combination, associative memory, cognitive aging, ERPs, imagery

## INTRODUCTION

A hallmark of the typical aging process is a reduced ability to learn and remember new information. However, aging does not uniformly affect all types of memory. Memory for associative or relational information (e.g., arbitrary pairings or groupings of stimuli, such as face-name pairings) is particularly susceptible to age-related decline, whereas the ability to remember single items is relatively spared (for a recent meta-analysis, see Old and Naveh-Benjamin, 2008). These behavioral

findings converge well with data from studies examining age effects on brain structure and function. Associative memory tasks place strong demands on the hippocampus and certain regions of the prefrontal cortex (Vargha-Khadem et al., 1997; Preston and Eichenbaum, 2013; Addis et al., 2014), both of which decrease in volume and integrity (Raz et al., 2005) as well as encoding-related activity (Dennis et al., 2008) across the adult lifespan. By contrast, item memory has been linked to the surrounding medial temporal lobe (MTL) cortex, particularly the perirhinal cortex (e.g., Davachi, 2006), which is less susceptible to changes with age (Dickerson et al., 2009).

In recent years, there has been a growing interest in the idea that individuals from populations with reduced hippocampal integrity may be able to develop strategies to increase their ability to rely on spared item memory to remember certain types of associations. Indeed, under the right circumstances, otherwise arbitrary associations can be represented in memory in a manner that is relatively *unitized* or item-like. For example, in a neuroimaging investigation of color-object associations (Diana et al., 2010), the perirhinal cortex was found to be more active during retrieval when the color was initially encoded as intrinsic to, rather than arbitrarily co-occurring with, the object. Other studies (e.g., Quamme et al., 2007; Haskins et al., 2008) have examined the extent to which a form of processing termed *conceptual combination* can be strategically applied to arbitrary word pairs in order to achieve relatively unitized memory representations. Conceptual combination refers to the processing of noun pairs as modifier-head dyads that together form new, emergent concepts. For example, applying conceptual combination to the word pair “dog spoon” might prompt an interpretation such as “a spoon that was designed specifically to feed dogs.”

The link between conceptual combination and unitization has been established in studies of patients with amnesia due to damage to the MTL. Quamme et al. (2007) asked both healthy participants and patients with MTL lesions to study unrelated word pairs under two conditions that either did or did not promote conceptual combination. In a so-called separate encoding condition, participants were shown each word pair along with a sentence with two corresponding blanks (e.g., “The \_\_\_\_ could be seen from the \_\_\_\_” for the word pair “cloud-lawn”). By contrast, in the compound encoding condition, each word pair was accompanied by an experimenter-generated definition that served to combine the words into a novel but meaningful concept (e.g., “a yard used for sky-gazing”). On a later test of associative recognition, participants with hippocampal damage performed markedly better when tested on items they studied in the conceptual combination condition. However, participants with more widespread temporal lobe injury that also encompassed the perirhinal cortex did not show this advantage, nor did healthy controls for whom both hippocampal and perirhinal processing were presumably intact (see Haskins et al., 2008, for converging evidence from fMRI).

Together, these findings suggest that the process of conceptually combining novel word pairs – when successful – can reduce the associative memory deficit observed in populations characterized by hippocampal decline, potentially including

older adults (Ahmad et al., 2014; Bastin et al., 2013). However, very little is known about how the conceptual combination process itself might change with age, particularly when applied in an *ad hoc*, flexible manner to word pairs that do not correspond to pre-existing definitions. Indeed, one recent study (Kamp et al., 2018) produced the counterintuitive finding that older adults’ associative memory for unrelated word pairs was worse when those word pairs were presented as part of an experimenter-defined compound phrase versus as part of a sentence – a pattern that is opposite to what has been found in patients with hippocampal amnesia.

Of note, there is evidence that age can impact processes related to semantic integration, or the construction and maintenance of message-level representations from language. Several studies using event-related potentials (ERPs, reviewed in Wlotko et al., 2010) have demonstrated that older adults show a reduced tendency to use sentential context to guide or constrain the processing of upcoming words. Analyses focusing on the N400, an electrophysiological index of the processing demands associated with semantic access, have been informative in this regard. N400 amplitudes are modulated by item-level lexical attributes, such as word frequency and orthographic neighborhood size, as well as the “fit” of incoming information with the preceding semantic or syntactic context (Kutas and Federmeier, 2011). Compared to younger individuals, older adults demonstrate a decreased sensitivity of N400 potentials to contextual information, combined with a spared or even increased sensitivity to lexical characteristics (Payne and Federmeier, 2018), suggesting a diminished capacity for rapid and flexible construction and/or use of context from semantic information.

It seems plausible that this reduced ability to build up contextual information and/or integrate it with incoming stimulus-based information could contribute to difficulties in implementing strategies that involve conceptual combination, particularly when applied to word pairs that are pre-experimentally unrelated. Of particular relevance, Huang and colleagues (Huang et al., 2010, 2012) demonstrated that age effects on semantic integration extend to simple two-word phrases, not unlike the stimuli used in many associative memory experiments. In these studies, ERPs were recorded while younger and older adults viewed a series of common nouns, each of which was alternately preceded by either a concrete or an abstract adjective (for example, “hilly farm” versus “productive farm”). Both age groups showed robust concreteness effects on ERP responses to the adjectives, consistent with reports that aging does not reduce sensitivity to lexical characteristics. In particular, relative to abstract adjectives, concrete adjectives elicited more negative N400 potentials, as well as enhanced amplitudes of a late frontal negativity, sometimes referred to as the N700, which has been linked to either visual imagery (West and Holcomb, 2000; Welcome et al., 2011; Gullick et al., 2013) or to a modality-independent feature integration process (Barber et al., 2013). Importantly, only in the young adults were concreteness effects evident at the compositional level, or in response to the same head nouns as a function of the concreteness of the preceding adjective. Mirroring world-level concreteness effects, nouns that

had been modified by a concrete adjective elicited smaller N400 potentials and larger N700 potentials than did the same nouns when modified by abstract adjectives. By contrast, compositional concreteness effects were absent in the older adults, suggesting a reduced ability to incorporate the features specified by the adjective into the meaning elicited by the head noun.

In the present study, we build on these findings by examining whether age differences are also present during a noun-noun conceptual combination task, similar to tasks that have been suggested to promote the formation of unitized memory representations. In a recent study (Lucas et al., 2017), we demonstrated that the N400 and N700 compositional concreteness effects found for adjective-noun processing were also evident during the processing of unrelated noun pairs (e.g., “road salad” versus “idea salad”). The later of these two compositional concreteness effects (N700) was found to be task specific, in that N700 differences were present on the second noun only when participants were encouraged to engage in conceptual combination by attempting to generate a sensible compound meaning for each word pair. When the same word pairs were processed in a task that involved comparing the relative frequency of the concepts denoted by the two words, only N400 compositional concreteness effects were present. Moreover, a subsequent free recall test revealed that word pairs that had been initially processed via conceptual combination were represented in memory in a more holistic manner, in that they tended to be recalled from memory as pairs rather than individual items, consistent with the notion that conceptual combination promoted unitization.

Together, these data suggest that (1) N700 concreteness effects, when present at the compositional level (e.g., on the same lexical item as a function of the concreteness of a preceding modifier), reflect an aspect of semantic integration that can be deployed in a top-down manner to support the ability to interpret novel concepts, and (2) doing so promotes the formation of strong and perhaps unitized associative memory representations in younger adults. As such, the design employed by Lucas et al. (2017) provides a starting point to examine the extent to which age-related decreases in integrative processes associated with language comprehension can impact the online conceptual combination process, thereby limiting older adults’ use of conceptual combination as a strategy to remediate associative memory deficits.

In this experiment, we used the same materials and procedures from Lucas et al. (2017) in a sample of healthy older adults. ERPs were recorded as participants studied unrelated noun pairs (either abstract-concrete or concrete-concrete) under instructions that either did or did not emphasize conceptual combination. In conceptual combination blocks, participants were asked to generate plausible compound definitions for each of the word pairs and provide trial-by-trial subjective ratings of their success in doing so. In frequency-comparison blocks, participants judged whether the concept denoted by the first word (W1) was one that they encountered more or less often relative to the concept denoted by the second word (W2). Given prior evidence for spared word-level processing in older adults, we predict that, regardless of task, older adults will show the

canonical pattern of lexical-level ERP concreteness effects on the W1s, including larger N400 responses and a sustained frontal negativity to concrete as compared with abstract words. However, in line with Huang et al. (2012), we do not expect that older adults will show the N700 compositional concreteness effects that were previously found in younger adults during conceptual combination. Rather, we expect that the N700 ERPs elicited by the W2s in older adults will be insensitive both to task demands (frequency comparison versus conceptual combination) and to the concreteness of the preceding W1. To examine age effects, we also compared older adults’ ERP data to those of the younger adults described in Lucas et al. (2017). We predict that significant age differences will be present in the magnitude of compositional, but not item-level, concreteness effects, and that these differences will be specific to the conceptual combination task.

In addition to examining encoding-related ERPs, we conducted multiple complementary analyses to better understand the underlying mechanisms and downstream effects of flexible conceptual combination in older adults. First, we tested associative recognition memory after each block to examine the extent to which the benefits enjoyed by young adults following conceptual combination are also present in older adults. Second, we examined whether participants’ trial-by-trial ratings of conceptual combination success predicted subsequent memory for word pairs studied during the conceptual combination task. To foreshadow these results, we found that participants from both age groups assigned significantly higher ratings to word pairs that they went on to remember versus those that they did not. This finding provides behavioral evidence for overlap between the online processes involved in flexible conceptual combination and those that facilitate associative memory formation. As such, we build on these results by using a single-trial analysis approach to assess relationships between ERPs linked to conceptual combination and perceived success across trials.

## MATERIALS AND METHODS

### Participants

Twenty-four older adults (13 female, mean age = 68.6 years, range = 61–79 years) from the Champaign–Urbana area participated in the study and were compensated \$10/hour. All were right-handed and native speakers of English. An additional five individuals completed the experiment but were excluded from analyses due to difficulty following task instructions ( $n = 1$ ), poor EEG data quality ( $n = 2$ ), or because they scored lower than a 51/57 on a modified version of the Mini Mental State Examination ( $n = 2$ ), which was administered prior to beginning the study. The average score for the included participants was 54.5 (range 51–57).

To examine age effects, key variables of interest were compared with the sample of 24 younger adults reported in Experiment 1 of Lucas et al. (2017). These participants were from the University of Illinois and surrounding areas. All were right-handed and native speakers of English. The mean age of this sample was 21 years (range = 18–24 years, 18 female). The study design, equipment, and procedures employed in the



young adult experiment were identical to those in the present experiment, except that the Mini Mental State Examination was not administered to the younger adults.

## Stimuli

The stimuli were the same as those used in Lucas et al. (2017) and consisted of 144 noun pairs (72 abstract-concrete pairs and 72 concrete-concrete pairs) formed from a set of 72 abstract nouns and 216 concrete nouns. Abstract and concrete nouns had a mean concreteness rating of 280 (range = 232–349) and 574 (range = 500–646), respectively, a mean imageability rating of 383 (range = 262–551) and 564 (range = 424–667), respectively, and a mean Kucera-Francis written frequency rating of 56 (range = 1–447) and 41 (range = 1–442), respectively, according to the Medical Research Council database described by Coltheart (2007, [http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa\\_mrc.htm](http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm)).

To construct the word pairs, each of the 72 abstract nouns was paired with a concrete noun of comparable frequency, familiarity, and length. These 144 nouns served as the first words (W1s). Each pair of yoked W1s was assigned two randomly chosen words from the remaining set of 144 concrete nouns to serve as second words (W2s). By creating two sets of word pairings in this manner, we were able to counterbalance the frequency with which each W2 was preceded by a concrete versus an abstract W1 across participants. The frequency with which each word pair was presented in a conceptual combination block versus a frequency-comparison block was also counterbalanced.

We manually inspected and adjusted noun pairings to eliminate pairs with clear pre-experimental meanings

(e.g., “flower store”). In addition, we used the University of South Florida Free Association Norms database (Nelson et al., 2004) to obtain free association data for 109 of the 144 W1s. For these words, we were able to confirm that the corresponding W2s were not forward associates of the W1s.

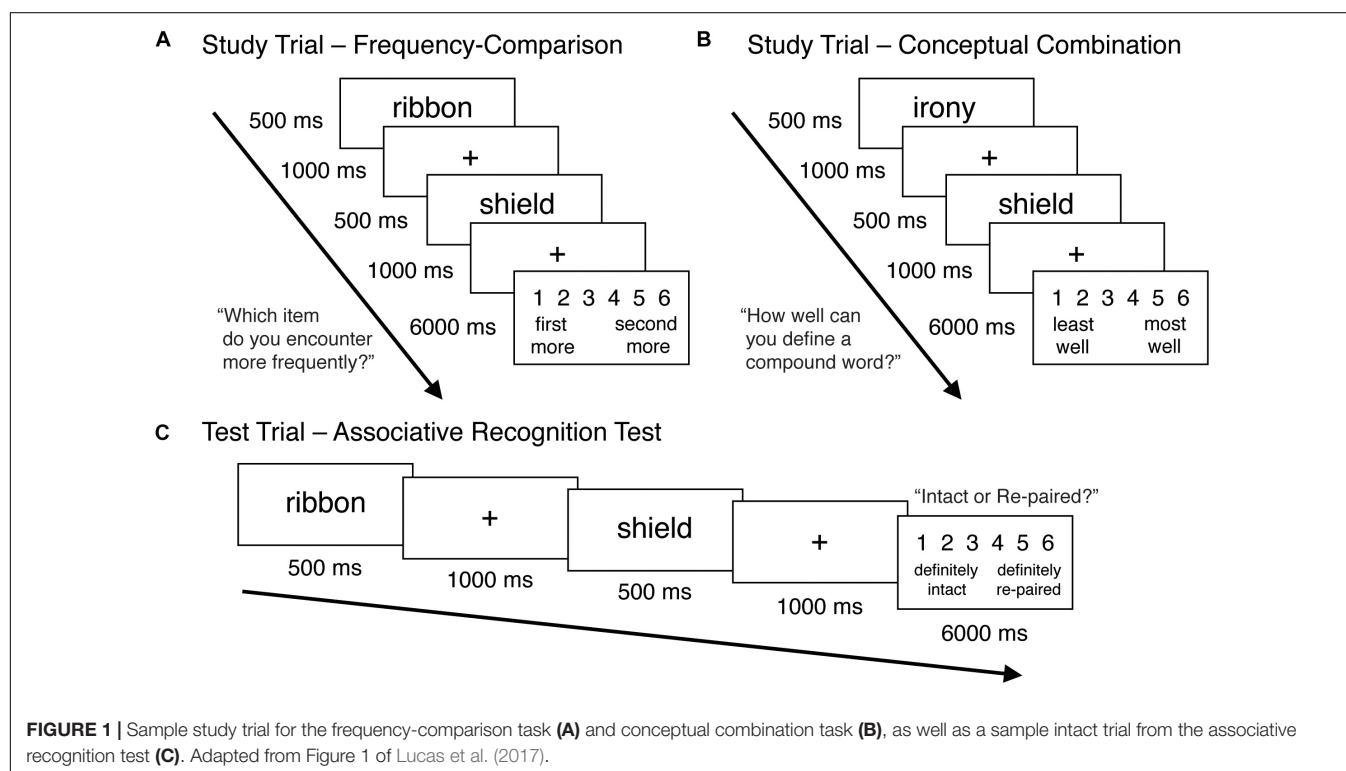
## Procedures

The procedures were the same as those of Lucas et al. (2017). Each participant completed four study-test blocks, two of which were conceptual combination blocks and two of which were frequency-comparison blocks. Both blocks of each type were completed consecutively, and presentation order was counterbalanced across participants.

## Study Phases

Sample study trials for the frequency-comparison and conceptual combination tasks are depicted in Figures 1A,B, respectively. Each study phase consisted of 36 word pairs (18 abstract-concrete and 18 concrete-concrete), which were presented in a random order, as well as one primacy and one recency buffer trial. In each study trial, a W1 was presented for 500 ms, followed by a fixation cross presented for 1000 ms, and then a 500 ms presentation of the W2. A fixation cross then appeared again for 1000 ms, followed by a 5000 ms prompt in which the appropriate rating scale was displayed, and participants were asked to make a response.

During the conceptual combination blocks, participants were asked to use a 1–6 scale to indicate the relative ease with which they could generate a definition for the compound phrase formed from the two words. For half of the participants, a rating of “1”



was assigned to pairs for which no meaning came to mind at all, and a rating of “6” corresponded to word pairs that were easiest to clearly define. Buttons 2–5 reflected intermediate levels of difficulty. These mappings were reversed for the remaining participants. To help ensure that the ratings would provide information about the relative ease of defining each word pair, we asked participants to make these ratings on a relative basis and to use the six buttons approximately evenly. We also emphasized that, while an attempt should be made on each trial to construct a meaningful definition, it was likely that for some trials no definition might come to mind, and that participants should press Button 1 (or 6, depending on counterbalancing) if they were unable to generate any definition.

During the frequency-comparison blocks, participants were instructed to use a 1–6 scale to indicate which of the items denoted by the two words is encountered more frequently. For half of the participants, a rating of “1” indicated a much greater frequency of the first word, while a rating of “6” indicated a much greater frequency of the second word. Intermediate buttons were used for less extreme frequency differences. Ratings were reversed for the remaining participants.

## Test Phases

In between the study and test phase, participants completed a brief (30 s) distractor task in which they were asked to count backwards in twos from a randomly generated number between 300 and 600. After the distractor task, an associative recognition test was administered. A sample test trial is depicted in **Figure 1C**. Each test phase consisted of the 36 pairs from the most recent study block, half of which were intact (presented in the same pairing as they had been during the study phase), and half of which were re-paired (paired with a different word from the same block). Re-pairings were determined randomly for each participant within each counterbalancing condition, with the following constraints: (1) W1 words remained first in each pair and were always re-paired with W2 words, (2) each W2 was paired with a W1 at test with the same concreteness status as its W1 from the study phase, and (3) an equal number of abstract-concrete and concrete-concrete word pairs were presented as intact or re-paired within each block. Test blocks began with the primacy and recency buffers of the previous study block, which were used as practice buffer trials and not included in analyses.

The timing and structure of the test trials were the same as the study trials, except that participants were instructed to provide confidence ratings as to whether each word pair was intact or re-paired on a scale of 1–6. For half of the participants, Button 1 corresponded to a high degree of confidence that a pair was intact, while Button 6 indicated a high degree of confidence that the pair was re-paired. Buttons 2 and 5 were used to indicate medium levels of confidence, and buttons 3 and 4 denoted low levels of confidence. These ratings were reversed for the remaining participants.

Before beginning the experiment, participants completed brief, six-trial practice blocks for each of the two study tasks, followed by a practice associative recognition test. During the practice conceptual combination block, participants were asked to verbally describe the definition they produced for each trial

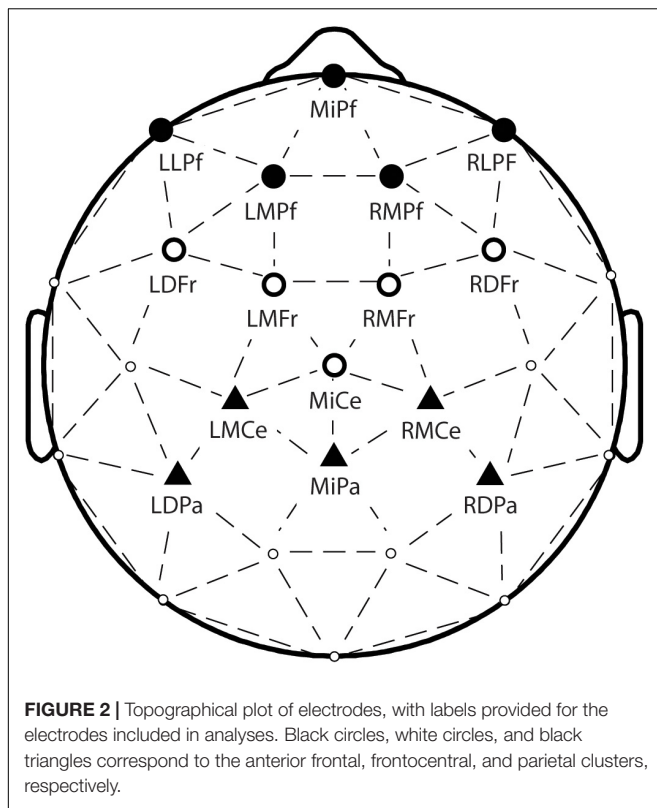
and explain their choice of button press. After the participant described their definition (or expressed an inability to come up with a definition), the experimenter offered multiple different examples of possible responses to reinforce the instructions.

## Electrophysiology

ERPs were extracted from scalp electroencephalographic recordings from 26 Ag/AgCl electrodes spaced evenly over the head. Voltages were referenced online to a left mastoid electrode and re-referenced offline to averaged left and right mastoids. Electrode impedances were kept below 5 k $\Omega$ . Signals were recorded with a bandpass filter of 0.01–100 Hz and sampled at a rate of 1000 Hz (BrainVision system). A bandpass filter of 0.1–30 Hz was applied offline prior to statistical analyses. An additional 10 Hz low-pass filter was applied to grand averages for display purposes only. Data preprocessing and analyses were conducted using EEGLAB (Delorme and Makeig, 2004).

Eye movements and blinks were recorded from three additional electrodes below the center of the left eye and on the outer canthus of each eye. Datasets in which more than 25% of epochs contained blink artifacts ( $n = 16$ ) were individually subjected to independent components analyses using Adaptive Mixture ICA (AMICA, Palmer et al., 2012), after which blink components were identified using a combination of manual inspection and by calculating the extent to which activity in each component correlated across time with activity in the bipolar eyeblink channel (calculated as the difference between the VEOG channel and channel LMPF, which is located at the front of the head above the left eye). Blink components were then removed from the EEG only on those trials that were identified as containing blink artifacts. For datasets in which <25% of trials contained blink artifacts ( $n = 8$ ), blink trials were excluded from analyses and no ICA was performed. We then screened for trials containing artifacts due to saccades, muscle activity, and residual eyeblink activity using a simple rejection threshold of  $\pm 75 \mu\text{V}$  on any scalp channel and a moving window rejection threshold  $\pm 40 \mu\text{V}$  (based on 200 ms windows and a window step of 10 ms) in the bipolar eyeblink and bipolar horizontal eye movement channels. These rejection decisions were then titrated individually using condition-blinded visual inspection to maximize correct rejection of artifacts and minimize loss of clean data. All told, an average of 8.3% of trials (range = 0.08–22.2) were excluded from analysis for each participant.

Each trial consisted of a 1000 ms epoch time-locked to stimulus onset. The mean amplitude of a 200 ms window prior to stimulus onset was subtracted to correct for baseline variability. As in Lucas et al. (2017), statistical comparisons were performed on amplitudes averaged over an anterior frontal electrode cluster (channels MiPf, LLPf, RLPf, LMPf, and RMPf), a frontocentral electrode cluster (channels LDFr, RDFr, LMFr, RMFr, and MiCe) and a parietal electrode cluster (MiPa, LMCE, RMCE, LDPA, RDPa). These channel locations are depicted in **Figure 2**. The first letter(s) of the abbreviations denote left (L), right (R), and midline electrodes (Mi). The second letter describes lateral (L), medial (M), and dorsal (D) locations. The final letters denote anteriority, as prefrontal (Pf), frontal (Fr), central (Ce), and parietal (Pa).



Occipital and temporal electrodes were not analyzed, as no experimental effects were expected to occur in these regions.

ERP comparisons were performed using repeated-measures ANOVAs (criterion  $p = 0.05$ ) over the windows of 300–500 ms (to capture N400 effects) and 700–1000 ms (to capture N700 effects), consistent with Lucas et al. (2017). The window of 300–500 ms was chosen for N400 analyses based on extensive prior research (e.g., Federmeier and Laszlo, 2009; Kutas and Federmeier, 2011). N700 latencies in previous studies have varied somewhat but tend to occur in the proximity of 700 ms (Holcomb et al., 1999; West and Holcomb, 2000; Huang et al., 2010; Welcome et al., 2011; Barber et al., 2013; Gullick et al., 2013). Mauchly's Test for Sphericity was used to test for sphericity violations when analyses involved three or more levels of a repeated measure, and the Greenhouse-Geisser correction was applied whenever non-sphericity was present.

We used a twofold strategy to compare ERP effects across age groups. Our first approach was to compute difference waves for each age group via a point-by-point subtraction of the ERP responses for the relevant contrast (e.g., concrete – abstract words) over the 300–500 ms and 700–1000 ms time windows, and then enter the results into a one-way, between-subjects ANOVA with age group as a factor. This analysis method is functionally the same as testing the age  $\times$  condition interaction.

Our second approach to examining age differences was to submit older and younger adults' difference waves to independent samples, two-tailed permutation tests based on the cluster mass statistic (Bullmore et al., 1999) using a family-wise alpha level

of 0.05. This approach has the advantage of being data-driven and less subject to bias or constraint by *a priori* selections of analysis windows and electrode sites while still maintaining statistical power to detect differences, although power is still lower relative to traditional mean-amplitude based analyses (Groppe et al., 2011). ERPs were first down-sampled to 100 Hz, creating 10 ms time bins. Permutation tests included all time bins between 300 and 1000 ms and all fifteen anterior frontal, frontocentral, and posterior electrodes. Analyses were conducted using the clustGRP function of the Mass Univariate ERP Toolbox (Groppe et al., 2011), which identifies spatiotemporal clusters that differ between conditions by conducting independent samples *t*-tests at each electrode and time bin using the original data and 2500 random between-subject permutations. Neighboring *t*-scores with uncorrected *p*-values of 0.05 or less are grouped into clusters, and all of the *t*-scores within a given cluster are summed together to calculate the cluster mass. Finally, a *p*-value is assigned to each cluster by comparing the cluster masses of the observed data with an estimate of the null distribution based on the maximal cluster masses of the random permutations. Electrodes within 5.44 cm of each other were considered spatial neighbors, and adjacent time points were considered temporal neighbors.

## RESULTS: BEHAVIOR

### Study-Phase Ratings

Table 1A shows the average proportion of abstract-concrete and concrete-concrete trials assigned each rating on the six-point ease-of-definition scale during the conceptual combination encoding blocks. A paired *t*-test revealed that the average rating assigned to concrete-concrete trials (mean = 4.12, se = 0.12) was significantly higher than that assigned to abstract-concrete trials [mean = 3.86, se = 0.14,  $t(23) = 3.81$ ,  $p < 0.001$ , Cohen's  $d = 0.78$ ].

Table 1B shows the average proportion of abstract-concrete and concrete-concrete trials assigned each rating on the six-point frequency-comparison scale during the frequency-comparison encoding blocks. Average ratings for concrete-concrete word pairs (mean = 3.52, se = 0.06) were significantly lower than were ratings assigned to abstract-concrete words pairs [mean = 3.81, se = 0.08,  $t(23) = 3.98$ ,  $p < 0.001$ , Cohen's  $d = 0.81$ ]. Thus, participants reported encountering the concepts denoted by the abstract W1s less frequently (when compared to the W2s) relative to the concepts denoted by the concrete W1s.

### Associative Recognition Performance

To assess associative recognition memory, discrimination sensitivity ( $d'$ ) was calculated separately for each combination of task and W1 concreteness level.  $D'$  measures the accuracy with which participants discriminate between studied and unstudied items, and is obtained by subtracting the *z*-transform of the “false alarm” rate (the proportion of re-paired items incorrectly endorsed as “intact”) from the *z*-transform of the “hit” rate (the proportion of intact items correctly endorsed as “intact”). The results are depicted in Figure 3. To formally assess the effects of the task and concreteness manipulations on

**TABLE 1 |** Mean proportion of abstract-concrete (abs-con) and concrete-concrete (con-con) trials assigned each rating on: (A) the six-point ease-of-definition scale in the conceptual combination task, or (B) the six-point relative frequency judgment in the frequency-comparison task.

(A)	Harder to define			Easier to define		
	"One"	"Two"	"Three"	"Four"	"Five"	"Six"
Abs-Con	0.15 (0.02)	0.17 (0.02)	0.11 (0.02)	0.08 (0.02)	0.20 (0.03)	0.28 (0.04)
Con-Con	0.13 (0.02)	0.15 (0.02)	0.09 (0.02)	0.10 (0.02)	0.21 (0.02)	0.33 (0.04)

(B)	First word more frequent			Second word more frequent		
	"One"	"Two"	"Three"	"Four"	"Five"	"Six"
Abs-Con	0.20 (0.03)	0.10 (0.02)	0.12 (0.02)	0.14 (0.03)	0.14 (0.02)	0.30 (0.04)
Con-Con	0.27 (0.03)	0.11 (0.02)	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)	0.28 (0.03)

associative memory, a 2 (task: compare/combine)  $\times$  2 (W1 type: abstract/concrete) repeated-measures ANOVA was performed on  $d'$  values. A significant main effect of task emerged [ $F(1, 23) = 50.64, p < 0.001, \eta_p^2 = 0.68$ ], indicating higher recognition accuracy in conceptual combination blocks relative to frequency-comparison blocks. In addition, a main effect of W1 type [ $F(1, 23) = 37.14, p < 0.001, \eta_p^2 = 0.62$ ] revealed greater accuracy for concrete-concrete relative to abstract-concrete word pairs. The task  $\times$  W1 type interaction was also significant [ $F(1, 23) = 10.58, p = 0.004, \eta_p^2 = 0.32$ ]. Follow-up paired  $t$ -tests indicated that the beneficial effect of W1 concreteness was significant in the conceptual combination condition [ $t(23) = 6.35, p < 0.001, \text{Cohen's } d = 1.28$ ], but only marginally significant in the frequency-comparison condition [ $t(23) = 2.00, p = 0.06, \text{Cohen's } d = 0.41$ ].

Unsurprisingly, older adults' recognition memory was lower than that of young adults in Lucas et al. (2017), for whom  $d'$  values were 1.18 and 1.66 for abstract-concrete and concrete-concrete word pairs in the frequency-comparison condition, and 2.48 and 3.08 for abstract-concrete and concrete-concrete word pairs in the conceptual combination condition, respectively. Indeed, combining these datasets by re-running the ANOVA with age included as a between-subjects factor produced a significant main effect [ $F(1, 46) = 40.17, p < 0.001, \eta_p^2 = 0.47$ ]. This comparison also revealed a significant 3-way interaction between age, task, and W1 concreteness [ $F(1, 46) = 4.78, p = 0.03, \eta_p^2 = 0.09$ ], suggesting that age differences in the beneficial effects of the conceptual combination over the frequency-comparison task differed across levels of W1 concreteness. Follow-up tests revealed that age differences in associative recognition for concrete-concrete word pairs were equivalent across the two study tasks, as indicated by a non-significant age  $\times$  task interaction [ $F(1, 46) = 0.00, p = 0.98, \eta_p^2 = 0.00$ ]. By contrast, this interaction was significant for abstract-concrete words, [ $F(1, 46) = 5.21, p = 0.03, \eta_p^2 = 0.10$ ], reflecting the fact that age differences were *larger* in the conceptual combination relative to the frequency task.

Finally, a series of independent-samples  $t$ -tests revealed that younger adults outperformed their older counterparts in all four task-W1 combinations [ $ps < 0.001$ ]. However, the effect size of this age difference was larger for abstract-concrete words in the conceptual combination task (Cohen's  $d = 1.76$ ) compared to the

other three conditions (1.46, 1.25, and 1.09 for concrete-concrete word pairs in the conceptual combination task, abstract-concrete pairs in the frequency-comparison task, and concrete-concrete pairs in the frequency-comparison task, respectively).

## Study-Phase Ratings and Associative Memory

An additional analysis assessed the relationship of participants' ease-of-definition ratings during the conceptual combination task to subsequent recognition accuracy. A one-way within-subjects ANOVA indicated that, as with the younger adults in Lucas et al. (2017), the average rating given at study by the older adults was significantly greater for items that were later remembered (mean rating = 4.24,  $se = 0.14$ ) relative to those that were later forgotten [mean rating = 3.11,  $se = 0.15, F(1, 23) = 60.27, p < 0.001, \eta_p^2 = 0.72$ ]. For comparison, the younger adults assigned mean ratings of 4.06 and 2.88 to later-remembered and later-forgotten items, respectively. Re-running the ANOVA with age as a between-subjects factor yielded only a main effect of subsequent memory [ $F(1, 42) = 69.65, p < 0.001, \eta_p^2 = 0.62$ ], indicating that subjective ratings of conceptual combination success were diagnostic of subsequent memory in both age groups.

Another way to examine the relationship between EoD ratings and memory is to use conditional probabilities to compute the likelihood of a word pair being successfully recalled at test contingent upon it having been assigned a certain rating at study. As shown in **Table 1**, on average, older adults assigned one of the highest two ratings (5 and 6) to 50% of the word pairs. Of these word pairs, an average of 83% were successfully recognized during the memory test, versus 62% of the word pairs given an EoD rating of 4 or lower. Younger adults assigned an EoD rating of 5 or 6 to an average of 44%<sup>1</sup> of the word pairs, and the probability of successful memory for these word pairs was 97% on average, versus 90% for word pairs given a rating of 4 or lower. A 2  $\times$  2 ANOVA

<sup>1</sup>The proportion of trials given a 5 or 6 rating did not differ significantly by age [ $t(39.97) = 1.27, p = 0.21, \text{Cohen's } d = 0.37$ ]. Of course, one cannot assume that the "absolute" difficulty level of word pairs assigned a given rating is equal between age groups, particularly since all participants were encouraged to use the entire rating scale.



with factors Age (YA/OA) and pooled EoD rating (High: 5–6/Low: 1–4) on recognition probability yielded significant main effects of both Age [ $F(1, 46) = 29.00$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.39$ ] and EoD [ $F(1, 46) = 46.50$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.50$ ] as well as a significant interaction [ $F(1, 46) = 10.41$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.19$ ]. These analyses suggest that older adults' memory performance dropped more precipitously than younger adults' memory performance as subjective ease-of-definition decreased, which could reflect a steeper drop-off in combination difficulty itself and/or a greater difficulty encoding the more difficult-to-combine pairs into memory.

## RESULTS: ELECTROPHYSIOLOGY

Effects of the task and W1 concreteness manipulations on ERPs were analyzed separately for first words (W1s) and second words (W2s) over the windows of 300–500 ms (N400) and 700–1000 ms (N700) respectively. Each analysis took the form of 2 (task: compare/combine)  $\times$  2 (concreteness: abstract/concrete)  $\times$  3 (electrode cluster: anterior frontal/frontocentral/parietal) ANOVAs. Effects involving electrode cluster are reported only in the context of interactions with other variables.

### Task and Concreteness Effects on First Words (W1s)

We first compared N400 and N700 ERPs to concrete and abstract W1s in the frequency-comparison and conceptual combination tasks. The resulting waveforms are depicted in **Figure 4**, and relevant topographical plots can be found in **Figure 5**. The mean

number of artifact-free trials per participant per condition was 33 (range = 22–36).

### 300–500 ms

From 300 to 500 ms (N400), a significant main effect of W1 concreteness emerged [ $F(1, 23) = 10.15$ ,  $p = 0.004$ ,  $\eta_p^2 = 0.31$ ], indicating more negative amplitudes for concrete than for abstract words. No other main effects or interactions were significant [ $F$ 's  $< 1.50$ ,  $p$ 's  $> 0.23$ ]. These results suggest that concreteness effects for W1s were similar across the two tasks and broadly distributed over the head<sup>2</sup>.

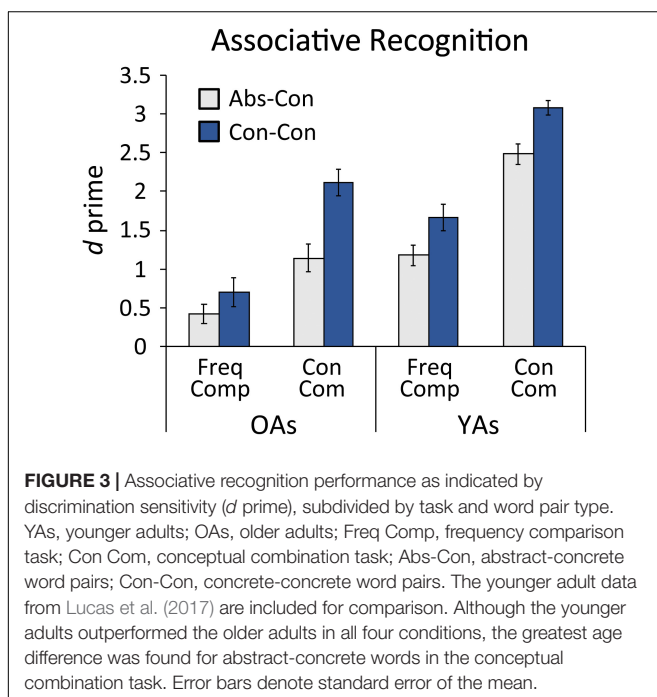
### 700–1000 ms

Analyses of ERPs over the 700–1000 ms (N700) window likewise revealed a significant main effect of W1 concreteness [ $F(1, 23) = 7.81$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.25$ ]. No other main effects or interactions approached significance, although a marginal Cluster  $\times$  Task interaction [ $F(1.28, 29.44) = 3.19$ ,  $p = 0.08$ ,  $\eta_p^2 = 0.11$ ] reflected a trend toward more negative ERPs in the combine relative to the compare condition that was larger at frontal relative to posterior electrodes.

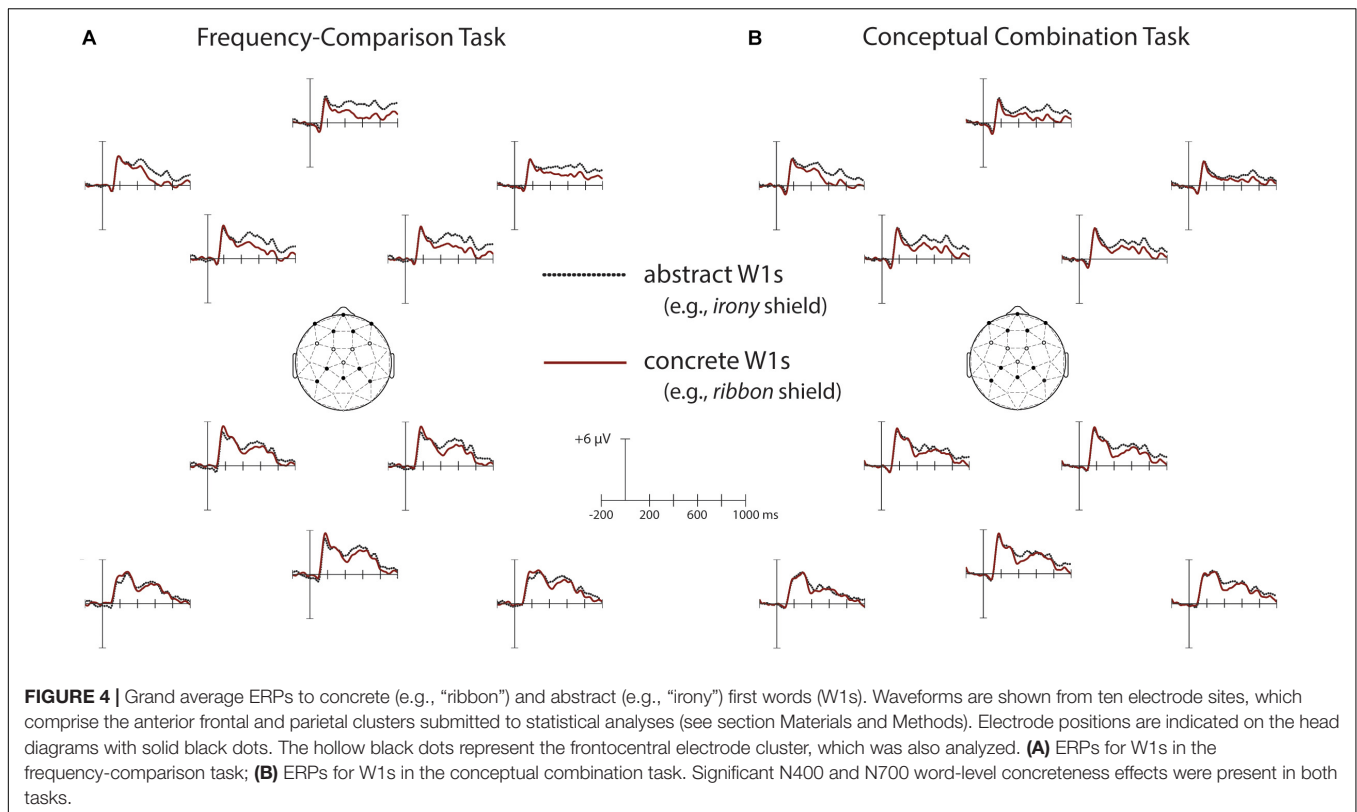
### Summary and Age Comparisons

In summary, older adults showed broadly-distributed concreteness effects for W1s that mirrored those previously found for young adults, in that ERPs were more negative for concrete than for abstract words from 300 to 500 ms regardless of task. In addition, concreteness effects in older adults persisted through the later time window of 700–1000 ms. To directly compare older adults' concreteness effects with those reported for young adults in Lucas et al. (2017), we computed difference waves via a point-by-point subtraction of the ERP response to abstract W1s from the ERP responses to concrete W1s, collapsed across task and electrode cluster. Mean amplitudes of these differences for each time window were entered into a one-way ANOVA with age as a between-subjects factor. Age effects were non-significant for both windows [ $F(1, 46) = 0.22$ ,  $p = 0.64$ ,  $\eta_p^2 = 0.01$  for 300–500 ms;  $F(1, 46) = 1.68$ ,  $p = 0.20$ ,  $\eta_p^2 = 0.04$  for 700–1000 ms]. Thus, despite the appearance of larger N700 concreteness effects in the older relative to the younger adults, this age difference did not reach statistical significance.

In addition to the above mean amplitude-based comparisons, older and younger adults' concreteness difference waves (ERP differences between concrete and abstract words) were entered into cluster mass permutation tests. Separate permutation tests were run on difference waves for the frequency comparison and conceptual combination tasks. All 10-ms time bins between 300 and 1000 ms and all fifteen electrodes were included. No significant differences were identified for either task (all cluster  $p > 0.41$ ). Together, these analyses corroborate



<sup>2</sup>Note that the sustained frontal concreteness effect captured in the N700 window appears to begin somewhat earlier, overlapping with the N400 window. Thus, it is possible that both components are contributing to the concreteness effects measured over this window.



previous evidence that word-level concreteness effects are spared with age.

## Task and Concreteness-Modification Effects on Second Words (W2s)

We next examined concreteness-modification effects present in both tasks by comparing ERPs to W2s that were preceded by concrete W1s with those that were preceded by abstract W1s. The resulting waveforms are depicted in **Figure 6**, and relevant topographical plots can be found in **Figure 7**. The mean number of artifact-free trials per participant was 32 for each condition (range = 19–36).

### 300–500 ms

A significant main effect of W1 concreteness emerged from 300 to 500 ms [ $F(1, 23) = 7.60, p = 0.01, \eta_p^2 = 0.25$ ], indicating more negative amplitudes for abstractly-modified than for concretely-modified words. Despite the appearance of larger concreteness-modification effects in the conceptual combination condition, the Task  $\times$  W1 concreteness interaction was non-significant [ $F(1, 23) = 0.005, p = 0.95, \eta_p^2 < 0.01$ ], nor was the three-way interaction [ $F(1.15, 26.39) = 0.74, p = 0.48, \eta_p^2 = 0.03$ ]. No other main effects or interactions approached significance, although a marginal W1 concreteness  $\times$  cluster interaction [ $F(1.32, 30.36) = 2.95, p = 0.08, \eta_p^2 = 0.11$ ], revealed a trend toward greater effects of W1 concreteness at anterior relative to posterior electrode clusters.

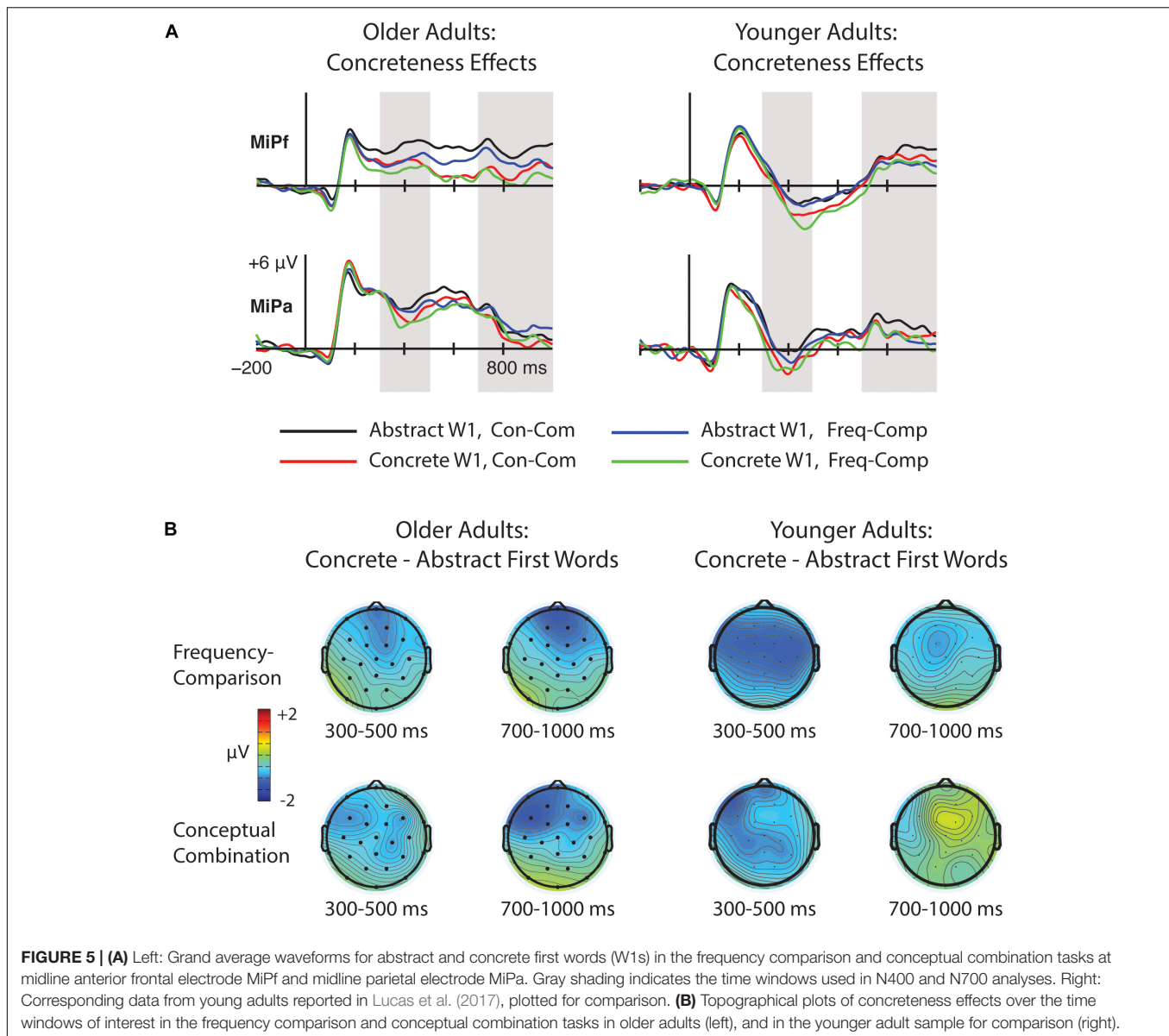
### 700–1000 ms

Analyses of ERPs to W2s from 700 to 1000 ms revealed no significant main effects or interactions for any variables in this time window ( $F$ 's  $< 1.65, p$ 's  $> 0.20$ ).

## Summary and Age Comparisons

In summary, compositional concreteness effects in older adults occurred in the N400 window, where they took the form of more negative ERPs to abstract versus concrete words. This pattern is similar to the N400 compositional concreteness effects observed in young adults in Lucas et al. (2017), and, indeed, a comparison of the mean difference in N400 amplitudes between concretely-modified and abstractly-modified W2s between age groups was non-significant [ $F(1, 46) = 0.12, p = 0.73$ ]. Note that no main effects or interactions involving encoding task were significant in either age group, meaning that N400 compositional concreteness effects were present regardless of whether or not the participants were attempting conceptual combination. As such, these effects appear to be largely stimulus driven, and may result from the tendency of concrete W1s to activate a wider range of semantic features and thereby induce “happenstance” feature overlap with W2s to a greater extent than abstract W1s (see also Experiment 1 Discussion of Lucas et al., 2017).

Importantly, the older adults here showed no evidence of compositional concreteness effects in the N700 window. By contrast, younger adults in our previous study showed N700 compositional concreteness effects that were selective to the conceptual combination task, in which ERPs over



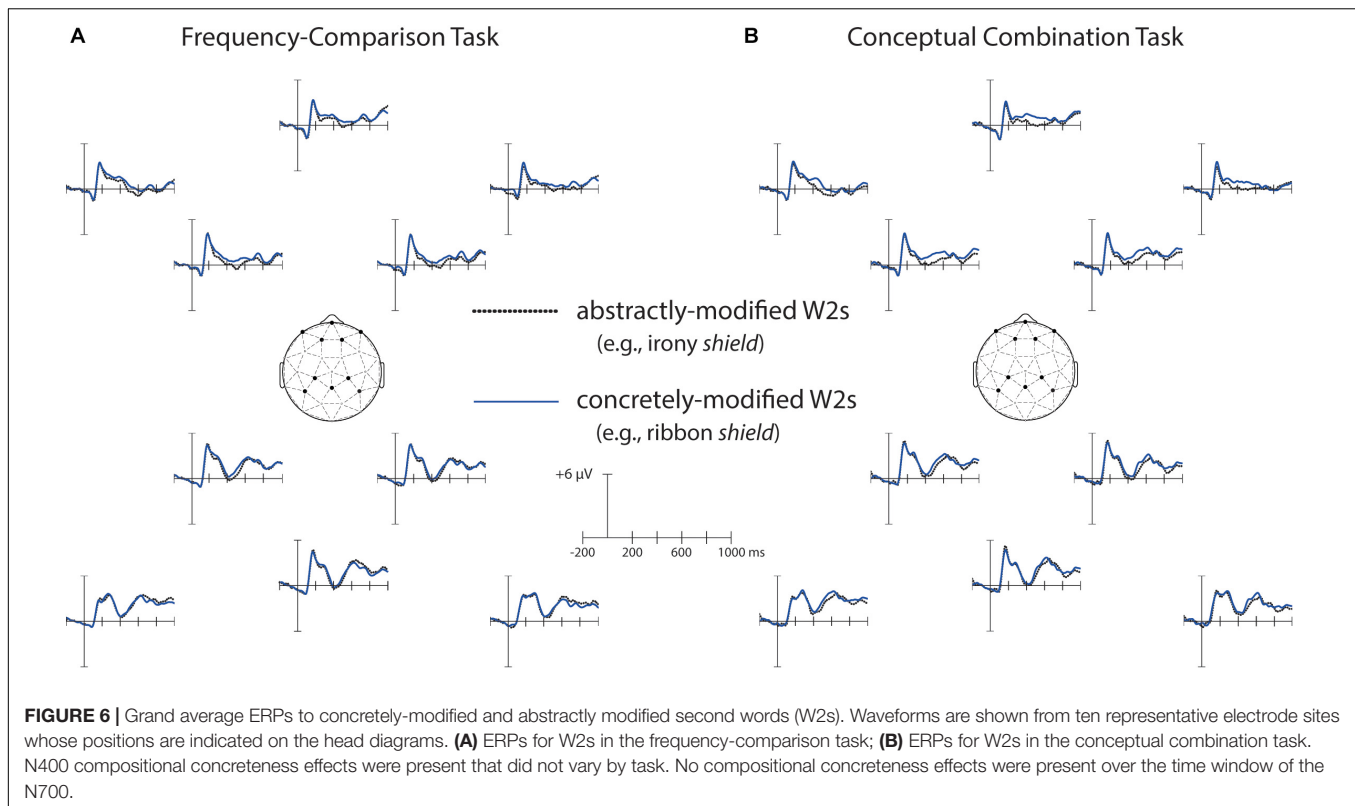
the anterior frontal electrode cluster were more negative for concretely-modified relative to abstractly-modified W2s. Thus, we employed a one-way ANOVA to examine age differences in N700 compositional concreteness effects during this task over the anterior frontal cluster. The effect of Age was significant [ $F(1, 46) = 10.02, p = 0.003, \eta_p^2 = 0.10$ ].

As with the word-level concreteness effects, we also examined age differences in compositional concreteness effects using cluster-based permutation analyses over all electrodes and all 10 ms time bins from 300 to 1000 ms separately for the frequency comparison and conceptual combination task. No difference was identified in the frequency comparison task ( $p > 0.51$ ). However, the permutation test for the conceptual combination task identified a significant difference between age groups ( $p = 0.03$ ) in the form of a negative frontally-focused cluster, which began around 850 ms and

continued to the end of the epoch (see **Figure 8**). Note that cluster-based permutation analyses control the type 1 error rate only with respect to whether or not the overall multivariate datasets differ (i.e., the null hypothesis is that the compositional concreteness effects of the older and younger adults are “interchangeable”), rather than for each individual time point/electrode pairing. Thus, these results should be interpreted as an approximate rather than exact spatiotemporal locus of the age difference (Sassenhagen and Draschkow, 2019). Nonetheless, the overall timing and topography of this cluster is consistent with the N700 effect found in the time window-based age comparison.

### Single-Trial Analyses

Like the young adults in Lucas et al. (2017), older adults here showed improvements in associative recognition memory in



the conceptual combination blocks relative to the frequency-comparison blocks. However, when comparing the recognition performance of the two age groups, it is apparent that conceptual combination did not reduce age-related memory deficits relative to the frequency-comparison task. Rather, age differences were numerically larger for concrete-concrete word pairs in the combination versus the frequency task, and were significantly larger for abstract-concrete word pairs, suggesting a reduction in the effectiveness of conceptual combination as a memorization strategy. Likewise, as in Huang et al. (2012), older adults did not show compositional concreteness effects on N700 ERPs, suggesting a reduced tendency or ability to use features of the W1 to modify the processing applied to the W2.

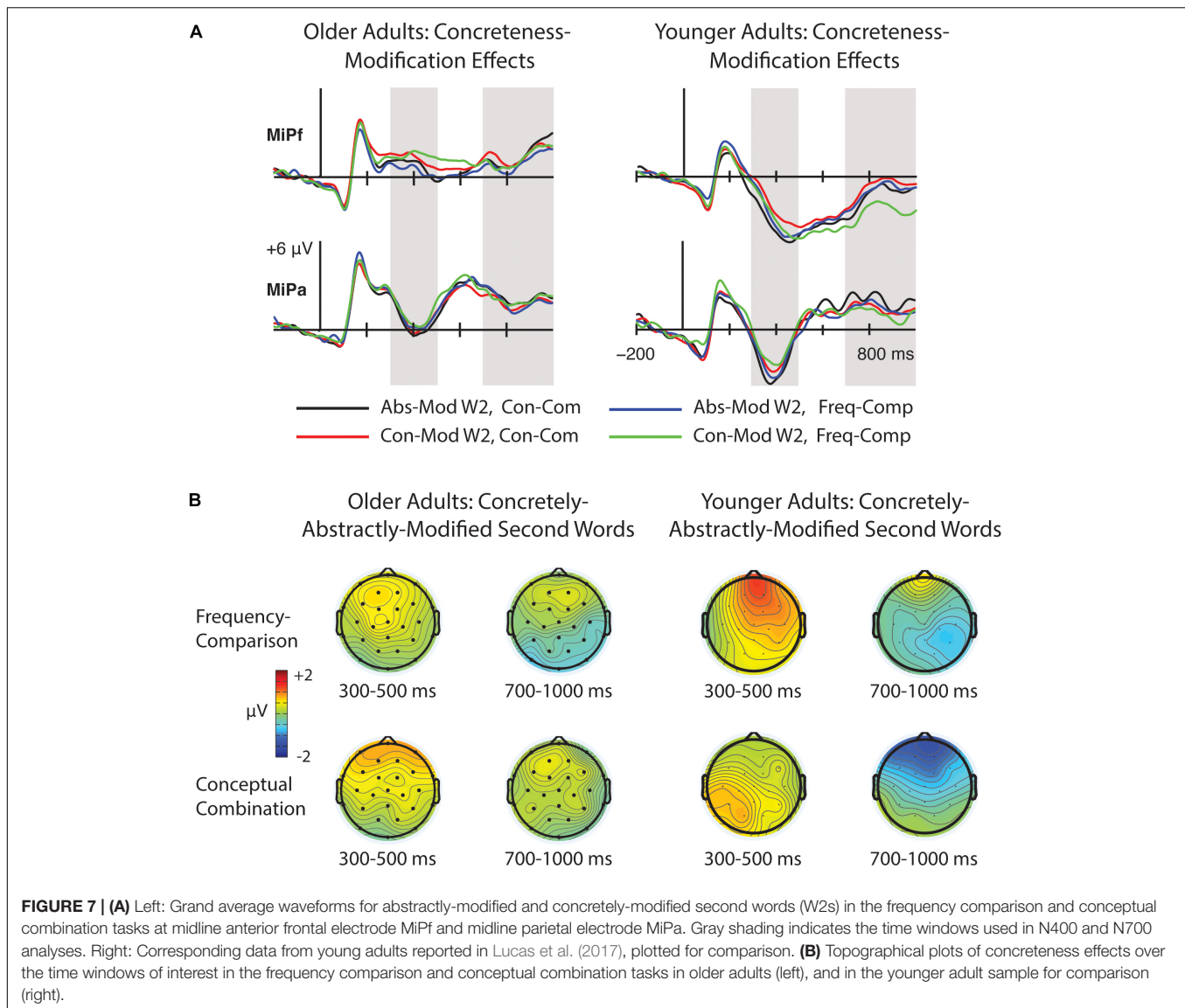
A clear next question concerns the nature of the relationship of the neurocognitive processing reflected in N700 potentials to conceptual combination and associative memory formation. One possibility is that N700 amplitudes track the relative success of conceptual combination and memory formation from trial to trial, such that amplitudes are more negative for trials on which conceptual combination is easiest. This pattern, together with the findings that word pairs with more imageable W1s were both rated as easier to combine and were better remembered, would seem intuitive in light of imagery-based accounts of N700 concreteness effects.

That said, other recent findings complicate the notion that the ease or vividness of the evoked imagery *per se* is the primary determinant of N700 concreteness effects. The “canonical” effect (more negative N700 amplitudes to concrete words) has typically been found in studies that ask participants to judge whether or

not a word is easily imageable (West and Holcomb, 2000; Gullick et al., 2013) or in tasks that require semantic processing without explicit reference to imagery (Holcomb et al., 1999; Huang et al., 2010). However, Welcome et al. (2011) obtained a reversal of this pattern when participants were asked to create an image for each and every word before moving on to the next, regardless of difficulty. In this context, N700 amplitudes were greater for abstract word pairs, suggesting a sensitivity to the amount of effort or cognitive resources put toward image generation on a given trial. In addition, Barber et al. (2013) reported evidence that N700 concreteness effects may not exclusively index processes related to visual imagery, but may reflect a more general top-down process of retrieving and integrating intrinsic features (visual and otherwise) that tends to be engaged to a greater extent during the processing of feature-rich concrete words. Together, these findings raise the possibility that more difficult to combine word pairs may evoke *larger* N700 amplitudes, at least in young adults, insofar as they trigger additional cognitive control processes to overcome the difficulty.

To gain traction on this issue, we used linear mixed-effects modeling to model W2 N700 amplitudes at the individual trial level based on participants’ ease-of-definition (EoD) ratings during the conceptual combination task. Amplitudes for each trial reflect the mean voltage from 700–1000 ms, averaged over all electrodes in the anterior frontal cluster. Analyses were carried out using the lme4 software package version 1.1-19 (Bates et al., 2015) and the afex package version 0.23-0 (Singmann et al., 2015). We first constructed a statistical model that included: (1) age and W1 concreteness as categorical



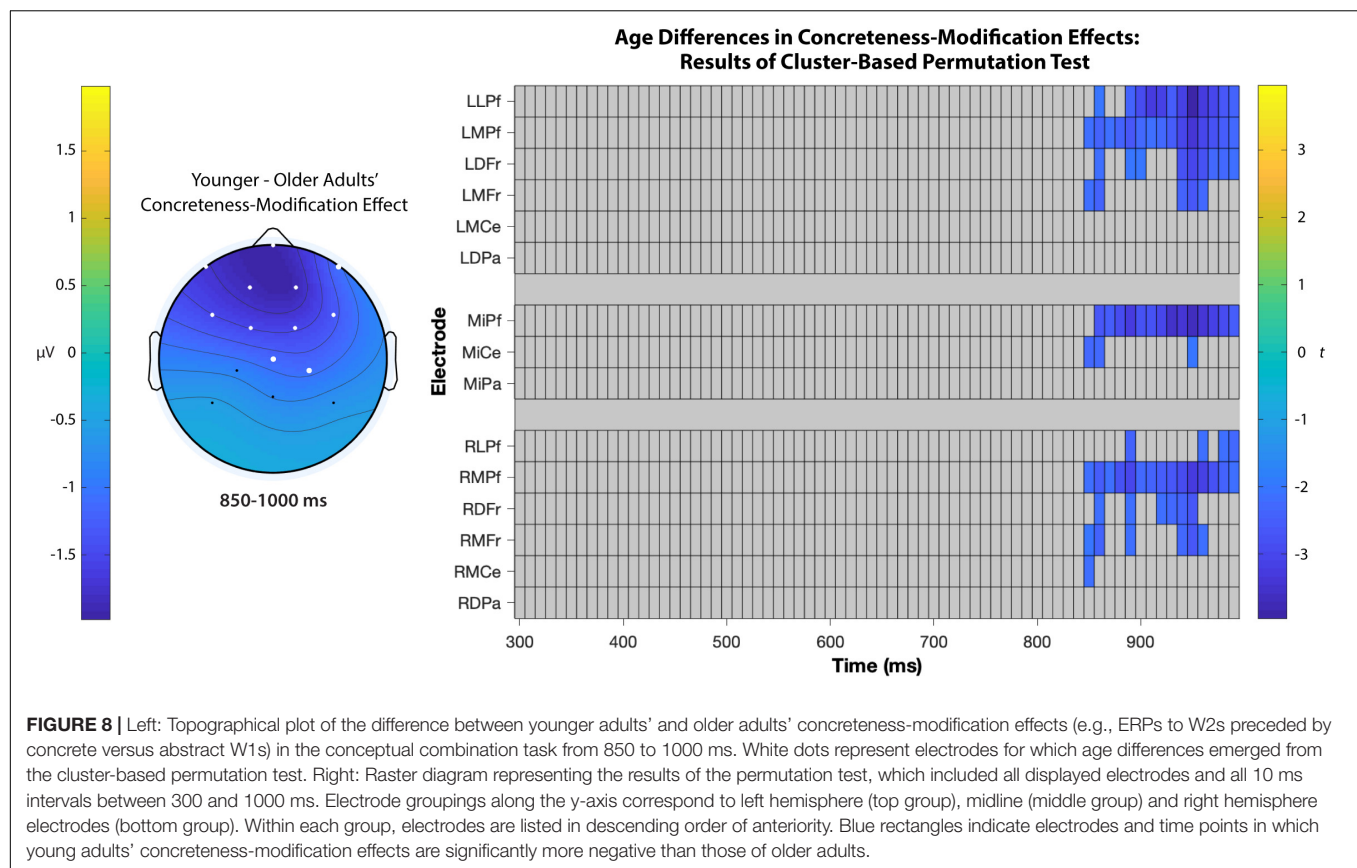


fixed-effect predictors, (2) EoD rating as a continuous fixed-effect predictor, with higher ratings indicating lower perceived difficulty, (3) interactions among fixed effects, and (4) a random intercept for participants. Random slopes were removed to facilitate model convergence. Prior to analyses, the continuous predictor was mean-centered within each age group and the nominal predictors were sum coded. Effect significance was evaluated using Satterthwaite's approximation as implemented in the lmerTest package (Kuznetsova et al., 2017).

The results are depicted in **Table 2**. As shown, W1 concreteness was a significant predictor of N700 amplitudes ( $\beta = -0.29$ ,  $t = 1.99$ ,  $p = 0.047$ ), with concrete items eliciting larger N700 amplitudes than abstract items. Consistent with the trial-aggregated analyses, this effect was qualified by a significant age  $\times$  W1 concreteness interaction ( $\beta = 0.35$ ,  $t = 2.40$ ,  $p = 0.02$ ). The main effect of EoD rating was also significant ( $\beta = 0.23$ ,  $t = 2.70$ ,  $p = 0.007$ ) and also interacted

with age ( $\beta = -0.18$ ,  $t = 2.08$ ,  $p = 0.04$ ). Note that the positive parameter estimate for the main effect indicates that N700 amplitudes were smaller (less negative) for trials that received higher ease-of-definition ratings. The EoD rating  $\times$  W1 concreteness interaction was not significant ( $\beta = -0.09$ ,  $t = -1.11$ ,  $p = 0.27$ ), nor was the three-way interaction ( $\beta = 0.01$ ,  $t = 0.15$ ,  $p = 0.88$ ).

To further examine the effects involving EoD rating, we tested separate models for each age group that included W1 concreteness and EoD rating as fixed effects and a random intercept for participant. **Figure 9A** depicts the parameter estimates and standard errors of the resulting models. As shown, both main effects were significant for the younger adults ( $\beta = -0.64$ ,  $t = 2.85$ ,  $p = 0.005$  for W1 concreteness;  $\beta = 0.41$ ,  $t = 2.99$ ,  $p = 0.003$  for EoD rating), whereas the interaction was non-significant ( $\beta = -0.11$ ,  $t = 0.77$ ,  $p = 0.44$ ). These results indicate that younger adults' W2 N700 amplitudes were: (1) more negative



for concrete-concrete than abstract-concrete trials, regardless of difficulty rating, and (2) more negative for trials that received more difficult ratings (e.g., lower EoD ratings), regardless of W1 concreteness. The analogous model in the older adults revealed no significant main effects, nor a significant interaction (all  $p$ 's > 0.40). For visualization purposes only, **Figure 9B** depicts W2 waveforms from the conceptual combination task in either age group subdivided by subjective rating (based on a split in which ratings 1:4 were grouped as "hard" and 5:6 as "easy," see figure caption).

**TABLE 2 |** Results of linear mixed-effects models predicting W2 N700 amplitudes during the conceptual combination task.

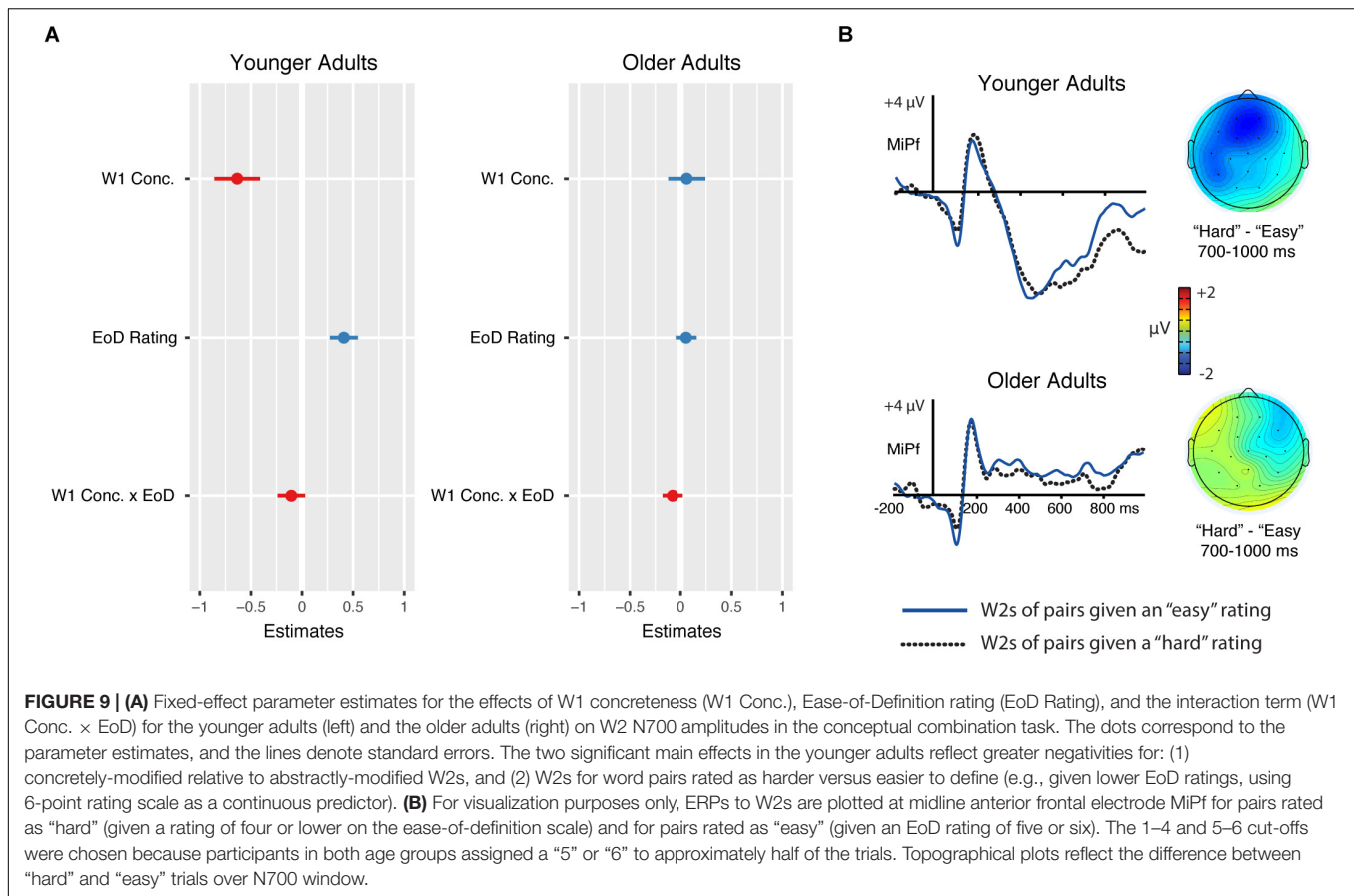
Parameter	Estimate	SE	t-value	P-value
Intercept	-0.36	0.43	0.83	0.41
Age	1.16	0.43	2.70	0.01**
W1 concreteness	-0.29	0.15	1.99	0.047*
EoD rating	0.23	0.09	2.71	0.007**
Age × W1 concreteness	0.35	0.15	2.40	0.02*
Age × EoD rating	-0.18	0.09	2.08	0.04*
W1 concreteness × EoD rating	-0.09	0.08	1.11	0.27
Age × W1 concreteness × EoD	0.01	0.08	0.15	0.88

Parameter estimates are for fixed effects involving age, W1 concreteness, and Ease-of-Definition (EoD) rating. T- and p-values obtained using Satterthwaite's approximation. SE, standard error; \* $p < 0.05$ ; \*\* $p < 0.01$ .

## DISCUSSION

Adult aging brings about changes to multiple domains of cognition, including aspects of memory and language processing. In the domain of memory, there is evidence that relational or associative memory is more vulnerable to the negative effects of age than is memory for single items (e.g., Old and Naveh-Benjamin, 2008). An analogous distinction is present in the literature on language and aging: while older and younger adults generally show similar lexical-level processing, aging has been linked to changes in compositional language processing (Wlotko et al., 2010). Thus far, however, these two facets of neurocognitive aging have been studied largely independently of one another. Indeed, despite decades of research on the importance of "deep" or meaning-based processing for memory, little attention has been paid to potential ways in which reduced online semantic integration may affect associative memory performance in older adults.

The present study addresses a question that falls squarely at the memory-language interface: the extent to which older and younger adults differ in implementing an associative memorization strategy that relies on conceptual combination. Previous work (Ahmad et al., 2014) has found that age-related associative memory deficits are smaller for noun pairs that corresponded to pre-experimentally familiar compound words (e.g., "store keeper") relative to those that do not, presumably because familiar pairs are represented in memory in a "unitized"



**FIGURE 9 | (A)** Fixed-effect parameter estimates for the effects of W1 concreteness (W1 Conc.), Ease-of-Definition rating (EoD Rating), and the interaction term (W1 Conc.  $\times$  EoD) for the younger adults (left) and the older adults (right) on W2 N700 amplitudes in the conceptual combination task. The dots correspond to the parameter estimates, and the lines denote standard errors. The two significant main effects in the younger adults reflect greater negativities for: (1) concretely-modified relative to abstractly-modified W2s, and (2) W2s for word pairs rated as harder versus easier to define (e.g., given lower EoD ratings, using 6-point rating scale as a continuous predictor). **(B)** For visualization purposes only, ERPs to W2s are plotted at midline anterior frontal electrode MiPf for pairs rated as "hard" (given a rating of four or lower on the ease-of-definition scale) and for pairs rated as "easy" (given an EoD rating of five or six). The 1–4 and 5–6 cut-offs were chosen because participants in both age groups assigned a "5" or "6" to approximately half of the trials. Topographical plots reflect the difference between "hard" and "easy" trials over N700 window.

or item-like manner. However, initial studies examining older adults' ability to engage in *ad hoc* unitization of non-pre-experimentally related stimuli have yielded mixed results (c.f. Bastin et al., 2013; Kamp et al., 2018). For example, a recent examination of older adults' ability to use experimenter-provided definitions to aid in the memorization of novel word pairs (Kamp et al., 2018) found that age differences in performance on subsequent associative memory tests were exaggerated rather than reduced relative to a control condition<sup>3</sup>. There is thus a need to better understand how aging intersects with the neurocognitive demands of flexible language processing in order to determine whether and when this and other language-mediated strategies are appropriate.

In the present study, we adapted a paradigm previously used to study the processing of adjective-noun modification relationships (Huang et al., 2010, 2012) to examine noun-noun combination for unrelated word pairs. A key feature of this design is the systematic manipulation of the concreteness of the first noun in each pair, which has been shown in young adults to induce compositional concreteness effects on a slow frontal potential termed the N700 (Lucas et al., 2017). Notably, in the present study, N700 compositional concreteness effects were absent in

older adults, despite the fact that word-level concreteness effects were age-invariant. Moreover, while older adults' associative memory was superior following the conceptual combination task relative to a non-combinatory encoding task, the magnitude of the age-related deficit was either similar between the two tasks (for concrete-concrete pairs) or greater in the conceptual combination task (for abstract-concrete pairs). As such, these data join others to suggest that there may be limits to older adults' ability to use conceptual combination as an encoding strategy.

As previously discussed, N700 amplitudes have been related in previous work to the retrieval and integration of semantic features – including, but perhaps not limited to, features involved in visual imagery (West and Holcomb, 2000; Welcome et al., 2011; Barber et al., 2013; Gullick et al., 2013). It therefore seems plausible that older adults may either be less able or less inclined to use compositional imagery or other semantic integration processes to aid in conceptual combination. Indeed, our single-trial analyses provide evidence that W2 N700 potentials are sensitive to variability in perceived difficulty of conceptual combination. In young adults, W1 concreteness and perceived difficulty accounted for independent variance in W2 N700 amplitudes: amplitudes were larger (more negative) for concrete-concrete word pairs regardless of rated difficulty and for word pairs that receive greater difficulty ratings regardless of W1 concreteness. This pattern of results coheres with proposals that attribute N700 concreteness effects to the greater number of

<sup>3</sup>Age differences were also found in ERPs recorded during encoding in the study by Kamp et al. (2018). However, comparisons with the present ERP effects are complicated by significant methodological differences between studies.

features intrinsic to concrete relative to abstract concepts (e.g., Wiemer-Hastings and Xu, 2005), which places greater demands on certain semantic integration processes. By this interpretation, the augmented N700 amplitudes to the W2s for more difficult word pairs (in the young adults) could reflect the engagement of additional processing resources to aid conceptual combination. Likewise, the absence of W2 N700 modulation by either variable in older adults is consistent with an overall impoverishment of semantic integration at the compositional level.

An important direction for future research will be to more directly examine how variability in N700 amplitudes relates to subsequent associative memory at the trial level. While both age groups were significantly more likely to remember word pairs that were given higher versus lower EoD ratings, recognition accuracy in the younger adults was quite high across the board. Accordingly, it is possible that the N700 enhancement in younger adults for more difficult pairs reflected a process that helped to “rescue” memory formation on these trials. This possibility could be tested in future work that compares encoding-related ERPs for later-remembered and later-forgotten trials in young adults separately for trials rated as easy versus difficult to combine, with the prediction that N700 potentials would selectively predict memory for difficult trials.

Interestingly, older adults’ subjective ratings and associative memory performance suggest disproportionate difficulty applying the conceptual combination strategy to the abstract-concrete word pairs. This is an unusual finding: although it is well-established that concrete words enjoy facilitated processing speed and memory relative to abstract words, older adults have generally shown concreteness effects that are either smaller than or equal to those of younger adults (e.g., Whitbourne and Slevin, 1978; Rissenberg and Glanzer, 1987; Dirks and Craik, 1992; Peters and Daum, 2008; Roxbury et al., 2015). As such, further research will be necessary to identify the neurocognitive locus of older adults’ increased difficulty creating combinations using abstract modifiers.

Cognitive theories of conceptual combination – such as the Competition Among Relations in Nominals (Gagné and Shoben, 1997) and Relational Interpretation Competitive Evaluation theories (Gagné and Spalding, 2013) – point to differences in the “combinatorial histories” of modifiers as a factor that may influence their amenability to flexible conceptual combination. According to these models, when a modifier is presented as part of a compound, multiple possible relations are activated based on that word’s history as a modifier, and competition among these relations must be resolved before a compound can be perceived as meaningful (see Schmidtke et al., 2016, for supporting evidence from two lexical decision megastudies). It seems plausible that the process of resolving relational competition may pose more difficulty for older adults, perhaps due to a combination of greater word knowledge (e.g., more competing meanings activated) and the need to engage in control processes to choose among competing meanings. As such, avenues for future research may include examinations of: (1) whether abstract and concrete modifiers differ in the number and/or availability of “suggested” possible relations, and (2) whether older adults are

disproportionately hindered by increases in competition among possible relations.

An additional consideration pertains to the *type* of relations suggested by abstract and concrete modifiers. Recent neuroimaging evidence (Boylan et al., 2017) suggests that partially distinct brain networks are involved in feature-based or attributive interpretations of the conceptual combinations (e.g., “monster truck,” which denotes a truck with monster-like features) relative to relationship-based interpretations (e.g., “soup spoon,” which denotes a modifier-head relationship that is not based on shared features). It is possible that abstract modifiers have a greater tendency to evoke relationship-based interpretations relative to more feature-rich concrete modifiers, and that it is this distinction that accounts for the age differences observed here. Future research could also examine this possibility, although at least one study (Taler et al., 2005) provides behavioral evidence against the idea that aging selectively impacts relationship-based interpretations of novel compounds.

Another goal for future research is to test the generality of these findings by juxtaposing conceptual combination with a wider range of alternative encoding strategies. We chose the frequency rating task as our starting point because it requires participants to engage in “deep” or meaning-based processing of both items in each pair without encouraging these meanings to be combined. The inclusion in future studies of strategies that vary in their neurocognitive overlap with conceptual combination will provide more precise insight into the underpinnings of the observed age differences, both in encoding-related ERPs and associative memory performance<sup>4</sup>. For example, it may be informative to compare the conceptual combination strategy with interactive imagery (e.g., imagining the two items interacting in some way), which arguably requires the use of compositional imagery without requiring the generation and selection of relational interpretations.

Multiple existing theories of age-related memory deficits point to a reduction in the use of strategies as a contributing factor to worsening associative memory (e.g., Glisky et al., 2001; Luo et al., 2007; Naveh-Benjamin et al., 2007; Craik and Rose, 2012; Hertzog et al., 2012). In particular, the Associative Deficit Hypothesis (Naveh-Benjamin, 2000; Naveh-Benjamin et al., 2007) implicates a reduced tendency of older adults to initiate intentional strategies (such as sentence-generation or interactive imagery) that aid in the formation of new associative memories. An area of debate within this literature concerns the extent to which these deficits are specific to the spontaneous initiation of strategies, rather than the ability to implement them when given explicit instructions to do so. In the present study, older adults were not only provided with strategy instructions, but were also offered a brief training and

<sup>4</sup>Another option is to include a “no strategy” condition in which participants are given no instructions other than to attempt to remember the word pairs. However, previous research suggests that the nature and extent of spontaneous strategy use can differ considerably between younger and older adults under these circumstances (e.g., Dunlosky and Hertzog, 2001; Naveh-Benjamin et al., 2007; Tournier and Postal, 2011; Craik and Rose, 2012). Thus, this type of design may be more suitable to answer questions about age effects on self-initiated strategy use.



practice in both the conceptual combination and frequency comparison strategies. As such, our results provide evidence that: (1) strategy-based deficits with age are not limited to spontaneous initiation, and (2) certain combinations of strategies and stimulus properties (e.g., conceptual combination for abstract-concrete word pairs) may present particular difficulties for older adults. More generally, this work underscores the idea that age effects on compositional processing in language and associative memory are not independent of one another and highlights a need for further research on interactivity between these cognitive domains across the lifespan.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board at the University of

Illinois Urbana-Champaign. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

HL, RG, RH, and KF contributed to the conception and design of the study, and wrote the manuscript. HL and RG collected the data. HL performed the statistical analysis. All authors contributed to the manuscript revision, and read and approved the submitted version.

## FUNDING

This work was supported by a National Institute on Aging Grant R01-AG026308, a James S. McDonnell Foundation Scholar Award to KF, and a Beckman Postdoctoral Fellowship to HL.

## ACKNOWLEDGMENTS

We thank Kathy Abusager for assistance with data collection.

## REFERENCES

- Addis, D. R., Giovanello, K. S., Vu, M.-A., and Schacter, D. L. (2014). Age-related changes in prefrontal and hippocampal contributions to relational encoding. *Neuroimage* 84, 19–26. doi: 10.1016/j.neuroimage.2013.08.033
- Ahmad, F. N., Fernandes, M., and Hockley, W. E. (2014). Improving associative memory in older adults with unitization. *Aging Neuropsychol. Cogn.* 22, 452–472. doi: 10.1080/13825585.2014.980216
- Barber, H. A., Otten, L. J., Koutsta, S.-T., and Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain Lang.* 125, 47–53. doi: 10.1016/j.bandl.2013.01.005
- Bastin, C., Diana, R. A., Simon, J., Collette, F., Yonelinas, A. P., and Salmon, E. (2013). Associative memory in aging: the effect of unitization on source memory. *Psychol. Aging* 28, 275–283. doi: 10.1037/a0031566
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Boylan, C., Trueswell, J. C., and Thompson-Schill, S. L. (2017). Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain Lang.* 169, 8–21. doi: 10.1016/j.bandl.2017.01.008
- Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., and Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 18, 32–42. doi: 10.1109/42.750253
- Coltheart, M. (2007). The MRC psycholinguistic database. *Q. J. Exp. Psychol. A* 33, 497–505. doi: 10.1080/14640748108400805
- Craik, F. I. M., and Rose, N. S. (2012). Memory encoding and aging: a neurocognitive perspective. *Neurosci. Biobehav. Rev.* 36, 1729–1739. doi: 10.1016/j.neubiorev.2011.11.007
- Dennis, N. A., Hayes, S. M., Prince, S. E., Madden, D. J., Huettel, S. A., and Cabeza, R. (2008). Effects of aging on the neural correlates of successful item and source memory encoding. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 791–808. doi: 10.1037/0278-7393.34.4.791
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* 16, 693–700. doi: 10.1016/j.conb.2006.10.012
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Diana, R. A., Yonelinas, A. P., and Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *J. Cogn. Neurosci.* 22, 1808–1818. doi: 10.1162/jocn.2009.21335
- Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., et al. (2009). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiol. Aging* 30, 432–440. doi: 10.1016/j.neurobiolaging.2007.07.022
- Dirkx, E., and Craik, F. I. (1992). Age-related differences in memory as a function of imagery processing. *Psychol. Aging* 7, 352–358. doi: 10.1037/0882-7974.7.3.352
- Dunlosky, J., and Hertzog, C. (2001). Measuring strategy production during associative learning: the relative utility of concurrent versus retrospective reports. *Mem. Cogn.* 29, 247–253. doi: 10.3758/bf03194918
- Federmeier, K. D., and Laszlo, S. (2009). Time for meaning: electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychol. Learn. Motiv.* 51, 1–44.
- Gagné, C. L., and Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 71–87. doi: 10.1037/0278-7393.23.1.71
- Gagné, C. L., and Spalding, T. L. (2013). *Conceptual Composition: The Role of Relational Competition in the Comprehension of Modifier-Noun Phrases and Noun-Noun Compounds in Psychology of Learning and Motivation*. Cambridge, MA: Academic Press.
- Glisky, E. L., Rubin, S. R., and Davidson, P. S. (2001). Source memory in older adults: an encoding or retrieval problem? *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1131–1146. doi: 10.1037/0278-7393.27.5.1131
- Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x
- Gullick, M. M., Mitra, P., and Coch, D. (2013). Imagining the truth and the moon: an electrophysiological study of abstract and concrete word processing. *Psychophysiology* 50, 431–440. doi: 10.1111/psyp.12033
- Haskins, A. L., Yonelinas, A. P., Quamme, J. R., and Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron* 59, 554–560. doi: 10.1016/j.neuron.2008.07.035
- Hertzog, C., Price, J., and Dunlosky, J. (2012). Age differences in the effects of experimenter-instructed versus self-generated strategy use. *Exp. Aging Res.* 38, 42–62. doi: 10.1080/0361073X.2012.637005
- Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an

- electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742. doi: 10.1037//0278-7393.25.3.721
- Huang, H.-W., Lee, C.-L., and Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *Neuroimage* 49, 1116–1123. doi: 10.1016/j.neuroimage.2009.07.031
- Huang, H.-W., Meyer, A. M., and Federmeier, K. D. (2012). A “concrete view” of aging: event related potentials reveal age-related changes in basic integrative processes in language. *Neuropsychologia* 50, 26–35. doi: 10.1016/j.neuropsychologia.2011.10.018
- Kamp, S.-M., Bader, R., and Mecklinger, A. (2018). Unitization of word pairs in young and older adults: encoding mechanisms and retrieval outcomes. *Psychol. Aging* 33, 497–511. doi: 10.1037/pag0000256
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lucas, H. D., Hubbard, R. J., and Federmeier, K. D. (2017). Flexible conceptual combination: electrophysiological correlates and consequences for associative memory. *Psychophysiology* 54, 833–847. doi: 10.1111/psyp.12840
- Luo, L., Hendriks, T., and Craik, F. I. M. (2007). Age differences in recollection: three patterns of enhanced encoding. *Psychol. Aging* 22, 269–280. doi: 10.1037/0882-7974.22.2.269
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1170–1187. doi: 10.1037//0278-7393.26.5.1170
- Naveh-Benjamin, M., Brav, T. K., and Levy, O. (2007). The associative memory deficit of older adults: the role of strategy utilization. *Psychol. Aging* 22, 202–208. doi: 10.1037/0882-7974.22.1.202
- 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
- Old, S. R., and Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychol. Aging* 23, 104–118. doi: 10.1037/0882-7974.23.1.104
- Palmer, J. A., Kreutz-Delgado, K., and Makeig, S. (2012). *An Adaptive Mixture of Independent Component Analyzers with Shared Components*. Available at: [https://scn.ucsd.edu/~jason/amica\\_a.pdf](https://scn.ucsd.edu/~jason/amica_a.pdf) (accessed May 20, 2019).
- Payne, B. R., and Federmeier, K. D. (2018). Contextual constraints on lexico-semantic processing in aging: evidence from single-word event-related brain potentials. *Brain Res.* 1687, 117–128. doi: 10.1016/j.brainres.2018.02.021
- Peters, J., and Daum, I. (2008). Differential effects of normal aging on recollection of concrete and abstract words. *Neuropsychology* 22, 255–261. doi: 10.1037/0894-4105.22.2.255
- Preston, A. R., and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* 23, R764–R773. doi: 10.1016/j.cub.2013.05.041
- Quamme, J. R., Yonelinas, A. P., and Norman, K. A. (2007). Effect of unitization on associative recognition in amnesia. *Hippocampus* 17, 192–200. doi: 10.1002/hipo.20257
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., et al. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cereb. Cortex* 15, 1676–1689. doi: 10.1093/cercor/bhi044
- Rissenberg, M., and Glanzer, M. (1987). Free recall and word finding ability in normal aging and senile dementia of the Alzheimer's type: the effect of item concreteness. *J. of Gerontol.* 42, 318–322. doi: 10.1093/geronj/42.3.318
- Roxbury, T., McMahon, K., Coulthard, A., and Copland, D. A. (2015). An fMRI study of concreteness effects during spoken word recognition in aging: preservation or attenuation? *Front. Aging Neurosci.* 7:240. doi: 10.3389/fnagi.2015.00240
- Sassenhagen, J., and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* 56:e13335. doi: 10.1111/psyp.13335
- Schmidtke, D., Kuperman, V., Gagné, C. L., and Spalding, T. L. (2016). Competition between conceptual relations affects compound recognition: the role of entropy. *Psychon. Bull. Rev.* 23, 556–570. doi: 10.3758/s13423-015-0926-0
- Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2015). *Afex: Analysis of Factorial Experiments. R Package Version 0.13–145*.
- Taler, V., Chertkow, H., and Saumier, D. (2005). The effect of Alzheimer's disease and aging on conceptual combination. *Brain Cogn.* 59, 306–309. doi: 10.1016/j.bandc.2004.08.049
- Tournier, I., and Postal, V. (2011). Strategy selection and aging: impact of item concreteness in paired-associate task. *Aging Neuropsychol. Cogn.* 18, 195–213. doi: 10.1080/13825585.2010.525623
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277, 376–380. doi: 10.1126/science.277.5324.376
- Welcome, S. E., Paivio, A., McRae, K., and Joanisse, M. F. (2011). An electrophysiological study of task demands on concreteness effects: evidence for dual coding theory. *Exp. Brain Res.* 212, 347–358. doi: 10.1007/s00221-011-2734-8
- West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037. doi: 10.1162/08989290051137558
- Whitbourne, S. K., and Slevin, A. E. (1978). Imagery and sentence retention in elderly and young adults. *J. Genet. Psychol.* 133, 287–298. doi: 10.1080/00221325.1978.10533386
- Wiemer-Hastings, K., and Xu, X. (2005). Content differences for abstract and concrete concepts. *Cogn. Sci.* 29, 719–736. doi: 10.1207/s15516709cog0000\_33
- Wlotko, E. W., Lee, C.-L., and Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Lang. Linguist. Compass* 4, 623–638. doi: 10.1111/j.1749-818x.2010.00224.x

**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 © 2019 Lucas, Gupta, Hubbard and Federmeier. 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.



# Naming and Knowing Revisited: Eyetracking Correlates of Anomia in Progressive Aphasia

Molly B. Ungrady<sup>1,2,3\*</sup>, Maurice Flurie<sup>2,3</sup>, Bonnie M. Zuckerman<sup>2,3</sup>, Daniel Mirman<sup>4</sup> and Jamie Reilly<sup>2,3\*</sup>

<sup>1</sup> Penn Frontotemporal Degeneration Center, University of Pennsylvania, Philadelphia, PA, United States, <sup>2</sup> Eleanor M. Safran Center for Cognitive Neuroscience, Temple University, Philadelphia, PA, United States, <sup>3</sup> Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA, United States, <sup>4</sup> Department of Psychology, The University of Edinburgh, Edinburgh, United Kingdom

## OPEN ACCESS

### Edited by:

Vitória Plai,  
Radboud University Nijmegen,  
Netherlands

### Reviewed by:

Laurel Brehm,  
Max Planck Institute  
for Psycholinguistics, Netherlands  
Jennifer E. Mack,  
University of Massachusetts Amherst,  
United States

### \*Correspondence:

Molly B. Ungrady  
molly.ungrady@temple.edu  
Jamie Reilly  
reillyj@temple.edu

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 20 June 2019

**Accepted:** 23 September 2019

**Published:** 11 October 2019

### Citation:

Ungrady MB, Flurie M,  
Zuckerman BM, Mirman D and  
Reilly J (2019) Naming and Knowing  
Revisited: Eyetracking Correlates  
of Anomia in Progressive Aphasia.  
*Front. Hum. Neurosci.* 13:354.  
doi: 10.3389/fnhum.2019.00354

Progressive naming impairment (i.e., anomia) is a core diagnostic symptom of numerous pathologies that impact anterior and inferior portions of the temporal lobe. For patients who experience such regional temporal lobe degeneration, patterns of language loss often parallel the degradation of semantic memory, an etiology of naming impairment known as semantic anomia. Previous studies of semantic anomia have focused extensively on the output of naming attempts by contrasting errors, omissions, and distortions as a function of item-level characteristics (e.g., prototypicality, semantic category). An alternative approach involves evaluating visual confrontation naming as the naming process unfolds. Techniques with high temporal resolution (e.g., eyetracking) offer a potentially sensitive mode of delineating the locus of impairment during naming. For example, a lexical retrieval disorder would hypothetically elicit normal gaze patterns associated with successful visual object recognition regardless of naming accuracy. In contrast, we hypothesize that semantic anomia would be distinguished by aberrant gaze patterns as a function of reduced top-down conceptually guided search. Here we examined visual object recognition during picture confrontation naming by contrasting gaze patterns time locked to stimulus onset. Patients included a cohort of patients with anomia associated with either primary progressive aphasia ( $N = 9$ ) or Alzheimer's disease ( $N = 1$ ) who attempted to name 200 pictures over the course of 18–24 months. We retrospectively isolated correct and incorrect naming attempts and contrasted gaze patterns for accurate vs. inaccurate attempts to discern whether gaze patterns are predictive of language forgetting. Patients tended to show a lower fixation count, higher saccade count, and slower saccade velocity for items that were named incorrectly. These results hold promise for the utility of eyetracking as a diagnostic and therapeutic index of language functioning.

**Keywords:** dementia, primary progressive aphasia, language treatment, language disorder, anomia, eye tracking

## INTRODUCTION

Neurotypical adults can name objects and people with high accuracy and little cognitive effort. However, the ease by which naming unfolds belies the complexity of the cognitive process. Successful confrontation naming (i.e., producing a target name when presented with a picture) demands the precise orchestration of a chain of interactive processes, beginning with visual object recognition, proceeding through semantic processing and lexical retrieval, ultimately resulting in overt articulation. Anomia, or the inability to name common objects and people, can result from disturbances at any stage of this process and is often one of the most functionally debilitating symptoms of living with a neurodegenerative disorder (Laine and Martin, 2006; Henry et al., 2008; Flanagan et al., 2016; Reilly, 2016).

The etiology of impairment in anomia often, but not always, manifests as a distinctive error pattern. The study of naming errors and the treatment of naming disorders in acquired neurogenic language disorders has historically fallen within the purview of aphasiology. Anomia in classical cortical aphasia syndromes is thought to reflect impaired linguistic access to otherwise intact conceptual knowledge (Warrington and Shallice, 1979, 1984; Shallice et al., 1983; Mirman and Britt, 2014). Anomia in post-stroke aphasia tends to manifest in a relatively inconsistent manner on a trial-by-trial basis (e.g., “dog” may be erroneously named as “doll” on one trial and successfully named on another trial). In contrast, anomia in dementia with progressive semantic impairment generally tends to have different properties in terms of its etiology, response consistency, and progression. Hereafter, we refer to this particular etiology of impairment as *semantic anomia*. Semantic anomia does not include a lexical access impairment, but rather is characterized by a loss of object knowledge. Hurley et al. (2012) also found these two distinct patterns of impairments in groups of PPA: lexical access and lexical semantic.

Although this etiology is common in disorders with progressive semantic impairment, it is also not unusual to see a combination of both semantic and lexical access impairments (Mesulam et al., 2009). In its mild stages, Mesulam et al. (2009) find that the semantic variant of Primary Progressive Aphasia (svPPA) starts out as a lexical access impairment, as the patients are able to recognize words but unable to produce the words. As the disease progresses, the patients lose their ability to even recognize the objects, and the semantic anomia becomes a prominent symptom. Thus, once the initial stages of the disease have passed, the inability to name an object likely indicates a semantic breakdown, even if lexical access impairments might be present as well.

One characteristic of semantic anomia, according to Hodges et al. (1996), is a dichotomy between “naming vs. knowing.” In this work, the authors examined the quality of concept definitions for items that patients with Alzheimer’s disease (AD) could successfully name relative to their anomic target items. The principal finding was that of globally impoverished concept definitions for items

that could not be named, a pattern that established a strong correlation between residual semantic knowledge and naming accuracy.

In a related task, Bozeat et al. (2003) examined the correlation between naming and knowing in a longitudinal study of object drawing in svPPA. Patients produced unique errors in the production of line drawings of concrete concepts (e.g., duck, lamp) to a verbally cued label. Comparisons of line drawings produced over time demonstrated a progressive loss of distinctive semantic features, consistent with feature dimming or averaging to the central tendency of a prototype. For example, as semantic impairment worsened, patients added two additional legs to the drawing of a duck, approximating its form as a prototypical four-legged animal. Patients showed a significant association between drawing performance, object naming, and word-to-picture matching. These results provide converging evidence that “knowing” deficits that occur in the presence of progressive semantic impairments compromise a range of verbal and non-verbal abilities, further differentiating progressive aphasia from stroke aphasia.

A reliably strong correlation between naming and knowing confers significant inference to naming ability. If indeed anomia observed in progressive aphasia is predominantly characterized by semantic impairment, then naming ability in this population can reasonably provide a proxy measure for the integrity of semantic knowledge (Reilly et al., 2011a,b). In contrast, inconsistency of naming errors in stroke aphasia and the nature of stroke as an access impairment preclude or at the very least jeopardize the reliability of such inference (Malt, 2019).

Much of the inference gleaned of semantic memory from the analysis of naming is derived from studies of output. Analyses of naming errors lend complementary detail about the processes and mechanisms underlying such output impairments. For example, a disproportionate impairment in the ability to name biological natural kinds relative to manufactured artifacts is one of the hallmarks of category-specific naming impairments associated with AD (Farah and McClelland, 1991; Caramazza and Shelton, 1998; Laws and Sartori, 2005; Lambon Ralph et al., 2007). Furthermore, referring to a knife as “you cut with it” or as a “kitchen thing” demonstrates preserved functional and thematic knowledge in the context of inaccurate retrieval. Despite the inferential value of error and item analyses, an exclusive focus on output affords limited inference about the mechanism underlying anomia. Consider, for example, a retrospective attempt to reconstruct the complex series of events leading up to a ruined recipe. You observe a ruined cake, but at what stage did the process fail? Often, the only possible way to answer to this question requires a perspective that evaluates success or failure of each step in real time as they are added to the recipe. In the domain of visual confrontation naming, eyetracking offers a powerful means of forward inference.

In past studies, eyetracking has provided insights into normal processes that underlie naming (e.g., picture identification, semantic categorization). The visual world paradigm (VWP), for example, involves analyzing gaze patterns to a particular scene or sequence of photographs while hearing verbal descriptors



(Cooper, 1974; Tanenhaus et al., 1995). Among the first to study the VWP was Cooper, who found that as we hear a phrase such as “my scatterbrained dog Scotty. . .” gaze tends to focus on a picture of a dog more than on other unrelated pictures. When they subsequently hear a phrase about a “photographic safari,” their gaze moves from the dog toward the picture of a camera. These findings confirmed that eyetracking tracks eye movement time-locked to verbal cues, and thus provides a unique, time sensitive window into cognitive processes. Similar to this connection between words we hear and gaze patterns, research shows strong connections between words we speak and gaze patterns, with fixation to an object occurring less than 1000 ms before the verbal description of an object (Meyer et al., 1998; Griffin and Bock, 2000). This informs us about the process of naming: first visualize the object, then comprehend the object, next choose the word from the mental lexicon, and finally produce its phonological form. It also demonstrates that the amount of time that a speaker spends fixating on a particular object is contingent on how long they need to complete this process (Griffin, 2001). For example, Meyer et al. (2003) demonstrated that neurotypical adults exhibited a longer fixation duration at an object with a longer and more difficult name than at an object with a shorter and easier name. These studies, among others, support that eyetracking is a tool that allows us insight into the complex process of naming.

In addition to normal or neurotypical processing, eyetracking has also proven useful as a metric for identifying and distinguishing between neurological disorders. For example, posterior cortical atrophy (PCA), also known as the visual variant of AD, is characterized by atypical plaque density within the primary visual cortex (Crutch et al., 2012). PCA patients tend to experience some degree of apperceptive and/or associative visual agnosia (Benson et al., 1988). Shakespeare et al. (2015) identified unique saccade behaviors in PCA patients when compared to typical AD and control subjects in a series of eyetracking assessments of stationary and moving fixation tasks. Specifically, PCA saccadic behaviors were slower and less efficient compared to typical amnesic AD patients. In regard to impairment in AD overall, both PCA and typical AD groups were characterized by reduced fixation stability (i.e., eccentricity in gaze around a focal point). When coupling eyetracking behaviors with volumetric MRI scans, authors were able to suggest different foundational causes for such aberrant fixation responses in each group. Reduced fixation stability in PCA patients was associated with a high frequency of large saccadic intrusions and reduced cortical thickness – suggesting a cognitive foundation for fixation impairment rooted in higher cortical processing. This work not only demonstrates the potential for eyetracking as a tool for identifying impairments but also as a sensitive measure of AD subtypes and their underlying relationships to cognitive processes.

Regarding the semantic anomie patients, studies have shown that eye gaze patterns reveal information about the underlying mechanisms of naming, and thus provide insight into the process of semantic representation. Rösler et al. (2000) demonstrated that in a visual search task where AD patients were asked to find a number or letter target amongst 79 number or letter

distractors, the AD patients exhibited a higher number of fixations, longer fixation durations, and a delayed response time compared to age-matched neurotypical controls. This suggests an inefficient visual search strategy in semantic anomie AD cases (Rösler et al., 2000), as they were unable to efficiently plan a search with minimal fixations organized by a top-down visual search strategy. In PPA, Seckin et al. (2016) recently evaluated the VWP as a means of exploring information processing in a word-to-object matching task. Individuals with PPA were asked to select an object from a circular array that matched a previously presented word (i.e., individuals observed a target word and were instructed to select the matching object image from a selection of 16 object probes). Results indicated that the PPA group demonstrated increased “back and forth” gazing behavior between related foils when compared to controls – offering evidence for bottom-up “probabilistic” mapping in selection rather than an efficient, decisive mapping between semantically matched object probes and word targets.

Previous investigations of naming and knowing have been guided by an offline, output-based empirical perspective such as analyzing correlations between naming with concept definitions, drawing, and word-to-picture matching. Here we employed eyetracking during confrontation naming as a means for affording forward inference about the locus of impairment in semantic anomia. We work from the assumption that anomia in the PPA and AD cases in the current study has a primary etiology of semantic loss, based on previous literature (Caramazza and Shelton, 1998; Reilly et al., 2011a). As such, we hypothesize that patients “know” less about words they cannot accurately name. This dearth of knowledge will impact patterns of visual search during picture naming such that patients will struggle to rapidly fixate on key diagnostic features of items they do not know and cannot name. Therefore, we predict that unnamed items will be subjected to an inefficient search path comprised of more fixations (e.g., looking at many irrelevant features), increased number of saccades (e.g., more undirected looking around), and slower saccade velocity (e.g., unguided search and thus slower to reach a feature).

## MATERIALS AND METHODS

### Overview

We tracked eye movements as participants with progressive naming impairment associated with either AD ( $N = 1$ ) or PPA ( $N = 9$ ) named common objects and familiar and famous people. In the first analysis we used a logistic mixed effects approach to isolate and contrast eye gaze patterns for words with accurate vs. inaccurate responses. In a second analysis we correlated neuropsychological measures of language and memory from the same time points with eyetracking and naming accuracy measures.

### Patients

We included patients with the primary amnesic variant of AD and PPA. Patients included nine patients with PPA and

one patient with AD tested over the span of 18–24 months. Demographic and neuropsychological data appear in **Table 1**.

Among the PPA patients, diagnoses were first established by experienced behavioral neurologists and later confirmed using a consensus approach based on the Gorno-Tempini et al. (2011) diagnostic criteria. The cohort included seven patients with svPPA and two patients with logopenic variant (lvPPA). One patient had a diagnosis of AD established using the McKhann et al. (2011) criteria. Each participant was enrolled for 18–24 months, completing baseline testing upon enrollment and then follow-up testing every 6 months. At the testing sessions, patients completed a battery of neuropsychological tasks, which included Digits Forward and Backward (Wechsler, 2009), Trails A and B (War Department Adjutant General's Office, 1944), the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005), the brief (15-item) form of the Boston Naming Test (BNT) (Kaplan et al., 1983; Mack et al., 1992), and Pyramids and Palm Trees (Howard and Patterson, 1992). We assessed naming at each time point for a combination of control (assigned randomly) and personalized (personal items and family members) picture stimuli ( $N = 200$ ).

## Eyetracking Procedures

We tracked eye movements using an infrared, laptop-mounted eyetracking system (SMI iView X RED eye-tracker) (SensoMotoric Instruments Inc., Boston, MA, United States). We presented picture stimuli using SMI's proprietary software (Experiment Center) and tracked movements of the right eye at a sampling rate of 120 Hz (spatial resolution  $< 0.03^\circ$ ). Patients were seated at a distance of 55–65 cm away from the infrared illuminator bar positioned at the bottom of the laptop monitor. Each eyetracking session initiated with a 5-point calibration and validation procedure. The SMI RED eyetracker uses a low-speed event detection algorithm to define fixations and saccades. This method considers fixations as its primary

event and derives information about saccades based on the fixations. This algorithm considers a group of consecutive points within a particular dispersion, over a defined amount of time as a fixation. We used the default parameters for this definition, with a fixation event defined as when the consecutive points have a maximum dispersion of 100px and a minimum duration of 80 ms.

## Picture Stimuli

Each participant selected a personal lexicon of 100 words that were used with high frequency in their daily lives, broken down into seven categories: people (e.g., their spouse), places (e.g., their church), foods (e.g., bananas), household items (e.g., television), hygiene items (e.g., toothbrush), clothes (e.g., shorts), and activities (e.g., exercising) (for item selection criteria see Reilly, 2016). Once these words were chosen by the participant and their families, pictures of these items were taken in their homes of their own personal items, henceforth referred to as “trained images.” They were then edited, adapted to a laptop, randomized, and presented individually, while the eye tracker recorded their gaze patterns. The onset of each picture was prompted via gaze contingency, where the eye tracker accrued gaze for 1000 ms within a rectangular area of interest (AOI) at the top of the screen prior to the onset of the next stimulus. In addition, each participant was assigned a set of 100 untrained images that served as a control condition. Trained and untrained stimuli were presented in separate blocks using the same stimulus presentation parameters as the trained images. Patients were assigned trained and untrained items using the procedures outlined by Reilly (2016). That is, patients together with their caregivers reviewed fixed lists of words blocked by semantic category (e.g., clothes, hygiene items). Approximately half of the words were chosen by the patient and their caregivers as training targets, whereas the remainder served as untrained controls. Thus, each patient had a different set of trained and untrained items depending personal

**TABLE 1 |** Patient demographics and neuropsychological tests.

Patient	Dx	Age	Year Onset	Edu	BNT	PPTpics	PPTwords	MoCA	DigitsF	DigitsB	TrailATime	TrailBTime
S01	svPPA	60–65	2011	15	1.4 (0.55)	17.2 (7.33)	13.25 (2.99)	9.6 (3.85)	5.8 (0.84)	5.2 (0.45)	105.68 (37.45)	281.6 (47.69)
S02	lvPPA	66–70	2012	13	10.33 (2.09)	25.34 (0.58)	25.34 (0.58)	15.5 (0.71)	6 (1.42)	5 (1.42)	58 (12.73)	136.5 (26.17)
S03	lvPPA	60–65	2011	14	12.2 (1.49)	22.6 (1.68)	23.8 (0.45)	12.8 (2.29)	3.25 (0.5)	2 (0)	59.86 (24.87)	289.5 (21)
S04	AD	76–80	2000	16	4.25 (0.5)	19.75 (3.41)	18.25 (2.07)	13.5 (3.11)	8 (1.42)	6.75 (0.5)	30.29 (6.77)	70.22 (20.06)
S05	svPPA	60–65	2012	18	3.2 (0.84)	21 (1.59)	22.8 (1.93)	19.4 (1.95)	6.8 (0.84)	7.4 (1.15)	32.64 (6.63)	89.24 (9.52)
S06	svPPA	60–65	2009	19	1.75 (0.5)	13.75 (1.5)	13.75 (1.71)	16.5 (1.3)	10.75 (0.96)	9 (0.82)	49.83 (5.8)	102.73 (10.14)
S07	svPPA	60–65	2013	12	4 (0)	22 (2.17)	18.8 (2.39)	16.6 (2.51)	5 (1.23)	6 (0.71)	19.62 (2.92)	49.74 (9.78)
S08	svPPA	66–70	2015	16	3.75 (0.5)	21.25 (2.07)	18.34 (1.16)	18.75 (2.22)	6.25 (0.5)	7.5 (1.74)	32.11 (8.83)	50.99 (6.59)
S09	svPPA	56–60	2011	12	1.5 (0.58)	23.25 (1.5)	17 (4.25)	19.25 (1.5)	8.67 (0.58)	8.34 (1.53)	26.47 (11.25)	63.88 (14.32)
S10	svPPA	60–65	2010	16	3.5 (0.71)	21.67 (0.58)	20.5 (2.13)	17.5 (2.13)	6 (1.42)	5.5 (0.71)	40.19 (0.15)	113.24 (6.7)
Max/Norm	NA	NA	NA	NA	15/13.2	26	26	30/26	9	8	NA	NA

Neuropsychological performance,  $M(SD)$  of patients across all time points. Dx, Diagnosis (svPPA, semantic variant Primary Progressive Aphasia; lvPPA, logopenic variant Primary Progressive Aphasia; AD, Alzheimer's Disease); Age, Age at baseline; YearOnset, Year diagnosed; Edu, Years of Education; BNT, Boston Naming Test Score; PPTpics, Pyramids and Palm Trees Test Score-Pictures; PPTwords, Pyramids and Palm Trees Test Score-Words; MoCA, Montreal Cognitive Assessment Score; DigitsF, Digit Span-Forward Score; DigitsB, Digit Span-Backward Score; TrailATime, Trail Making Test-A time to complete (seconds); TrailBTime, Trail Making Test-B time to complete (seconds). In the last row, we included the maximum score for each test and gave means based on normative data where applicable (Mack et al., 1992; Nasreddine et al., 2005).

preference. Both conditions, trained and control images, were included in analysis and collapsed across conditions.

## Naming Procedures and Scoring

Patients were asked to verbally state the name of the items in all of the pictures after the presentation of the image. They were allowed unlimited time to provide an answer. When necessary patients were cued semantically first, phonologically second, however, only the spontaneous response was scored as either accurate or inaccurate. If the participant self-corrected their spontaneous response, we considered their self-correction as the response to score. We utilized a binary scoring protocol, where responses were either correct or incorrect. Patients were asked to name both sets of pictures, the personalized lexicon and the canonical images, at baseline, and then every 6 ( $\pm 2$ ) months for up to 2 years.

## Eyetracking Metrics

All eyetracking data were windowed to  $2750 \pm 250$  ms upon presentation of the stimulus in order to control for differences in patterns that could result from analyzing a wide range of presentation time (i.e., a stimulus that was viewed for 200 ms vs. a stimulus that was viewed for 5000 ms). When extracting the data from SMI BeGaze, we exported the eyetracking data from the first 3000 ms, and then further restricted the presentation time in RStudio to  $2750 \pm 250$  ms prior to analysis.

Eyetracking studies of scene viewing tend to encompass measures of depth and breadth of visual attention (e.g., where is someone looking). Depth of visual attention is typically indexed by fixation measures (e.g., count, duration) which are thought to quantify deep processing of particular elements of scenes (e.g., faces). In contrast, breadth of search is indexed by saccade measures (e.g., count, amplitude). Since we are interested both in the depth and breadth of visual search, we analyzed a range of fixation and saccade measures commonly used in scene perception research. These included fixation count, fixation duration total, fixation dispersion total, saccade count, saccade duration total, saccade amplitude total, and saccade velocity total. Upon visual inspection of the data, we isolated four eye gaze metrics. The first measure was fixation count, which includes the total number of fixations that occurred within the windowed timeframe. Second, we assessed the fixation dispersion. However, this measure was highly correlated with fixation count ( $r = 0.95$ ) and due to potential multicollinearity was not included in the analysis. Next we assessed the number of saccades (i.e., saccade count). Finally, we evaluated saccade velocity, defined by the change in eye position (degrees) divided by seconds. For this study, we divided the saccade velocity by 100 in order to keep the unit (ms) between each of our variables consistent.

## Data Analysis and Statistical Procedures

We employed a logistic mixed effects model to assess predictors of item-level accuracy using the “lme4” package within R, collapsing across condition (i.e., trained or untrained pictures) and time points. Fixed effects included: number of fixations, number of saccades, and saccade velocity. Random effects included

participant and item. To evaluate the unique contribution of each of the fixation measures to model fit, each measure was iteratively removed from the model while leaving the other two fixed effects in the model. That is, each fixation measure's unique contribution to model fit was assessed while controlling for the other fixation measures. Changes in goodness of model fit were assessed by the likelihood ratio test: two times the change in log-likelihood, which is distributed as  $\chi^2$  with degrees of freedom corresponding to the difference in number of parameters (one in each of these comparisons).

In order to characterize the association of eyetracking behaviors with various measures of cognition, we performed a correlation analysis between eyetracking performance and neuropsychological performance. Pearson correlations were generated based on mean performance of each participant at each time-point. Incomplete observations were removed from analysis.

The dataset for this study can be found in Open Science Framework<sup>1</sup>.

## RESULTS

Each of the eyetracking measures were significant predictors of naming accuracy. **Table 2** reflects output of the logistic mixed effects model predicting item accuracy.

Total number of fixations was significantly lower for incorrectly named items (mean = 7.61, SD = 2.75) than for correctly named items (mean = 7.65, SD = 2.70;  $p < 0.001$ ). The total number of saccades produced during the viewing of incorrectly named images (mean = 8.20, SD = 3.22) was significantly higher than the number of saccades for correct responses (mean = 8.06, SD = 2.97;  $p < 0.001$ ). Finally, saccade velocity for the incorrectly named items (mean = 890.76, SD = 3.22) was significantly slower than for the correctly named items (mean = 937.34, SD = 549.08;  $p < 0.001$ ). Model-predicted associations between each of these measures and accuracy are shown in **Figure 1A**. To further describe the data, we added a violin plot (**Figure 1B**) of each eyetracking variable.

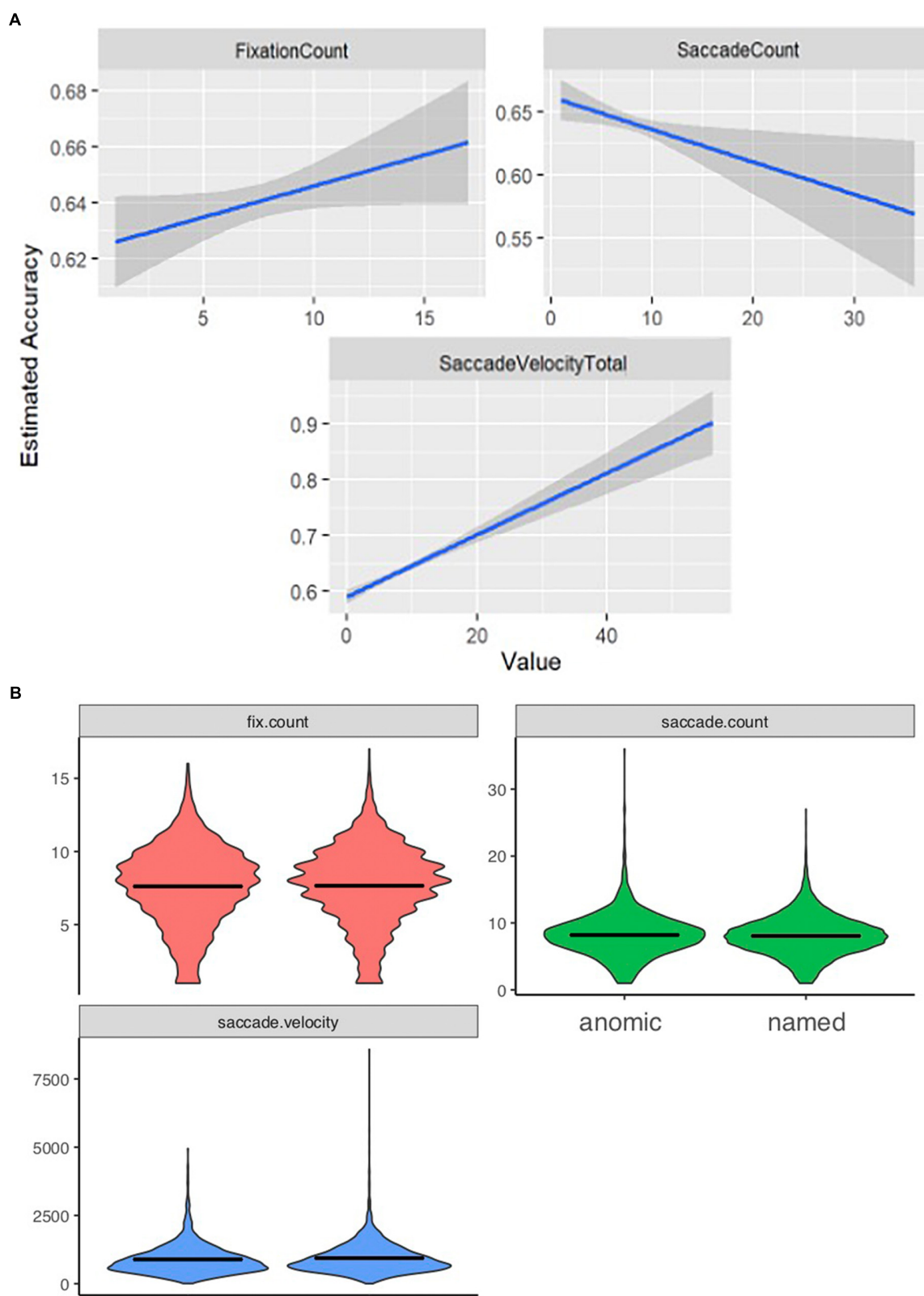
We see that a lower fixation count, a slower saccade velocity and a higher saccade count are significant predictors of lower accuracy, and thus impaired knowing.

<sup>1</sup><https://mfr.osf.io/render?url=https://osf.io/eutr2/?action=download%26mode=render>

**TABLE 2 |** Mixed logistic regression.

	Estimate (SE)	$\chi^2(1)$	<i>p</i>
Number of fixations	0.065 (0.018)	12.6	<0.001
Number of saccades	−0.0919 (0.018)	26.0	<0.0001
Saccade velocity	0.0496 (0.0085)	33.3	<0.0001

*This table shows the results of the mixed logistic regression. The number of fixations, the number of saccades, and the saccade velocity were all significant predictors of inaccurate naming. This analysis was completed in R and the syntax for the model was “glmer(Acc ~ Fixation Count + Saccade Count + Saccade Velocity + (1| Patient) + (1| Item), data = , family = binomial).”*



**FIGURE 1 |** Eyetracking patterns based on accuracy. Panels **(A)** shows how the eyetracking measures predict accuracy. As accuracy is lower patients displayed a lower fixation count; a higher saccade count, and a slower saccade velocity. **(B)** Represents the range and distribution of data observed for each eyetracking measure [fix.count (fixation count), saccade count, and saccade velocity] between the anomic (items inaccurately named) and named (items accurately named) items.



## Neuropsychological Performance and Naming Inaccuracy

**Figure 2** represents a correlation matrix describing relationships between gaze metrics and offline neuropsychological measures for incorrectly named items. Global cognitive performance as measured by the MoCA was moderately positively correlated with total number of fixations ( $r = -0.35$ ,  $p = 0.03$ ), total number of saccades ( $r = -0.24$ ,  $p = 0.03$ ), and total saccade velocity ( $r = 0.47$ ,  $p = 0.01$ ), indicating that reductions in global cognition were associated with more diffuse gaze, less focused attention, and slower patterns of looking at points between attentional fixation. Executive functioning as indexed by Digit Span-Backward was positively correlated with saccade velocity total ( $r = 0.48$ ,  $p = 0.03$ ).

We observed no significant correlations between offline measures of naming performance, semantic memory, and gaze.

## DISCUSSION

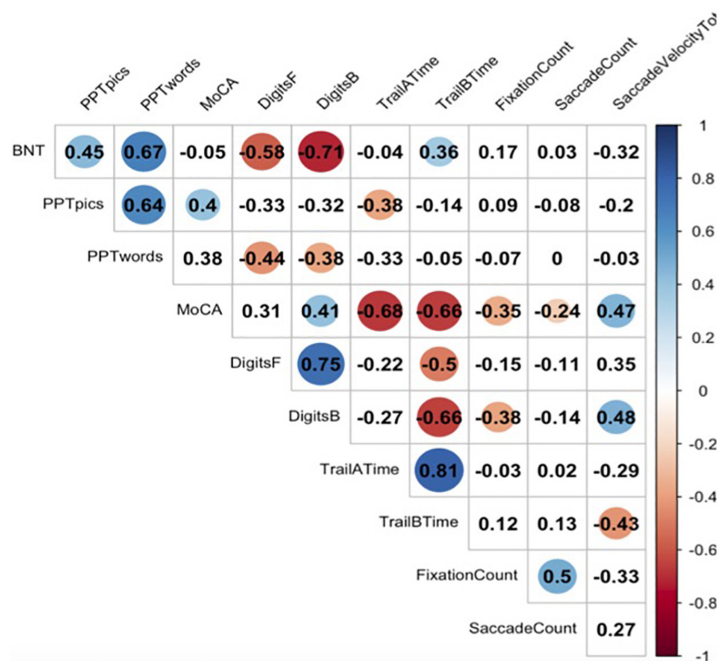
We examined eyetracking as a sensitive measure for evaluating naming in the context of progressive semantic impairment. Little is known about how semantic impairment impacts visual object recognition and how these associated gaze patterns predict naming accuracy. Our working hypothesis is that visual confrontation naming involves a combination of top-down expectancies and bottom-up, salience driven processing. This hypothesis was also tested by Choi et al. (2017) who found that neurotypical older adults were just as easily able to access

top-down and bottom-up strategies as younger adults in order to optimize reading strategy.

A global deficit in semantic processing would, therefore, reduce the top-down contribution, forcing reliance upon bottom-up salience. This division of labor between conceptual expectancy and sensory-driven visual search has been reported in other experimental paradigms (Yarbus, 1967). However, a comprehensive account of gaze behaviors that parallels semantic degradation during naming is lacking.

The following gaze patterns differentiated known (named accurately) from forgotten (anomic) items: saccade count, fixation count, and saccade velocity. Specifically, forgotten items were associated with a lower fixation count, slower saccade velocity, and an increased number of saccades. Scan paths for forgotten items appeared unguided and disorganized with unstable gaze patterns. That is, the eyes are not staying still long enough to constitute a fixation, and instead moving around enough to be counted as saccades.

A few aberrant gaze patterns could likely arise as the result of impoverished semantic knowledge in picture naming. Due to a loss of top-down knowledge of an object, patients experiencing semantic anomia might engage in a bottom-up driven search, thus resulting in disorganized searching around the image to identify the diagnostic and salient features of an object. This would result in increased fixation and saccadic occurrences. Alternatively, the inability to identify diagnostic features of an object might result in a slowed and unguided visual search when attempting to name an object. Originally, we hypothesized the first pattern of behaviors and predicted a more sporadic visual search approach. However, our results suggest that for words



**FIGURE 2 |** Correlations between eyetracking metrics and neuropsychological tasks. Here we see significant correlations between neuropsychological measures and eyetracking measures. A blue dot indicates a positive correlation, and a red dot indicates a negative correlation. Numbers reflect Pearson correlation coefficients.

that were named incorrectly patients demonstrated the second pattern of behavior described, showing a slowed search strategy with fewer fixations in the given window.

The observed gaze patterns differ from our original predictions and from previous work in several respects. First, patients fixated less often for items they could not name, whereas related work using more complex visual arrays has demonstrated *more* fixations for forgotten items (Seckin et al., 2016). This discrepancy between studies could have resulted from different task demands. Consider the tasks involved in studies discussed: Rösler et al. (2000) visual-search task among letters and numbers, and Seckin et al. (2016) picture-word matching task. Both of these paradigms required patients to select an item from an array of competing stimuli. Reduced top-down semantic support for naming compelled patients to make a probabilistic selection based on bottom-up guided visual search. As a result, patient selections were characterized by numerous revisits among competing stimuli. In contrast, visual confrontation naming in our study involved only a single stimulus per trial with no extrinsic competition between picture stimuli on any individual trial. This methodological discrepancy between studies may account for the observed differences in the amount of fixation. In the current study, inaccurate naming (i.e., forgetting) was associated with fewer fixations. This pattern may reflect inability to effectively seek and focus on diagnostic semantic features necessary to accurately name a target picture. A similar interpretation of “feature dimming” was offered by Bozeat et al. (2003) in reference to altered picture drawing (e.g., drawing a duck with four legs) documented as their longitudinal patient cohort declined over time.

Another possible account of our finding that fewer fixations predicts when items are unknown relates to a division of labor in visual object recognition between global visual form (e.g., shape) vs. local visual detail (e.g., texture, facial features) (Bar, 2003; Bar et al., 2006). During picture viewing, simultaneous detail is available both from low and high spatial frequencies (Bar, 2003; Bar et al., 2006). Low spatial frequency details about global visual form (e.g., shape) can be assessed without fixating on the picture. In contrast, local visual form (e.g., texture) is represented by high spatial frequency detail, often requiring fixation(s) in numerous places. A neurotypical person names pictures both rapidly and accurately because they can effectively integrate low and high spatial frequency information. In contrast, the lack of top-down semantic support might compel the patient to conduct an unfocused search, attending to low spatial frequencies. Such a search strategy would be characterized by a high saccade count with a correspondingly low fixation count.

In a companion study we observed the longitudinal eye gaze patterns of objects that are consistently known, vulnerable to being forgotten, and objects that are consistently forgotten over the course of a 2-year study. The stimuli and procedure of presentation were identical to the current study. This companion analysis found a u-shaped pattern of eyetracking as objects go from known to vulnerable to forgotten. When the objects are known, the fixation count is low suggesting a streamlined and efficient top-down visual search. For words that are vulnerable to being forgotten the fixation count spikes and indicates that the

patients are attempting to name the item by fixating on many different places to find the important features. However, once the words progress from being vulnerable to being completely forgotten, the fixation count drops below that of known words. This finding suggests that a low fixation count is a behavior that can result from two different stages of progressive anomia: a streamlined and organized visual search resulting in effective naming, or a slowed and unguided search resulting in incorrect naming (Reilly et al., under review). The eyetracking pattern for the latter stage of progressive anomia supports our finding that a low fixation count can in fact predict unknown words, although not words that are vulnerable and on the trajectory of being forgotten.

## Neuropsychological Correlations With Eyetracking

We assessed global cognition, working memory, language, semantic memory, and attention using a variety of offline neuropsychological measures. Patients showed correlations between online measures of eyetracking during naming with several of these offline neuropsychological measures (see **Figure 2**). Significant correlations were observed between saccade velocity MoCA score, digits backward, and Trail B time (in seconds). Significant correlations were also observed between saccade count and MoCA score. Additionally, a significant correlation was found between the MoCA and fixation count. However, this correlation did not follow a linear distribution and should be interpreted with caution.

Slowed saccade velocity predicted naming accuracy, and gaze slowing occurred in conjunction with declines in global cognition. Changes in saccade velocity could have either a cognitive or motor etiology. Lueck et al. (2000) found that patients with AD exhibit irregular saccades (e.g., more forward saccades per line and more saccadic regressions) during text reading compared to controls. Lueck et al. (2000), among other authors, have also found that increased saccadic abnormalities are correlated with a more severe cognitive impairment (Schewe et al., 1999). These studies link saccade behavior to difficulties with lexical-semantic access in AD. In contrast, a relative minority of studies have linked abnormal saccade behavior in AD to oculomotor dysfunction (Hutton et al., 1979; Pirozzolo and Hansch, 1981). Although oculomotor dysfunction is a plausible cause of saccade slowing, the observed correlations with declining global cognition suggest more of a cognitive etiology in our patient cohort (see also Scinto et al., 1994).

## Limitations

We did not observe expected correlations between neuropsychological tasks and eyetracking data, as we predicted a positive correlation between eyetracking patterns and the tests measuring semantic knowledge and naming ability. While the BNT showed some variation (1–14), this variation came from only two patients out of this cohort. Eight patients performed consistently at floor performance with little variance. This indicates that the patients included in this study began with an impaired semantic understanding. We did in fact see a

correlation between eyetracking patterns and the MoCA. Since the MoCA is a measure of global cognition that assesses more than just the semantic impairment, these scores might not have started low but rather showed a progressive decline over time as patients became more impaired.

Secondly, although previous studies have ruled out oculomotor difficulties in their eyetracking studies (Scinto et al., 1994), we did not have our own experiment to rule out this possibility in our own cohort of patients.

In the current paper we dichotomized the data as either known (e.g., accurate) or unknown (e.g., inaccurate) and collapsed across time points in order to determine if there are eyetracking patterns that can predict accuracy. We recognize that it does not explicitly examine the change over time, although we believe it has important implications for such a longitudinal investigation. Reilly et al. (under review), described above, conducted this longitudinal analysis in the same cohort of patients.

Furthermore, we acknowledge that access-based anomie errors are common in PPA (Mesulam et al., 2009). While it is clear that the patients in our cohort exhibit progressive anomia, it is not as clear whether this is due to a semantic impairment or other causes (e.g., lexical access). In **Table 1** we report scores for the neuropsychological task, Pyramid and Palm Trees (PPT), that assess semantic knowledge. Further examination of semantic vs. lexical access impairments would be useful in determining the cause of anomia in this cohort of patients and strengthen our findings.

## Clinical Implication

In all, these results support the notion that individuals with progressive anomia demonstrate specific gaze patterns for preserved concepts vs. impoverished concepts in naming tasks.

Though previous studies have characterized naming capabilities in PPA, no studies until now have identified eyetracking behaviors uniquely associated with known vs. forgotten items in progressive anomia. Such findings hold promise for the use of eyetracking as a clinical tool capable of identifying impoverished concept knowledge in progressive anomia. This finding has vast clinical implications for personalized language interventions. Recent work has advocated for the use of maintenance-based interventions over compensatory or restorative interventions, as *maintaining* a lexicon is more efficacious than *relearning* a lexicon for patients with progressive semantic degradation (Reilly, 2016). Using the approach of eyetracking during picture naming, therapists may be able to create patient-specific inventories of “at-risk” target words at the onset of treatment. With the ability to reliably predict which words will drop from a patient’s lexicon, interventions could adjust focus on an item-specific basis. Such treatments could maximize the prolongation of preserved concept knowledge and provide patients and their families with a personalized treatment that would help to maintain communication for as long as possible.

## Future Directions

Future work ought to specifically examine item-specific gaze patterns associated with items as they transition from known to unknown. There remains to be a comprehensive account of gaze patterns illustrating the progression of concept degradation, which inevitably leads to naming impairment. This work could lead to the use of gaze metrics as a cost-effective, mobile tool for preclinical identification of semantic impairment. Pairing this personalized treatment with a non-invasive brain stimulation that has been shown to increase naming speed and improve visual search (Binney et al., 2018), might further augment the benefit that this eyetracking treatment would exhibit on the naming performance of the patients. Binney et al. (2018) demonstrated that the use of transcortical direct current stimulation (tDCS) improves patients’ ability to locate salient features of an object during confrontation naming.

## CONCLUSION

This study demonstrates that eyetracking is a useful tool to detect the degradation of concept knowledge, as our results show that saccade velocity and the amount of fixations and saccades are significant predictors of unknown items. This information could be used to develop clinical therapies for progressive anomia; a devastating symptom with currently very few treatments.

## DATA AVAILABILITY STATEMENT

The dataset for this study can be found in Open Science Framework (<https://mfr.osf.io/render?url=https://osf.io/eutr2/?action=download%26mode=render>).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Boards of Temple University and the University of Pennsylvania.

## AUTHOR CONTRIBUTIONS

JR and DM contributed to the design of the study. MU collected the data. MU and MF organized the database. JR, DM, and MF performed the statistical analyses. BZ contributed to the statistical analyses. MU, MF, DM, and JR contributed to the writing and editing of the drafts.

## FUNDING

This study was supported by a grant NIH/NIDCD DC013063.

## REFERENCES

- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454.
- Benson, D. F., Davis, R. J., and Snyder, B. D. (1988). Posterior cortical atrophy. *Arch. Neurol.* 45, 789–793.
- Binney, R. J., Ashaie, S. A., Zuckerman, B. M., Hung, J., and Reilly, J. (2018). Frontotemporal stimulation modulates semantically-guided visual search during confrontation naming: a combined tDCS and eye tracking investigation. *Brain Lang.* 180, 14–23. doi: 10.1016/j.bandl.2018.04.004
- Bozeat, S., Lambon Ralph, M. A., Graham, K. S., Patterson, K., Wilkin, H., Rowland, J., et al. (2003). A duck with four legs: investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cogn. Neuropsychol.* 20, 27–47. doi: 10.1080/02643290244000176
- Caramazza, A., and Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *J. Cogn. Neurosci.* 10, 1–34. doi: 10.1162/089892998563752
- Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., and Henderson, J. M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: a comparison between young and older adults. *Psychol. Aging* 32, 232–242. doi: 10.1037/pag0000160
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cogn. Psychol.* 6, 84–107. doi: 10.1016/0010-0285(74)90005-x
- Crutch, S. J., Lehmann, M., Schott, J. M., Rabinovici, G. D., Rossor, M. N., and Fox, N. C. (2012). Posterior cortical atrophy. *Lancet Neurol.* 11, 170–178.
- Farah, M. J., and McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *J. Exp. Psychol. Gen.* 120, 339–357. doi: 10.1037/0096-3445.120.4.339
- Flanagan, K. J., Copland, D. A., van Hees, S., Byrne, G. J., and Angwin, A. J. (2016). Semantic feature training for the treatment of anomia in Alzheimer disease: a preliminary investigation. *Cogn. Behav. Neurol.* 29, 32–43. doi: 10.1097/WNN.0000000000000088
- Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014. doi: 10.1212/WNL.0b013e31821103e6
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition* 82, B1–B14.
- Griffin, Z. M., and Bock, K. (2000). What the eyes say about speaking. *Psychol. Sci.* 11, 274–279. doi: 10.1111/1467-9280.00255
- Henry, M. L., Beeson, P. M., and Rapcsak, S. Z. (2008). Treatment for lexical retrieval in progressive aphasia. *Aphasiology* 22, 826–838. doi: 10.1080/02687030701820055
- Hodges, J. R., Patterson, K., Graham, N., and Dawson, K. (1996). Naming and knowing in dementia of Alzheimer's type. *Brain Lang.* 54, 302–325. doi: 10.1006/brln.1996.0077
- Howard, D., and Patterson, K. (1992). *The Pyramids and Palm Trees Test: A Test of Semantic Access from Words and Pictures*. Bury St. Edmunds: Thames Valley Test Company.
- Hurley, R. S., Paller, K. A., Rogalski, E. J., and Mesulam, M. M. (2012). Neural mechanisms of object naming and word comprehension in primary progressive aphasia. *J. Neurosci.* 32, 4848–4855. doi: 10.1523/JNEUROSCI.5984-11.2012
- Hutton, J. T., Johnston, C. W., Shapiro, I., and Pirozzolo, F. J. (1979). Oculomotor programming disturbances in the dementia syndrome. *Percept. Motor Skills* 49, 312–314. doi: 10.2466/pms.1979.49.1.312
- Kaplan, E., Goodglass, H., and Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia: Lea and Febiger.
- Laine, M., and Martin, N. (2006). *Anomia: Clinical and Theoretical Aspects*. New York, NY: Psychology Press.
- Lambon Ralph, M. A., Lowe, C., and Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia. *HSVE Neural Netw. Model. Brain* 130(Pt 4), 1127–1137. doi: 10.1093/brain/awm025
- Laws, K. R., and Sartori, G. (2005). Category deficits and paradoxical dissociations in alzheimer's disease and herpes simplex encephalitis. *J. Cogn. Neurosci.* 17, 1453–1459. doi: 10.1162/0898929054985428
- Lueck, K. L., Mendez, M. F., and Perryman, K. M. (2000). Eye movement abnormalities during reading in patients with Alzheimer disease. *Neuropsychiatry Neuropsychol. Behav. Neurol.* 13, 77–82.
- Mack, W. J., Freed, D. M., Williams, B. W., and Henderson, V. W. (1992). Boston naming test: shortened versions for use in Alzheimer's disease. *J. Gerontol.* 47, 154–158.
- Malt, B. C. (2019). Words, thoughts, and brains. *Cogn. Neuropsychol.* 9, 1–13. doi: 10.1080/02643294.2019.1599335
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dementia* 7, 263–269.
- Mesulam, M. M., Rogalski, E., Wieneke, C., Cobia, D., Rademaker, A., Thompson, C., et al. (2009). Neurology of anomia in the semantic variant of primary progressive aphasia. *Brain* 132, 2553–2565. doi: 10.1093/brain/awp138
- Meyer, A. S., Roelofs, A., and Levelt, W. J. M. (2003). Word length effects in object naming: the role of a response criterion. *J. Mem. Lang.* 48, 131–147. doi: 10.1016/s0749-596x(02)00509-0
- Meyer, A. S., Sleiderink, A. M., and Levelt, W. J. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition* 66, B25–B33.
- Mirman, D., and Britt, A. E. (2014). What we talk about when we talk about access deficits. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20120388.
- Nasreddine, Z. S., Phillips, N. A., Badirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x
- Pirozzolo, F. J., and Hansch, E. C. (1981). Oculomotor reaction time in dementia reflects degree of cerebral dysfunction. *Science* 214, 349–351. doi: 10.1126/science.7280699
- Reilly, J. (2016). How to constrain and maintain a lexicon for the treatment of progressive semantic naming deficits: principles of item selection for formal semantic therapy. *Neuropsychol. Rehabil.* 26, 126–156. doi: 10.1080/09602011.2014.1003947
- Reilly, J., Peelle, J. E., Antonucci, S. M., and Grossman, M. (2011a). Anomia as a marker of distinct semantic memory impairments in Alzheimer's disease and semantic dementia. *Neuropsychology* 25, 413–426. doi: 10.1037/a0022738
- Reilly, J., Rodriguez, A. D., Peelle, J. E., and Grossman, M. (2011b). Frontal lobe damage impairs process and content in semantic memory: evidence from category-specific effects in progressive non-fluent aphasia. *Cortex* 47, 645–658. doi: 10.1016/j.cortex.2010.05.005
- Rösler, A., Mapstone, M. E., Hays, A. K., Mesulam, M., Rademaker, A., Gitelman, D. R., et al. (2000). Alterations of visual search strategy in Alzheimer's disease and aging. *Neuropsychology* 14, 398–408. doi: 10.1037//0894-4105.14.3.398
- Schewe, H. J., Uebelhack, R., and Vohs, K. (1999). Abnormality in saccadic eye movement in dementia. *Eur. Psychiatry* 14, 52–53. doi: 10.1016/s0924-9338(99)80716-0
- Scinto, L. F., Daffner, K. R., Dressler, D., Ransil, B. I., Rentz, D., Weintraub, S., et al. (1994). A potential noninvasive neurobiological test for Alzheimer's disease. *Science* 266, 1051–1054. doi: 10.1126/science.7973660
- Seckin, M., Mesulam, M.-M., Voss, J. L., Huang, W., Rogalski, E. J., and Hurley, R. S. (2016). Am I looking at a cat or a dog? Gaze in the semantic variant of primary progressive aphasia is subject to excessive taxonomic capture. *J. Neurolinguistics* 37, 68–81. doi: 10.1016/j.jneuroling.2015.09.003
- Shakespeare, T. J., Kaski, D., Yong, K. X., Paterson, R. W., Slattery, C. F., Ryan, N. S., et al. (2015). Abnormalities of fixation, saccade and pursuit in posterior cortical atrophy. *Brain* 138, 1976–1991. doi: 10.1093/brain/awv103
- Shallice, T., Warrington, E. K., and McCarthy, R. A. (1983). Reading without semantics. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 35A, 111–138.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863



- War Department Adjutant General's Office, (1944). *Army Individual Test Battery: Manual of directions and scoring*. Washington, DC: War Deartment Adjutant General's Office.
- Warrington, E. K., and Shallice, T. (1979). Semantic access dyslexia. *Brain* 102, 42–63.
- Warrington, E. K., and Shallice, T. (1984). Category specific semantic impairments. *Brain* 107, 829–854.
- Wechsler, D. (2009). *Wechsler Memory Scale*, 4th Edn. San Antonio, TX: Pearson Inc.
- Yarbus, A. L. (1967). *Eye Movements and Vision*. New York, NY: Plenum Press.

**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 © 2019 Ungrady, Flurie, Zuckerman, Mirman and Reilly. 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.



# Input Complexity Affects Long-Term Retention of Statistically Learned Regularities in an Artificial Language Learning Task

Ethan Jost<sup>1</sup>, Katherine Brill-Schuetz<sup>2</sup>, Kara Morgan-Short<sup>2</sup> and Morten H. Christiansen<sup>1\*</sup>

<sup>1</sup> Department of Psychology, Cornell University, Ithaca, NY, United States, <sup>2</sup> Department of Psychology, University of Illinois at Chicago, Chicago, IL, United States

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Jessica Hall,  
University of Arizona, United States  
Eleonore H. Smalle,  
Ghent University, Belgium  
Zhenghan Qi,  
University of Delaware, United States

### \*Correspondence:

Morten H. Christiansen  
christiansen@cornell.edu

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 14 June 2019

**Accepted:** 26 September 2019

**Published:** 15 October 2019

### Citation:

Jost E, Brill-Schuetz K,  
Morgan-Short K and Christiansen MH  
(2019) Input Complexity Affects  
Long-Term Retention of Statistically  
Learned Regularities in an Artificial  
Language Learning Task.  
Front. Hum. Neurosci. 13:358.  
doi: 10.3389/fnhum.2019.00358

Statistical learning (SL) involving sensitivity to distributional regularities in the environment has been suggested to be an important factor in many aspects of cognition, including language. However, the degree to which statistically-learned information is retained over time is not well understood. To establish whether or not learners are able to preserve such regularities over time, we examined performance on an artificial second language learning task both immediately after training and also at a follow-up session 2 weeks later. Participants were exposed to an artificial language (Brocanto2), half of them receiving simplified training items in which only 20% of sequences contained complex structures, whereas the other half were exposed to a training set in which 80% of the items were composed of complex sequences. Overall, participants showed signs of learning at the first session and retention at the second, but the degree of learning was affected by the nature of the training they received. Participants exposed to the simplified input outperformed those in the more complex training condition. A GLMM was used to model the relationship between stimulus properties and participants' endorsement strategies across both sessions. The results indicate that participants in the complex training condition relied more on an item's chunk strength than those in the simple training condition. Taken together, this set of findings shows that statistically learned regularities are retained over the course of 2 weeks. The results also demonstrate that training on input featuring simple items leads to improved learning and retention of grammatical regularities.

**Keywords:** statistical learning, artificial language learning, second language learning, retention, memory

## INTRODUCTION

Statistical learning (SL) has been identified as a domain-general cognitive ability that is integral to language processing, acquisition, and evolution (see Armstrong et al., 2017, for an overview). For the purposes of this study, we can define SL as the process by which learners uncover the structure of the input from its distributional properties (Frost et al., 2015). However, little is known about the extent to which statistically learned information is retained over time (see Gomez, 2017, for a review), particularly in adult learners.

Initial studies of SL focused on the rapidity with which human infants could learn from predictable, structured sequences of input (e.g., Saffran et al., 1996). As a result, the literature has

remained quite focused on measuring the ability of participants to learn these regularities within a single session, usually with a test-phase following some sort of training. However, there have been only a handful of studies examining the ability of adult participants to retain statistically learned information over long periods of time.

Previous work that has focused on adult SL includes Kim et al. (2009). This study demonstrated that participants implicitly learned the statistical relationships governing a sequence of rapidly presented visual stimuli, and that this learning was retained over the course of 24 h. Other work has shown that adults possess the ability to maintain information about the underlying relationship between visual stimuli in an SL task, with testing periods at 30 min, 1, 2, 4, and 24 h delays (Arciuli and Simpson, 2012). Participants showed no difference in their ability to correctly identify grammatical sequences, suggesting that at least over the course of a day, the information gleaned within a(n) SL task is relatively robustly retained. The authors of this study notably suggest that their findings do not indicate any sort of enhancement in retention for participants who slept between training and test.

Additional research, however, has attempted to examine the role of sleep in the consolidation of associations learned within a(n) SL paradigm, and a few of these findings have a bearing on retention and SL more generally. Although the present study does not seek to examine the effects of sleep on SL, this is perhaps the most well-studied aspect of long-term retention within the literature. Napping appears to improve consolidation, as participants who slept during a 4 h delay period between training and test outperformed those who did not during an auditory discrimination task (Durrant et al., 2011). Interestingly, this enhancement was positively correlated with the amount of slow-wave sleep obtained by the participant. Similarly, researchers have shown that participants who slept were more likely to apply statistically learned constraints in a speech production task (Gaskell et al., 2014). This study also demonstrated a positive relationship between slow-wave sleep and learning effects. In general, it seems that over a relatively brief period of time, knowledge gained in a(n) SL task can be retained, and that this retention may even be enhanced by sleep in some instances. More recently, another sleep study by Frost and Monaghan (2017) demonstrated that participants who underwent a period of sleep between training and test within a non-adjacency SL paradigm outperformed those who stayed awake at both word learning and generalizing the rules of the grammar to new sequences that had not been seen during training.

Two studies in particular stand out as examples of investigations into a more traditional definition of long-term retention and consolidation of sequence learning abilities. Romano et al. (2010) demonstrated that participants seemed to retain sequence-specific learning and general skill effects a year after training on a serial reaction time task (pressing a key corresponding to a target circle's location as targets appeared on the computer screen). Retention was observed across a variety of training groups: younger adults, older adults, experienced musicians, and video game players. In effect, participants recalled frequent triplets more quickly than they did low-probability

control trials, showing a learning effect over the course of the tasks at session one that persisted at session two, 1 year later.

Kobor et al. (2017) attempted to extend these findings in a task designed to test consolidation along with retention, as they were interested in uncovering the core mechanism(s) that underlie long-term memory formation in such a task. This study investigated the role of retroactive interference in forgetting by training participants on a new set of items with an alternate statistical structure 1 day after the initial test session. The second test session in this study, which tested long-term retention of the initially trained patterns, took place a full year later, similar to the Romano et al. (2010) study. Again, the researchers found learning effects for highly frequent items relative to infrequent items, and also found no effect for the potentially interfering materials. An additional test demonstrated that the knowledge gained in this task seemed to be implicit in nature. They took this to mean that long-term memory for statistically learned sequences does undergo a process of consolidation that appears to be robust and resistant to some kinds of interference. Moreover, learning scores were reported as relatively stable between the first session, the interference training session the next day, and the final session a year later.

While the previous two studies demonstrate the persistence of sequence learning abilities over a long stretch of time, they are still limited in a few important ways. First, the tasks in both studies required only visuospatial to motor mappings without any auditory or verbal component, limiting the degree to which their findings might generalize to language itself. Second, and relatedly, the learned sequences did not contain any kind of meaning. While this is a common practice within the SL literature, it limits the study's ecological validity when it comes to addressing the mechanisms thought to underlie language learning (Morgan-Short, in press). Third, the statistical structure underlying the training items was not very complex in either study, again somewhat undermining the claim that the kinds of relationships learned between items in a sequence are characteristic of those in natural language. Finally, neither of these studies demonstrated a quintessential feature of learning in SL and artificial grammar learning (AGL), the generalization of learned regularities to new items. The test sets in each task contained exclusively items on which participants had already been trained.

A second set of studies that has focused more directly on natural and artificial second language learning has also provided evidence of retention over periods of time. Indeed, the delayed posttest as a measure of retention is not uncommon in second language acquisition research as shown in a meta-analysis (Norris and Ortega, 2000), as nearly half of all studies featured some kind of follow-up test phase more than a week after training. This meta-analysis revealed a robust learning effect (Cohen's  $d = 1.02$ ) at delayed testing. Specifically addressing the question of second language retention, Morgan-Short et al. (2012a) reported behavioral and neurophysiological evidence of retention over the course of 3–6 months for the same artificially learned language used in this study, Brocanto2. Using an artificial, as opposed to natural, language allowed for control of prior experience and extra-experimental exposure to the second language. Participants in Morgan-Short et al. (2012b) achieved high levels of proficiency

(~95%) by completing three training and practice sessions, in which they completed a total of 36 comprehension and production practice modules comprised of 20 practice items each. When tested for retention 3–6 months later, there was no evidence of a decline in their performance (Morgan-Short et al., 2012a). Because of the extensive practice provided in Morgan-Short et al. (2012b), it is not possible to determine if exposure itself led to retention. Thus, the current study aimed to examine retention of grammatical regularities after shorter periods of exposure to input but without practice. It also leveraged the artificial language paradigm Brocanto2 to control exposure to the second language and to manipulate the input (see below).

Given the frequently described links between SL and language, it would seem likely that the associations learned in commonly used SL paradigms should persist over longer periods of time than we currently have robust evidence for, an issue the current study seeks to address with the hypothesis that participants will retain learned information over the course of 2 weeks. In other words, if statistically learned information truly undergirds our language learning abilities, it must be retained beyond an immediate post-test in the lab.

## Input Complexity Affects the Learning of Statistical Regularities

The idea that learners process the co-occurrence statistics of the input in the service of acquiring more abstract grammatical regularities is not new (Elman, 1990; Altmann, 2002). This processing ability has been proposed to develop as we learn the most basic information available from the input first, as suggested by the “less is more,” hypothesis: that is, beginning to learn without fully developed cognitive abilities could convey an advantage to children (Newport, 1990). This notion has been applied to the process of language learning and has been pointed out as a potential reason for the existence of sensitive periods in language acquisition (Johnson and Newport, 1989; Newport, 1990, 2016). The corresponding idea that “starting small” may be advantageous for learners shares similar longevity within the literature (Elman, 1993; Elman et al., 1996), and emphasizes the possible benefit that reduced complexity within the learner’s input (e.g., in terms of length or syntactic complexity) has on learning.

Although the evidence for these hypotheses has subsequently become somewhat less straightforward (for example, see Rohde and Plaut, 1999; Siegelman and Arnon, 2015), new research is emerging that, within the context of artificial language learning, participant performance may benefit from training that becomes progressively more challenging (Kersten and Earles, 2001; Lai and Poletiek, 2011). A recent study has shown that starting small leads to better learning of recursive structures, with the primary facilitation coming from a gradual increase in stimuli complexity rather than simply the effect of reduced length (Poletiek et al., 2018).

Other work has also shown that artificially biasing the kinds of chunks that adults form to be more simplified can lead to improved learning in a Hebb-repetition paradigm (Smalle et al., 2016). Smalle et al. (2018) expanded upon this idea

by showing that children exhibited better retention of implicitly learned phonological sequences within a Hebb-repetition task than adults in a longitudinal design with a year between the first and last test sessions. This study demonstrated the long-term retention of input containing probabilistic dependencies, highlighting the importance of chunking as a potential factor for both learning and retention within such paradigms. However, this study left open the relative importance of input complexity on learning, and did not test for generalization of learned knowledge to novel test items.

Taken in conjunction with other recent ideas, chunking can be seen as an integral component of the SL process as it applies to language (Isbilen and Christiansen, 2018; Christiansen, 2019). Rapidly recoding and compressing information by chunking may allow learners to more efficiently process input, and to do so at higher levels of abstraction. In fact, stronger learners may show a decreased reliance on surface-level fragment information when tested due to the fact that they have already used that information to internalize the higher-order regularities, and no longer rely on them as a crutch.

## The Current Study

The present study seeks to examine the different ways in which learners retain knowledge about the grammatical regularities of an artificial language, Brocanto2 (Morgan-Short et al., 2010, 2012b), through the process of SL. To that end, we conducted original analyses of unpublished data from Brill-Schuetz (2016). In this study, training conditions differed by the amount of exposure to complex stimuli presented in the training, where complexity was related to the cognitive demands needed to process an item. In the Simple condition, half of the participants in this study received a more simplified set of training items generated by the grammar. This manipulation attempted to mimic the constraints placed on young learners by the simplified input they tend to receive (Cameron-Faulkner et al., 2003). Training sets with progressively increasing difficulty have been used in past AGL and SL studies for similar reasons (e.g., Conway et al., 2003; Christiansen et al., 2012; Poletiek et al., 2018). Those in the simple training condition were eventually exposed to complex items, but the extensive experience they received with simple items before moving on to the more complex ones is expected to boost performance in the test phase of the experiment. Therefore, our first prediction is that learners, particularly those who receive simple training, will retain knowledge from training over the span of 2 weeks.

In the Complex condition, the other half of participants received far less training with simplified items prior to exposure to the set of complex items yet obtained the same amount of total experience in terms of number of trials. These participants are thus predicted to have more trouble learning, and subsequently retaining, the rules of the artificial language as they would have insufficient experience processing simple constructions before encountering the more difficult complex items. This may lead them to adopt poor learning strategies, disrupting their extraction of the relevant statistical structure embedded within the sequence. In short, more exposure to simple items (i.e., the simple training condition) should confer an advantage to the



learning of the grammar rules that govern Brocanto2 both in the short- and long-term (Brill-Schuetz, 2016).

We are also interested in finding out how the different training groups approach the task of endorsing items as grammatical, by looking into what features of the test items are most relevant to such judgments. While the Brocanto2 artificial language learning paradigm was not designed to test SL, the underlying distributional information embedded within its dependencies offers a potential window into the ways that learners use statistical regularities to learn language. Examining endorsement strategies is expected to provide insight into what each group of participants retained from the task across both sessions, and what kinds of information they are sensitive to. The specific cues that participants rely on to make grammaticality judgments might vary between the training groups, and if participants in the complex training condition show the reduced sensitivity to the grammatical regularities of the language that we predict, we hypothesize that they will instead be found to rely more on fragment information, such as chunk strength. On the other hand, the simple training group will likely not be as distracted by surface-level similarities between training and test items and will rather demonstrate knowledge of the higher-order grammatical regularities.

## MATERIALS AND METHODS

### Participants

Participants ( $N = 47$ ; Male = 10) were young adult students at a large, Midwestern university, ranging in age from 18 to 24 ( $M = 19.43$ ,  $SD = 1.98$ ). Recruitment for the first session was conducted through a psychology department subject pool where participation earned class credit. For the second session, some participants received additional credit through a subject pool and others received monetary compensation (\$5). Selection criteria limited participants to those who had no hearing, learning, or speaking impairments, and to native speakers of English. All participants provided written consent before beginning the study<sup>1</sup>.





The second session took place approximately 2 weeks after the original training session. Although every effort was made to schedule the delayed post-test exactly 2 weeks from the original session, the actual range was between 12 and 14 days from the training session. At this second session, some ( $N = 33$ ) participants also completed an additional battery of cognitive tests.

### Materials

#### Artificial Language

The artificial language learned by participants was Brocanto2<sup>2</sup> (Morgan-Short et al., 2010, 2012b), which was adapted from the original version, Brocanto (Friederici et al., 2002). Brocanto2 follows basic patterns typical of many natural languages and is

**TABLE 1 |** Complete list of words used within the artificial language learning task.

Word category	Brocanto2 word	Symbol/meaning
Noun	Pleck <sub>m</sub>	
	Neep <sub>m</sub>	
	Blom <sub>f</sub>	
	Vode <sub>f</sub>	
Adjectives	Trois(e <sub>m</sub> /o <sub>f</sub> )	Circle
	Neim(e <sub>m</sub> /o <sub>f</sub> )	Square
Determiners	li <sub>m</sub> /lu <sub>f</sub>	The
Verbs	klin <sub>intran</sub>	Move
	praz <sub>tran</sub>	Switch
	nim <sub>tran/intran</sub>	Capture
	yab <sub>tran/intran</sub>	Release
Adverbs	Noyka	Vertically
	Zayma	Horizontally

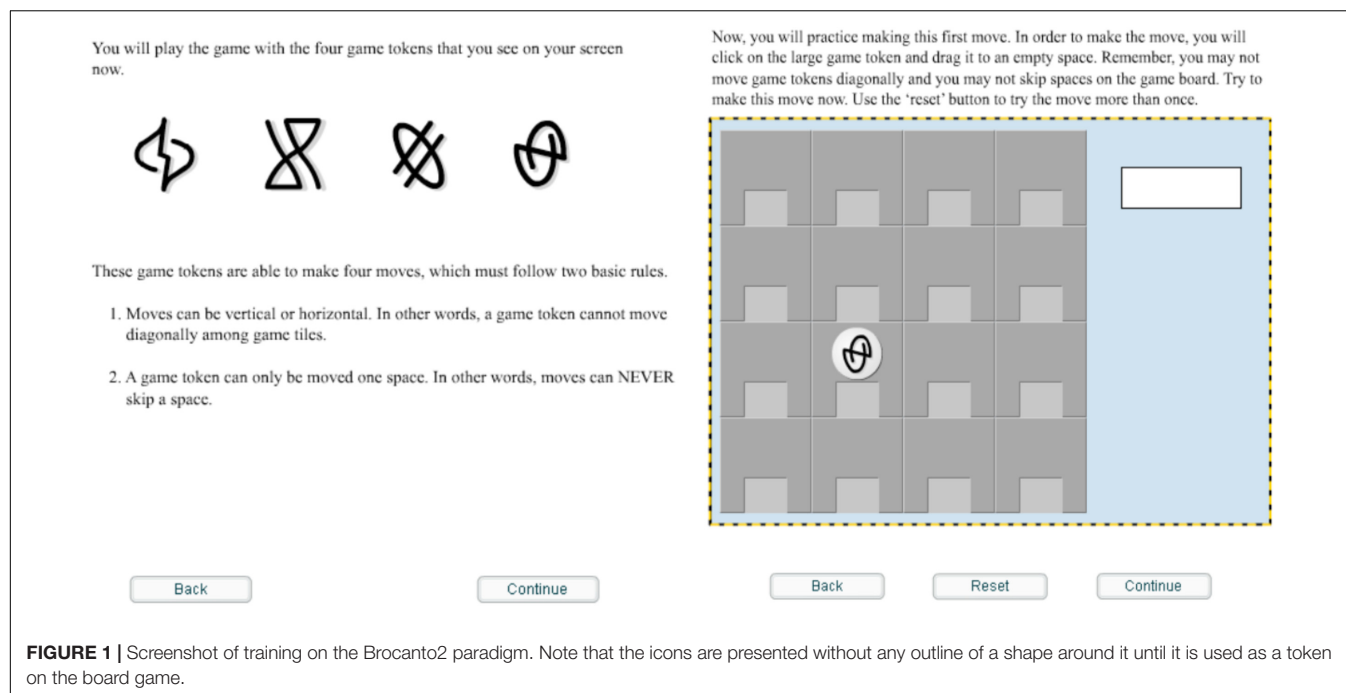
*Subscripts denote the gender of each noun and determiner along the corresponding marking for each adjective, and also the transitive nature of each verb. The adjectives described the shape of the area bordering the game piece, such as the circle that can be seen in Figure 1. Table adapted from Morgan-Short (2007).*

fully productive; it consists of 14 novel words: four nouns, two adjectives, two articles, four verbs, and two adverbs (see Table 1 for a list of all words and their meanings). The grammatical structure of this language follows a syntactic pattern different from that of English; while English follows a subject-verb-object order, Brocanto2 follows a subject-object-verb order, which is found in languages such as Hindi and Japanese. For example, the Brocanto2 sentence “Blom neimo lu neep troise li praz zayma” corresponds to “Blom-piece square the neep-piece round the switch horizontally” and would be translated into English as “The square blom-piece switches with the round neep-piece horizontally.” Participants learned this artificial language in order to play a computer-based game in which the tokens can move according to dictation in Brocanto2 (see Figure 1).

These sentences could be either simple or complex in nature; simple stimuli were limited to words from three of the word categories (noun, article, verb) and could consist of three to five lexical items. Complex stimuli consisted of words from all five of the categories allowed in Brocanto2 (noun, adjective, article, verb, adverb) and a complex sentence could contain five to eight lexical items (Brill-Schuetz, 2016). For example, the sample sentence given above would be classified as a complex item due to the inclusion of the adjectives and the adverb, a difference highlighted within Table 2. The presentation of each sentence was consistent in that all the noun phrases were simple or complex and all verb phrases were either simple or complex; for example, a sentence would not have a simple noun phrase followed by a complex verb phrase. See Table 2 for examples of both complex and simple sentences. During the beginning of training (but not test), the simple and complex stimuli included noun phrases presented without a corresponding verb or adverb. The simple phrases had only a noun and a determiner, while the complex phrases included noun, adjective, and determiner. Figure 2 illustrates all possible word class combinations and identifies the two kinds of

<sup>1</sup>This research was approved by the University Human Subjects Institutional Review Board under protocol 2008–0496.

<sup>2</sup>Brocanto2 is available upon request by email to karams@uic.edu.



phrases and four kinds of sentences that could be generated by the Brocanto2 grammar.

## Procedure

### Brocanto2 Artificial Language Learning Paradigm

Participants were taught the Brocanto2 vocabulary to 100% accuracy prior to starting any other aspects of the study. Vocabulary training consisted of a self-paced PowerPoint presentation that paired Brocanto2 audio with the symbols for nouns or a general animation to signify that an action was happening. At no point during the vocabulary training was explicit information given regarding spelling, translations, or parts of speech. The vocabulary assessment was a second PowerPoint presentation that replicated the training PowerPoint with one important difference: participants had to generate the correct Brocanto2 word for each slide. Therefore, participants

had to self-generate 100% of the vocabulary before progressing to the next phase of the training.

The participants were then presented with game training that consisted of an introduction to the computerized board game they would be playing at a later point, thus providing a meaningful context for the artificial language on which they were subsequently trained<sup>3</sup>. Participants read the rules of the game and viewed the four possible types of game moves (move, switch, capture, or release). They were then asked to practice making each move on the game board by selecting game tokens with a mouse and repeating the move that had just been visually presented, as illustrated by **Figure 3**. At no point were explicit translations of the symbols or movements provided. After becoming familiar with the rules of the game, participants continued on to language training. Note that all participants received the same vocabulary and game training – it was not part of the manipulation.

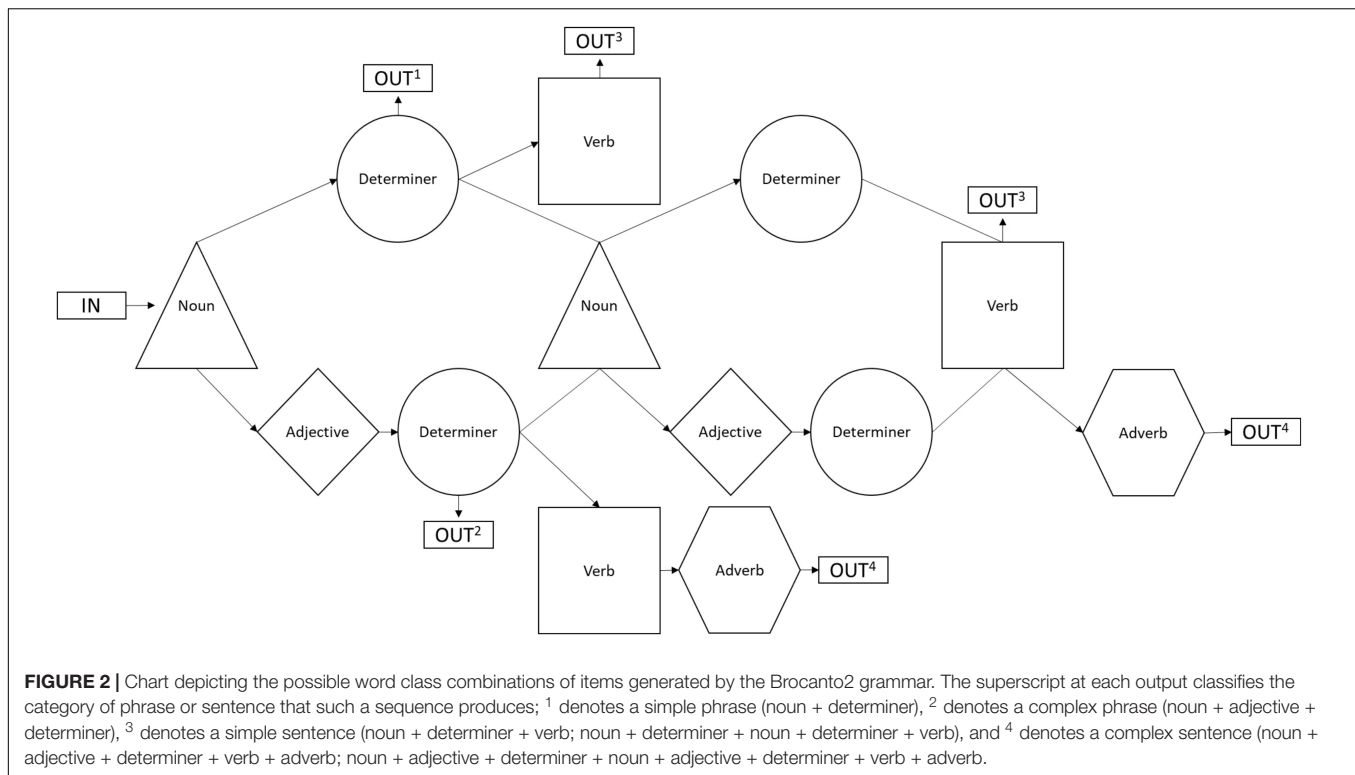
Before beginning the task, participants were instructed that they would receive training on an artificial language and would be presented with words, phrases, and sentences that would correspond to still and moving images on the game board. Participants were told they would complete a short quiz to test their memory and they would not be able to review this information again. They were also informed that they would then use the artificial language to play a board game at a later point. No other instructions were given; therefore, training can be viewed as implicit or uninstructed (not incidental) due to the lack of explicit information or explanation of the Brocanto2 language rules. Note, however, that the implicit, uninstructed format of

**TABLE 2 |** Examples of simple and complex input for klin and praz in Brocanto2.

	Brocanto2 sentence	Word categories
<b>Simple input</b>		
Klin <sup>^</sup>	Blom lu klin	N Det + V
Praz <sup>+</sup>	Blom lu neep li praz	N Det + N Det + V
<b>Complex input</b>		
Klin <sup>^</sup>	Blom neimo lu klin noyka	N Adj Det + V Adv
Praz <sup>+</sup>	Blom neimo lu neep troise li praz noyka	N Adj Det + N Adj Det + V Adv

Example sentences from both complexity conditions containing the two verbs that could not be both transitive and intransitive. Noun, N; determiner, Det; verb, V; adjective, Adj; adverb, Adv; <sup>^</sup>denotes intransitive verb and <sup>+</sup> denotes transitive verb.

<sup>3</sup>Participants played the computerized board game as part of a comprehension assessment that followed the GJT, the results of which are not reported here.



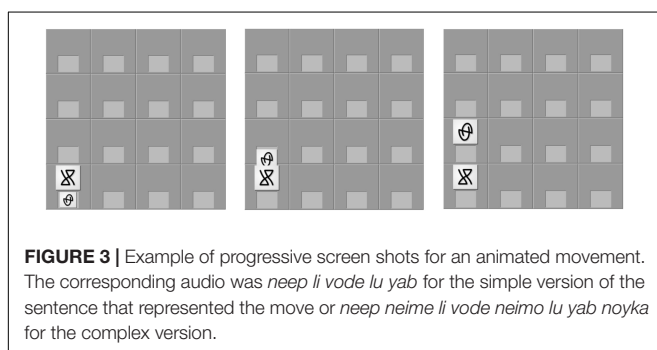
the training does not entail that learning is necessarily implicit in nature.

Participants were pseudorandomly assigned to either simple ( $N = 24$ ) or complex ( $N = 23$ ) input conditions, with every other learner assigned to the simple condition. All participants received training phrases and sentences featuring identical nouns and verbs, presented either in a simple or complex format (100 items). Thirty-six of the training items were phrases, while 64 were sentences.

In the “simple” training condition, 80% of the sentences that participants received were simple while the other 20% were complex; in the “complex” training condition, 80% of the sentences were complex while 20% were simple. This particular ratio of stimuli was utilized so that participants would be exposed to every word category in Brocanto2 and its function in a sentence while still presenting a vast majority of one particular type of stimuli. Furthermore, a 1:4 ratio has also been used

in previous cognitive linguistics studies examining the learning and generalization of grammatical regularities for novel verbs (e.g., Casenhiser and Goldberg, 2005).

Participants were presented with the Brocanto2 stimuli aurally and *always* received simple stimuli *before* complex stimuli regardless of the training condition. Each training condition began with the visual presentation of the 36 individual symbols that corresponded to Brocanto2 noun phrases (simple and complex) and progressed to 64 fully animated moves with corresponding sentences (simple and complex). That is, all participants received the training items in the following order: simple phrases, complex phrases, simple sentences, complex sentences. This ordering of phrases being presented before full sentences follows the structure of previous Brocanto2 studies (e.g., Morgan-Short et al., 2012b, 2014) and that of studies exploring the starting small hypothesis (e.g., Kersten and Earles, 2001; Conway et al., 2003; Poletiek et al., 2018). Presentations for each noun phrase consisted of a single, static game piece while the audio was played. An animated movement involving one or more pieces on the game board accompanied the presentation of sentences, and in this case, the audio was played before the animated movement occurred. At the conclusion of each noun phrase or animation, there was a 1 s break before the next item appeared on screen. The game pieces and animations presented to participants were identical across the two conditions—the training only varied in terms of the audio. More specifically, participants in the simple training condition were presented with twenty-nine simple noun phrases followed by seven complex noun phrases, and then fifty-one simple sentences followed by thirteen complex sentences. In the complex condition, the overall



order of sentence types would remain the same, but participants would instead be trained on seven simple noun phrases followed by twenty-nine complex noun phrases, and then thirteen simple sentences followed by fifty-one complex sentences.

The primary language assessment in this study consisted of a grammaticality judgment task (GJT). The GJT requires the participant to make a judgment regarding the grammaticality (yes or no) of a sentence and is commonly used across second-language learning literature (cf. Loewen, 2009; Plonsky et al., 2019). The GJT consisted of 72 novel sentences, half (36) of the stimuli were simple sentences and half were complex. Of the 36 simple sentences, half were correct and half contained a violation; this was also the case for the complex sentences. The same set of GJT items was used at each session.

Grammatical sentences for the GJT were novel, i.e., correct sentences that were not presented during training. In general, ungrammatical items were generated by introducing violations in the novel, correct sentences. However, four ungrammatical simple sentences had to be created using violations of sentences that appeared in training due to the limited number of such sentences that could be generated by the grammar. There were an equal number of word order (6), verb argument (6), and gender (6) violations in both the simple and complex GJT stimuli. Word order violations were created by replacing a word from one of the five word categories (e.g., noun) with a word from a different category (e.g., adjective, article, verb, adverb). Verb argument violations were created by replacing a transitive verb with an intransitive verb and vice versa, therefore these violations were constrained to the verbs *klin* and *praz*. Grammatical gender violations were created by replacing a feminine adjective or article with a masculine adjective or article, and vice versa. Violations never occurred on the first or final word, and violation position among words was distributed as evenly as possible. Word frequency within each grammatical category was also as equally distributed as possible across all sentences. Examples of each type of violation sentence can be found in **Table 3**.

The GJT was programmed in SuperLab 5 and the stimuli (the Brocanto2 sentences) were randomized. The GJT began by guiding participants through the instructions; all directions were presented in white font (size 30) on a black background. The initial screen informed participants that the task was to make a series of judgments regarding new sentences in the artificial language. These judgments were, in order: grammaticality (good or bad), confidence rating (confident or not confident), and source attribution (rule, memory, intuition, or guess). Participants were asked to make each judgment as quickly and accurately as possible. Although the confidence rating and source attribution data is not analyzed for the current study, the full methodology is presented so that the reader fully understands the task demands and to acknowledge that this could have influenced other results (see Brill-Schuetz, 2016, for analyses of the confidence ratings and source attributions).

## Hypotheses and Planned Analyses

This experimental design enables us to examine three separate main hypotheses. The first, that participants will exhibit retention of statistically learned sequences within an artificial language over

**TABLE 3 |** Example correct Brocanto2 sentences and violations thereof.

Sentence type	Brocanto2 sentence
Correct sentence	<i>Blom neimo lu neep neime li praz</i> Blom-piece square the neep-piece square the switch "The square blom-piece switches with the square neep-piece."
Word Order (Syntactic)	<i>Blom *nim lu neep neime li praz</i> Blom-piece *capture the neep-piece square the switch
Violation sentence	"The *capture blom-piece switches with the square neep-piece."
Verb Argument	<i>Blom neimo lu *praz</i> Blom-piece square the *switches
Violation sentence	"The square blom switches *(missing object)"
Morphosyntactic (Gender Agreement)	<i>Blom neimo lu neep *neimo li praz</i> Blom square the neep-piece (fem) *square(masc) the switches
Violation sentence	The square blom switches with the square <sub>(m)</sub> neep <sub>(f)</sub>

\*Denotes the location of the violation.

the course of 2 weeks, will be tested by examining whether or not participants' accuracy is above chance on the GJT at the second session. In addition, we will examine the degree to which this performance is maintained across sessions, as perfect retention is not expected. As a reminder, participants were trained under what can be considered implicit training conditions, meaning they received repeated exposure to the language without any explicit instruction on the rules of the grammar. The second hypothesis, derived from the "starting-small" literature, is that participants trained on the simpler set of items will outperform their peers in the complex condition overall. Better learning due to the reduced input complexity received in training will lead to better memory both in the short- and long-term for those in the simple training condition. This will be evaluated by looking at the relative performance (accuracy) on the GJT of each group on the GJT at both the first and second sessions. Specifically, participants in the simple training condition should show stronger learning at session one, and this advantage will be carried forward to session two as well.

The third hypothesis consists of multiple parts; the first part of hypothesis three is that the complex training group will rely more on chunk strength when judging the grammaticality of test items than the simple training group; thus, there may be an interaction effect between chunk strength and group. Hypothesis three will be assessed by modeling the relationship between the properties of each test item (e.g., chunk strength) and the likelihood that participants from each group would choose to endorse that item as grammatical. For these analyses we will rely on measures of endorsement rather than accuracy in order to better isolate the aspects of the knowledge participants used to discern between grammatical and ungrammatical test items. Given our interest in better understanding the retention of the grammar embedded within the artificial language, we also wanted to examine how the hypothesized effect of chunk strength on participants' grammaticality judgments manifested



itself across sessions, expecting that its influence might diminish over time. Therefore, the second part of hypothesis three predicts that there will be an interaction between chunk strength and time. We planned to model this interaction using a GLMM, following them up with a series of correlational analyses. Overall, we hope to show that participants in the complex training group rely to a greater extent on chunk strength when endorsing items, and that this reliance changes to some degree over time.

## RESULTS

### Participant Performance and Retention

Before presenting analyses related directly to our hypotheses and research questions, t-tests were conducted to validate that participants exhibited evidence of learning from the two training conditions. As shown in **Table 4**, those in the simple training condition demonstrated above chance performance at both the first [ $t(23) = 4.018, p = 0.001, d = 0.83$ ] and second [ $t(23) = 3.835, p = 0.001, d = 0.75$ ] sessions. Those in the complex condition showed above chance accuracy at session one [ $t(22) = 2.907, p = 0.008, d = 0.60$ ], but not at session two [ $t(22) = -0.172, p = 0.865, d = 0.03$ ]. These results are taken to support that learning had taken place in both the simple and complex training condition.

Related to the first hypothesis about retention, overall participant accuracy was above chance (i.e., 50%) when judging items as grammatical or ungrammatical at both sessions one [ $t(46) = 4.774, p < 0.001, d = 0.69$ ; mean: 56.9% correct; standard deviation: 0.10; 95% CI: 54.1–60.0%] and two [ $t(46) = 2.452, p = 0.018, d = 0.36$ ; mean: 53.6% correct; standard deviation: 0.10; 95% CI: 50.7–56.6%]. A paired t-test to examine how GJT accuracy degraded between sessions showed that while participants did show above chance performance at session two, it was significantly lower than their performance at session one [ $t(46) = -3.0, p = 0.004, d = 0.33$ ]. This demonstrates that participants retained knowledge of the pattern of the artificial language's grammatical regularities over the course of 2 weeks, although this retention was not perfect.

In regard to the second hypothesis about what whether the type of training affected accuracy and retention, we examined how each group of participants performed on the GJT across both sessions. Looking deeper to see what aspects of training affected accuracy and retention, a 2 (session)  $\times$  2 (training condition) mixed ANOVA analyzing accuracy showed significant main

effects for both session [ $F(1, 45) = 9.058, p = 0.004, \eta_p^2 = 0.168$ ] and training group [ $F(1, 45) = 6.872, p = 0.012, \eta_p^2 = 0.132$ ], while the interaction effect did not reach significance [ $F(1, 45) = 0.796, p = 0.377, \eta_p^2 = 0.017$ ]. In spite of the non-significant interaction term, a follow-up on group differences was performed in order to clarify the different pattern of results found between groups, which should not be over-interpreted. A set of paired t-tests showed that participants in the simple training condition did not exhibit a statistically significant change in performance between sessions [ $t(23) = 1.46, p = 0.158, d = 0.23$ ], while those in the complex training condition performed significantly better at session one than they did at session two [ $t(22) = 2.84, p = 0.009, d = 0.55$ ].

While this set of results indicates that those in the complex training condition did not exhibit learning or retention as well as those in the simple training condition, it is also possible that they were sensitive to other aspects of the items besides their grammaticality. That is, it is possible that they learned some features of the training set besides the grammar and used those as cues when accepting or rejecting items. Other analyses that are specific to performance related to test items, complexity (simple vs. complex) and grammatical structure (syntax, morphosyntax, and verb argument), are reported in Brill-Schuetz (2016).

### Modeling Predictors of Item Endorsement

In relation to hypothesis three, in order to get a clearer picture of the type(s) of information to which participants in either group showed sensitivity, we used the dependent variable of endorsement rates rather than accuracy. Endorsement rates were calculated by looking at the proportion of “yes” responses when participants were asked if they thought a GJT test item was grammatical, even when it was not. More specifically, when a participant responded “yes” to either a grammatical or ungrammatical item, they would receive a score of “1” whereas when they responded “no” to either a grammatical or ungrammatical item, they would receive a “0” instead. Endorsement rates for each group at both sessions can be found in **Table 5**. Using endorsement rates for particular items will allow us to figure out how each group may have used the information they statistically learned when performing the GJT in a way that just looking at the group's mean performance (percent correct) cannot. For each item, we can connect the chunk strength of that item to the likelihood that it was endorsed by the participants; thus we will be able to determine the sub-features of the items that most strongly led participants to say “yes” and “no” to them when making grammaticality judgments at test.

After calculating these endorsement rates, we investigated what fixed factors were the strongest predictors of item endorsement. To do this we used a series of generalized linear mixed effect models (GLMMs) to examine the effects of training condition, chunk strength, and time (session) on item endorsement using the LME4 package in R (Bates et al., 2014). The model included as fixed effects: training group (complex vs. simple), chunk strength of GJT item

**TABLE 4 |** GJT accuracy performance by training condition across sessions.

	Simple training		Complex training	
	Session 1	Session 2	Session 1	Session 2
Correct (SD)	59.9% (0.12)	57.5% (0.09)	54.2% (0.07)	49.7% (0.09)
95% CI	54.8–65.0%	53.4–61.5%	51.2–57.1%	45.6–53.7%

Mean percent correct on the GJT for participants in each training condition at both sessions, along with standard deviations in parentheses. 95% Confidence Intervals are reported beneath each mean. These statistics were calculated by-subject.

**TABLE 5 |** GJT response patterns by training condition for item endorsement across sessions.

	Simple training		Complex training	
	Session 1	Session 2	Session 1	Session 2
Grammatical (SD)	64.4% (0.13)	64.4% (0.14)	55.6% (0.28)	58.0% (0.17)
95% CI	60.3–69.0%	59.6–69.1%	46.1–65.1%	52.2–63.9%
Ungrammatical (SD)	45.5% (0.15)	49.5% (0.17)	48.2% (0.20)	58.5% (0.17)
95% CI	40.4–50.5%	43.8–55.2%	41.6–54.9%	52.7–64.4%

Endorsement rates on the GJT for participants in each training condition at both sessions for both grammatical and ungrammatical items, along with standard deviations in parentheses. 95% Confidence Intervals are reported beneath each mean. These statistics were calculated by-item.

(continuous), and time (session one vs. session two). We included as a random effect the intercepts for GJT endorsement by subject. This controlled for individual differences in response bias, making it easier to detect fixed effects of our variables of interest.

The chunk strength of each item was calculated in order to determine the extent to which the participants used this kind of fragment information when endorsing items. The chunk strength referenced here was measured as the sum of the frequency of occurrence in the training items of each of the fragments in a test item, weighted by the number of fragments in that item (Knowlton and Squire, 1994). For example, the associative chunk strength of the item ZVX would be calculated as the sum of the frequencies of the fragments ZV, VX, and ZVX divided by 3. A higher number indicates that a test item is well supported by chunk information in the training items. Chunk strength thus captures the repeated use of 2- and 3-element chunks in a sequence, allowing for generalization from known sequences to novel ones. So, just because a test item did not occur in training that does not mean that some portion of it did not appear as part of a training item. If “the brown cat” is a training item while “the brown cow” is a test item, the chunk “the brown” appeared in both, and therefore would contribute to the chunk strength of the test item.

With the sets of training and test items used in this study, chunk strength actually was significantly greater for grammatical vs. ungrammatical over all test items, meaning that it was a potentially useful cue for performing accurately on this task for both the simple [ $t(70) = 2.268, p = 0.026, d = 0.53$ ] and complex [ $t(70) = 2.396, p = 0.019, d = 0.56$ ] training groups; this was coincidental, as chunk strength was not factored in when creating the stimuli for this experiment. Note that these comparisons were computed separately given that the two groups had different training sets, even though the test sets were exactly the same. Descriptive statistics for the chunk strength of both grammatical and ungrammatical test items for each training group can be found in Table 6.

The initial model (Model 1) with separate fixed effects is reported in Table 7. However, due to the nature of the manipulation and the variables of interest, another model with three two-way interaction terms was built. This model (Model 2) was primarily built in order to appropriately control for the

two-way interactions’ inclusion in the final model containing the key three-way interaction term. Additionally, we hypothesized that the effect of chunk strength on item endorsement may potentially degrade with time due to the nature of memory, thus we included an interaction term between these variables. The results for Model 2, which include these interaction terms, are also reported in Table 7. To test if the inclusion of interaction terms improved upon Model 1, a deviance test was conducted (Singer and Willett, 2003). The interaction terms improved model fit,  $\chi^2(3) = 76.681, p < 0.0001$ .

A further desire to also include a potential three-way interaction between training condition, session, and chunk strength led to the creation of Model 3. This model outperformed Model 2 [ $\chi^2(1) = 13.716, p = 0.0002$ ], supporting the hypothesis that the effect of training on retention would differ between groups. Importantly, we know that the three-way interaction

**TABLE 6 |** Average chunk strength of test items for each training group.

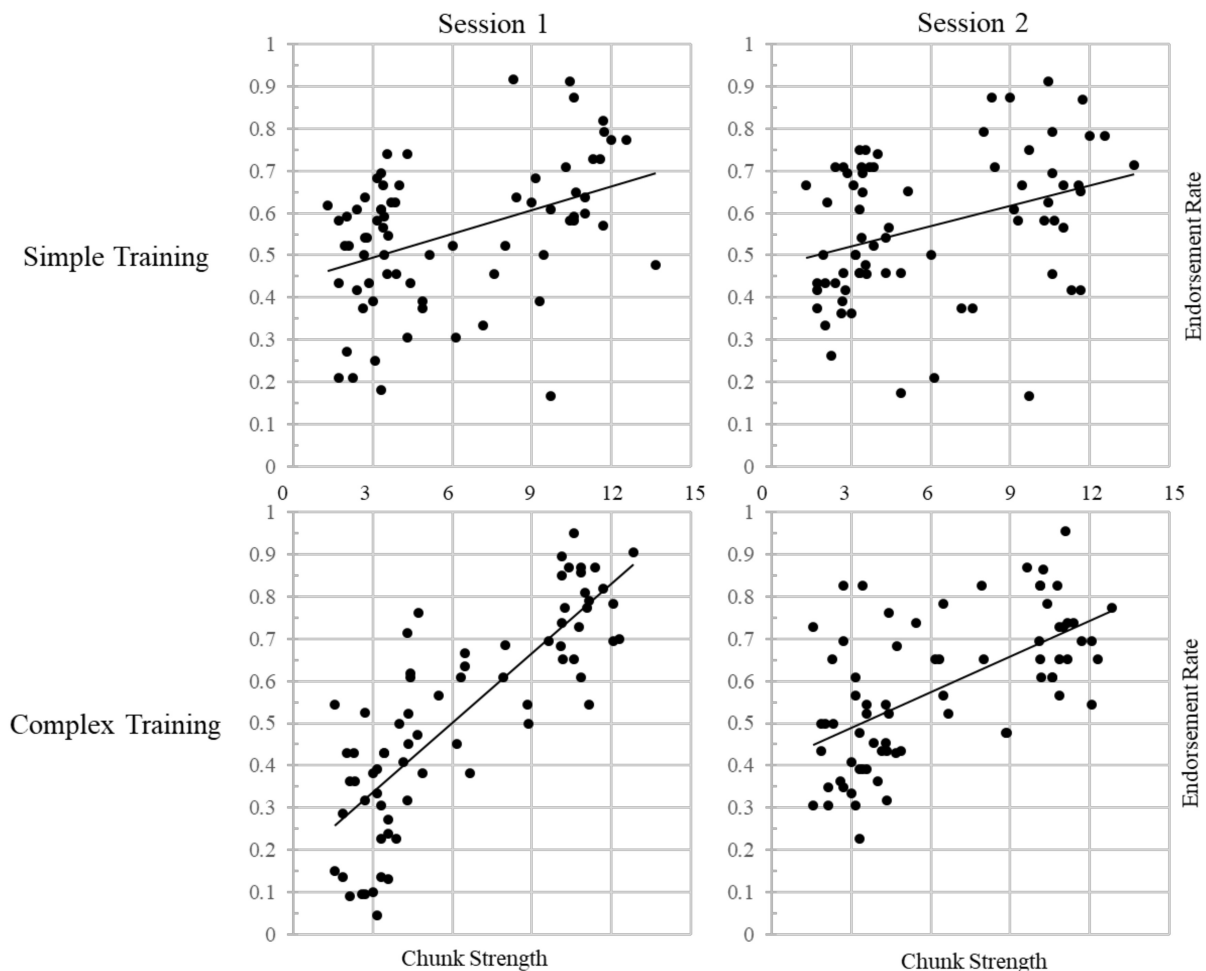
	Simple training		Complex training	
	Grammatical	Ungrammatical	Grammatical	Ungrammatical
CS (SD)	6.93 (3.74)	5.02 (3.40)	7.34 (3.79)	5.36 (3.20)
95% CI	5.67–8.20	3.87–6.17	6.05–8.62	4.27–6.44

Mean chunk strength of grammatical and ungrammatical test items for each training condition, along with standard deviations in parentheses. 95% Confidence Intervals are reported beneath each mean. CS, chunk strength.

**TABLE 7 |** Summaries of the two generalized linear mixed effects models.

Fixed Effects	Model 1	Model 2	Model 3
Intercept	−0.847*** (0.134)	−1.961*** (0.217)	−2.527*** (0.268)
Group	0.095 (0.137)	1.093*** (0.229)	2.100*** (0.357)
Chunk	0.126*** (0.007)	0.281*** (0.025)	0.374*** (0.036)
Time	0.180*** (0.052)	0.682*** (0.118)	1.05*** (0.155)
Group*Chunk		−0.110*** (0.015)	−0.279*** (0.048)
Group*Time		−0.227* (0.105)	−0.887*** (0.207)
Chunk*Time		−0.064*** (0.015)	−0.125*** (0.022)
Group*Chunk*Time			0.111*** (0.030)
<b>Random Effects</b>			
Subject (Intercept)	0.189 (0.435)	0.192 (0.438)	0.193 (0.439)
<b>Goodness Of Fit</b>			
Log likelihood	−4231.2	−4192.9	−4186.0
AIC	8472.5	8401.8	8390.1
BIC	8506.4	8456.0	8451.1

Estimated coefficients are listed while standard errors are reported in parentheses. \* $p < 0.05$ , \*\*\* $p < 0.001$ .



**FIGURE 4 |** Endorsement rates correlated with chunk strength across sessions for each training group, illustrating the three-way interaction effect in Model 3. Trend lines represents linear lines of best fit.

term is solely responsible for the improvement in the model's fit due to the inclusion of all three two-way interaction terms in Model 2. Visual inspection of **Figure 4** demonstrates this interaction nicely, showing that the effect of chunk strength on item endorsement decreases over time and illustrating the greater impact of chunk strength on endorsement for participants in the complex training condition; additional correlational analyses will attempt to verify the directionality of the interaction. Note that including Item as a random effect resulted in a model that failed to converge when also including the critical three-way interaction.

With the aim of extending the GLMM's findings, we also chose to examine the ways in which accuracy and endorsement varied depending on the surface-level features of each test item at both sessions within either training group. In order to do so, we conducted subsequent analyses on by-items data rather than collapsing across participants. As described in the methods, this meant that the twenty-three participants in the complex training condition and 24 in the simple training condition constituted the number of observations across the seventy-two

test items, and due to the differing fragment statistics for each training condition, all subsequent analyses treated these groups separately.

To further explore the results of the GLMMs, traditional, frequentist analyses were conducted. Both training groups exhibited a correlation between an item's chunk strength and their endorsement rate. Notably, while the simple training group showed small to moderate correlations at both sessions one ( $r = 0.409$ ,  $p < 0.001$ ) and two ( $r = 0.342$ ,  $p = 0.003$ ), the complex training group showed an extremely strong correlation at session one ( $r = 0.819$ ,  $p < 0.001$ ), as well as a moderately strong correlation at session two ( $r = 0.598$ ,  $p < 0.001$ ), suggesting that this pattern drove the three-way interaction above. A Fisher's  $r$  to  $z$  comparison of these correlation coefficients shows that the two groups' correlations are significantly different from one another at both sessions one ( $z = -4.23$ ,  $p < 0.001$ ) and two ( $z = -1.96$ ,  $p = 0.05$ ).

To verify the validity of these contrasts, we examined whether there was inherently a stronger relationship between the test items' chunk strength and their grammaticality for

the complex group than for the simple group. If that were the case, then the meaningfulness of the difference between the groups' correlations would be reduced – it would have just been the case that for one group these two variables tracked one another more closely and was not driven by the differential effects of their training. However, this was not the case, as the mean chunk strength of grammatical items was not significantly different between the simple and complex training conditions [ $t(70) = -0.456$ ,  $p = 0.649$ ,  $d = 0.11$ ], a pattern that also held true for ungrammatical items [ $t(70) = -0.429$ ,  $p = 0.670$ ,  $d = 0.10$ ]. Refer back to **Table 6** to find the relevant means, standard deviations, and confidence intervals. This shows that chunk strength was not a stronger cue for either group of participants, suggesting that the complex group's reliance on it was not merely because it was more useful for them in terms of differentiating grammatical and ungrammatical items at test.

A key difference between training groups also emerged when looking at how the chunk strength of each item correlated with participants' accuracy when judging the grammaticality of that item. Only participants in the complex training condition showed a statistically significant relationship between accuracy and chunk strength, and they did so at session one ( $r = 0.300$ ,  $p = 0.010$ ), as well as at session two ( $r = 0.248$ ,  $p = 0.035$ ), whereas those in the simple training condition did not at either session one ( $r = 0.187$ ,  $p = 0.116$ ) or session two ( $r = 0.139$ ,  $p = 0.244$ ). This underscores the complex training group's reliance on the surface level properties of the test stimuli when engaged in the GJT.

## DISCUSSION

The set of results described above demonstrates that first, learners overall seem to be able to retain the regularities of an artificial language over the span of 2 weeks. While retention was not perfect, as performance degraded over time, a sufficient degree of knowledge was maintained to show a learning effect at the second test session. This is a longer time interval than what is typically found in the extant literature on SL, which typically only looks at retention after a period of hours or days. Extensive research on other types of learning and memory has found that participants can recall learned items at rather long intervals (Tulving et al., 1982; Schacter, 1987; Roediger, 1990; Mitchell, 2006). Note that the test items in this study were not present during training and were only seen once previously during a test session using a randomized presentation, where half of the trials were foils. This suggests that instead of recalling previous answers, participants were able to use learned knowledge to respond to test items.

The ability of participants to retain their knowledge of statistically learned dependencies over time is crucial to understanding the way in which experience with linguistic constructions affects later processing (Real and Christiansen, 2007; Wells et al., 2009). In order for SL to impact language processing in the way it has long been hypothesized (Saffran, 2001), the learned statistical patterns must be retained in memory. Our findings demonstrate that such retention is possible and adds support for such theories. Determining the limits of

retention for statistically learned regularities should be a priority for future research, as the SL literature has long rested on the assumption that such associations form a key foundation for language learning.

There is an interesting pattern of results that speak to both the SL literature (directly above) and the “starting small” theories. Firstly, the fact that those trained extensively on simple items exhibited above-chance accuracy performance at both sessions provides evidence that “starting small” with extensively scaffolded, staged training leads to more accurate learning and retention of grammatical regularities – their performance did not show a statistically significant decline between sessions. Whereas both training conditions within the present study started small, participants in the simple training condition were given significantly more time to learn from the simpler items. Intentionally reducing the problem space for learners during the early phases of acquisition seemed to improve learning outcomes in this study (see also Conway et al., 2003). Poletiek et al. (2018) have recently demonstrated that participants are able to use their memory of previously encoded, simple structures to facilitate their learning of newer, more complex ones. They also point out the importance of incrementally exposing learners to increasingly complex items, rather than simply longer ones.

The present research also shows a similar trend to other studies that demonstrate how overrepresenting simplified input early on during training can lead to improved learning (Pine, 1994; Perfors et al., 2011). Scaffolding reflects the way in which young learners typically acquire language, however, the results here suggest that forcing adults to adopt more immature strategies when learning a novel language may confer benefits. Future research into the relationship between second language learning in adults and intentionally constrained input could be important to shaping adult pedagogical strategies and our understanding of language acquisition more generally.

Conversely, participants in the complex training condition showed above-chance accuracy on grammaticality judgments in session one, yet they did not match the performance of the simple training group. However, what is interesting is that participants in the complex training condition showed evidence of relying more heavily on chunk strength, which captures basic frequency information. While this suggests that the complex training condition promoted simple learning of frequency patterns, it may not have been enough exposure for participants to induce the more complex probabilistic patterns underlying grammatical regularities. Future research may increase the length of exposure to complex stimuli to investigate if this does improve overall performance.

The overall set of findings fits in well with recent proposals about how the constraints placed on learning by our cognitive abilities shape the way in which we process, and thereby learn, language (Christiansen and Chater, 2008, 2016). The proposed “Now-or-Never bottleneck” refers to the process by which language users must continuously recode and compress linguistic input in order to keep up with comprehension. In this framework, language processing *is* language learning; during comprehension, we must effectively process the input as quickly and accurately



as possible before it is overwritten or interfered with by new incoming information. Learners take the information that makes it through the bottleneck as far as they can – in the simple training condition of the present study, more exposure to simple items may have allowed them to process subregularities more efficiently and thereby better deal with similar patterns in the more complex items, whereas those in the complex condition were only able to rely on the more surface-level information contained within the chunks that they learned and retained.

In sum, participants in this study showed the ability to retain information learned within an artificial language learning paradigm over the course of 2 weeks. It also appears that increasing exposure to simplified grammatical structures in beginning stages of learning confers benefits to adult learners. Importantly, some of the grammatical regularities of this artificial language are retained in long-term memory in a way that has not been shown previously in SL research. This falls in line with theories about both first- and second-language acquisition, and also with new ideas concerning the role of processing constraints on language learning. Overly challenging and complex input seems to derail learners and affects the kind of information they are sensitive to, leading them to rely more on simple fragment frequency rather than higher-order associations between them. This pattern of results contrasts with learners who were provided scaffolded input, as they demonstrated better acquisition of the higher-order regularities and relied less on basic frequency cues when choosing to endorse items as grammatical or ungrammatical.

## REFERENCES

- Altmann, G. T. M. (2002). Statistical learning in infants. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15250–15251.
- Arciuli, J., and Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neurosci. Lett.* 517, 133–135. doi: 10.1016/j.neulet.2012.04.045
- Armstrong, B. C., Frost, R., and Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philos. Trans. R. Soc. B* 372:20160047. doi: 10.1098/rstb.2016.0047
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Brill-Schuetz, K. A. (2016). *The Complexity of Input and How It Contributes to Adult Second Language Acquisition*. Ph.D. thesis, University of Illinois at Chicago, Chicago, IL.
- Cameron-Faulkner, T., Lieven, E., and Tomasello, M. (2003). A construction based analysis of child directed speech. *Cogn. Sci.* 27, 843–873. doi: 10.1016/j.cogsci.2003.06.001
- Casenhiser, D., and Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Dev. Sci.* 8, 500–508. doi: 10.1111/j.1467-7687.2005.00441.x
- Christiansen, M. H. (2019). Implicit-statistical learning: a tale of two literatures. *Top. Cogn. Sci.* 11, 468–481. doi: 10.1111/tops.12332
- Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–508. doi: 10.1017/S0140525X08004998
- Christiansen, M. H., and Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* 39:e62. doi: 10.1017/S0140525X1500031X
- Christiansen, M. H., Conway, C. M., and Onnis, L. (2012). Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Lang. Cogn. Process.* 27, 233–256. doi: 10.1080/01690965.2011.606666
- Conway, C. M., Ellefson, M. R., and Christiansen, M. H. (2003). “When less is less and when less is more: starting small with staged input,” in *Proceedings*

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UIC Institutional Review Board (protocol 2008–0496). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

KB-S and KM-S conceived of the original study. KB-S collected the data. EJ, KB-S, KM-S, and MC conceived of the current analysis, analyzed the data, and revised the manuscript. EJ wrote the initial draft.

## ACKNOWLEDGMENTS

The data reported in this study was originally collected as part of the doctoral dissertation by Brill-Schuetz (2016). A special thank you goes to Rex Dayola and Mallory Webber for their essential contributions to this study.

*of the 25th Annual Conference of the Cognitive Science Society*, New York, NY, 270–275.

- Durrant, S. J., Taylor, C., Cairney, S., and Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia* 49, 1322–1331. doi: 10.1016/j.neuropsychologia.2011.02.015
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402\_1
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99. doi: 10.1016/0010-0277(93)90058-4
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 529–534. doi: 10.1073/pnas.012611199
- Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* 19, 117–125. doi: 10.1016/j.tics.2014.12.010
- Frost, R. L. A., and Monaghan, P. (2017). Sleep-driven computations in speech processing. *PLoS One* 12:e0169538. doi: 10.1371/journal.pone.0169538
- Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., et al. (2014). Sleep underpins the plasticity of language production. *Psychol. Sci.* 25, 1457–1465. doi: 10.1177/0956797614535937
- Gomez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160054. doi: 10.1098/rstb.2016.0054
- Isbilen, E. S., and Christiansen, M. H. (2018). Chunk-based memory constraints on the cultural evolution of language. *Topics Cogn. Sci.* doi: 10.1111/tops.12376 [Epub ahead of print].
- Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0

- Kersten, A. W., and Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *J. Mem. Lang.* 44, 250–273. doi: 10.1006/jmla.2000.2751
- Kim, R., Seitz, A., Feenstra, H., and Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neurosci. Lett.* 461, 145–149. doi: 10.1016/j.neulet.2009.06.030
- Knowlton, B. J., and Squire, L. R. (1994). The information acquired during artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 79–91. doi: 10.1037/0278-7393.20.1.79
- Kobor, A., Janacek, K., Takacs, A., and Nemeth, D. (2017). Statistical learning leads to persistent memory: evidence for one-year consolidation. *Sci. Rep.* 7, 760. doi: 10.1038/s41598-017-00807-3
- Lai, J., and Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition* 118, 265–273. doi: 10.1016/j.cognition.2010.11.011
- Loewen, S. (2009). “Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge,” in *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, eds R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, and H. Reinders, (Bristol: Multilingual Matters), 94–112. doi: 10.21832/9781847691767-006
- Mitchell, D. B. (2006). Nonconscious priming after 17 years. *Psychol. Sci.* 17, 925–929. doi: 10.1111/j.1467-9280.2006.01805.x
- Morgan-Short, K. (in press). Insights into the neural mechanisms of becoming bilingual: a brief synthesis of second language research with artificial linguistic systems. *Bilingualism: Language and Cognition*.
- Morgan-Short, K. (2007). *A Neurolinguistic Investigation of Late-Learned Second Language Knowledge: The Effects of Explicit and Implicit Conditions*. Unpublished dissertation, Georgetown University, Washington, DC.
- Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., and Wong, P. C. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism Lang. Cogn.* 17, 56–72. doi: 10.1017/s1366728912000715
- Morgan-Short, K., Finger, I., Grey, S., and Ullman, M. T. (2012a). Second language processing shows increased native-like neural responses after months of no exposure. *PLoS One* 7:e32974. doi: 10.1371/journal.pone.0032974
- Morgan-Short, K., Steinhauer, K., Sanz, C., and Ullman, M. T. (2012b). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *J. Cogn. Neurosci.* 24, 933–947. doi: 10.1162/jocn\_a\_00119
- Morgan-Short, K., Sanz, C., Steinhauer, K., and Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: an event-related potential study. *Lang. Learn.* 60, 154–193. doi: 10.1111/j.1467-9922.2009.00554.x
- Newport, E. L. (1990). Maturation constraints on language learning. *Cogn. Sci.* 14, 11–28. doi: 10.1016/0364-0213(90)90024-q
- Newport, E. L. (2016). Statistical language learning: computational, maturational, and linguistic constraints. *Lang. Cogn.* 8, 447–461. doi: 10.1017/langcog.2016.20
- Norris, J. M., and Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Lang. Learn.* 50, 417–528. doi: 10.1111/0023-8333.00136
- Perfors, A., Tenenbaum, J. B., and Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition* 118, 306–338. doi: 10.1016/j.cognition.2010.11.001
- Pine, J. M. (1994). “The language of primary caregivers,” in *Input and Interaction in Language Acquisition*, eds C. Gallaway, and B. J. Richards, (Cambridge: Cambridge University Press), 109–149.
- Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., and Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Lang. Res.* doi: 10.1177/0267658319828413
- Poletiek, F., Conway, C. M., Ellefson, M. R., Lai, J., Bocanegra, B. R., and Christiansen, M. H. (2018). Under what conditions can recursion be learned? Effects of starting small in artificial grammar learning of recursive structure. *Cogn. Sci.* 42, 2855–2889. doi: 10.1111/cogs.12685
- Real, F., and Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *J. Mem. Lang.* 57, 1–23. doi: 10.1016/j.jml.2006.08.014
- Roediger, H. L. (1990). Implicit memory. Retention without remembering. *Am. Psychol.* 45, 1043–1056. doi: 10.1037/0003-066X.45.9.1043
- Rohde, D. L. T., and Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition* 72, 67–109. doi: 10.1016/S0010-0277(99)00031-1
- Romano, J. C., Howard, J. H., and Howard, D. V. (2010). One-year retention of general and sequence-specific skills in a probabilistic, serial reaction time task. *Memory* 18, 427–441. doi: 10.1080/09658211003742680
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *J. Mem. Lang.* 44, 493–515. doi: 10.1006/jmla.2000.2759
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Schacter, D. L. (1987). Implicit memory: history and current status. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 501–518. doi: 10.1037/0278-7393.13.3.501
- Siegelman, N., and Arnon, I. (2015). The advantage of starting big: learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *J. Mem. Lang.* 85, 60–75. doi: 10.1016/j.jml.2015.07.003
- Singer, J. D., and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.
- Smalle, E. H., Page, M. P., Duyck, W., Edwards, M., and Szmalec, A. (2018). Children retain implicitly learned phonological sequences better than adults: a longitudinal study. *Dev. Sci.* 21:e12634. doi: 10.1111/desc.12634
- Smalle, E. H. M., Bogaerts, L., Simonis, M., Duyck, W., Page, M. P. A., Edwards, M. G., et al. (2016). Can chunk size differences explain developmental changes in lexical learning? *Front. Psychol.* 6:1925. doi: 10.3389/fpsyg.2015.01925
- Tulving, E., Schacter, D. L., and Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 8, 336–342. doi: 10.1037/0278-7393.8.4.336
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., and MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cogn. Psychol.* 58, 250–271. doi: 10.1016/j.cogpsych.2008.08.002

**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.

Copyright © 2019 Jost, Brill-Schuetz, Morgan-Short and Christiansen. 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.



# Do Changes in Language Context Affect Visual Memory in Bilinguals?

Scott R. Schroeder\*

Department of Speech-Language-Hearing Sciences, Hofstra University, Hempstead, NY, United States

## OPEN ACCESS

### Edited by:

Vitória Piai,  
Radboud University Nijmegen,  
Netherlands

### Reviewed by:

Angela De Bruin,  
Basque Center on Cognition, Brain  
and Language, Spain  
Greg Poarch,  
University of Münster, Germany

### \*Correspondence:

Scott R. Schroeder  
scott.r.schroeder@hofstra.edu

### Specialty section:

This article was submitted to Speech  
and Language, a section of the  
journal *Frontiers in Human  
Neuroscience*

**Received:** 30 June 2019

**Accepted:** 27 September 2019

**Published:** 17 October 2019

### Citation:

Schroeder SR (2019) Do Changes in  
Language Context Affect Visual  
Memory in Bilinguals?  
*Front. Hum. Neurosci.* 13:364.  
doi: 10.3389/fnhum.2019.00364

Language is often present when people are encoding visual memories. For bilinguals, this language context can have different forms (i.e., Language A, Language B, or both Language A and B), and can change over the course of events. The current study examined whether a change in language context during a visual event or between visual events affects a bilingual's ability to remember visual information. English-Spanish bilinguals and control participants encoded three lists of novel shapes amid different task-irrelevant language contexts. Following each list, participants completed a free recall test in which they drew the novel shapes they remembered. Results indicated that a change in language context between events, but not during events, affected visual memory. Specifically, the switch in language context between the second and third event (such as an English context in list 2 switching to a Spanish context in list 3) produced a reliable memory advantage for the English-Spanish bilinguals (relative to the control participants). The results offer preliminary evidence that task-irrelevant language context can influence a bilingual's ability to remember non-linguistic information, as well as further evidence for context effects and multi-sensory effects in memory.

**Keywords:** language, memory, bilingualism, context, multisensory

## INTRODUCTION

Considering that we likely hear upwards of hundreds of words per hour (Hart and Risley, 2003), much of our encoding of visual memories occurs in the context of spoken language. For most people, this spoken language context consists of one or both of their two languages, as bilingualism is the norm worldwide (Grosjean, 2015). Despite the prevalence and despite considerable research on bilingual memory (e.g., Marian and Neisser, 2000; Pu and Tse, 2014; Basnight-Brown and Altarriba, 2016; Heredia and Cieśllicka, 2019), there appears to be no published research examining whether a bilingual's language context influences visual memory (or, more broadly, memory for non-linguistic information). The current study provides an initial examination, by assessing whether a change in language context during or between events influences English-Spanish bilinguals' ability to remember novel shapes.

As a framework to guide us, we can draw upon research from the event processing literature (for reviews, see Zacks and Swallow, 2007; Kurby and Zacks, 2008; Radvansky and Zacks, 2017). According to event processing work (such the Event Horizon Model and Event Segmentation Theory), people parse their ongoing experiences into events and subevents. In other words, people segment continuous activity (e.g., watching a movie or going to the grocery store) into parts and subparts. Possibly, event segmentation could be assisted by a change in language context (for example, a switch from English into Spanish). Specifically, a change in language context *within*

an event could facilitate segmentation into subevents (i.e., ending one subevent and starting a new subevent), and a change in language context *between* events could facilitate segmentation into separate events (i.e., ending one event and starting a new event; see **Figure 1** below for a depiction of these two possibilities). As event segmentation is highly correlated with later event memory (Kurby and Zacks, 2008; Swallow et al., 2009; Sargent et al., 2013), changes in language context over the course of events could affect (even improve) event memory. These two possibilities—i.e., within-event and between-event changes in language context affecting visual event memory—are now fleshed out, with the within-event change discussed first and the between-event change discussed second.

How could a change in language context *within* an event boost visual memory? In other words, why might visual memory be enhanced when there is a switch in language context at one or more points during the course of a visual event? Theories of event processing, such as the Event Horizon Model, posit that more subevents (each with fewer elements) will lead to better memory than fewer subevents (each with more elements; Radvansky, 2012; Pettijohn et al., 2016). Consistent with this hypothesis, recent work found that memory (e.g., remembering word lists) was improved when the encoding event contained a segmenting cue (e.g., walking through a doorway from one room to another, or closing and then opening a computer window), which divided the encoding event into subevents (Pettijohn et al., 2016). Furthermore, more segmenting cues (and thus more subevents) led to better memory. The explanation is that segmenting an event into subevents provides an organizational structure and breaks the to-be-remembered information into smaller (and thus easier to remember) chunks. Following from this research, the first hypothesis assessed in the current study is that changes in language context *within* a visual event (thereby creating subevents) will enhance a bilingual's visual memory.

It is also possible that a change in language context *between* visual events could boost visual memory (the second hypothesis). This hypothesis is rooted in the well-established finding that, when multiple similar events occur over time, there can be proactive interference from the first event to subsequent events, leading to worse memory for subsequent events (than would otherwise be observed; Postman et al., 1968; Anderson and Neely, 1996; Kane and Engle, 2000). In other words, if the first event and second event are similar, memory for the second event can be hindered. However, evidence suggests that proactive interference can be reduced if there is a change in context between the first event and the second event (or subsequent events; Sahakyan and Kelley, 2002; Pastötter and Bäuml, 2007; Bäuml and Kliegl, 2013). With the first and second events having distinct contexts, presumably the two events become easier to differentiate, thereby reducing competitive interference. Extending this line of thinking to language contexts in bilinguals, if the language context (e.g., English) in the first event differs from the language context (e.g., Spanish) in the second event, then proactive interference from the first event to the second event may be diminished. There is some evidence consistent with this hypothesis, as studies assessing word-list memory in bilinguals have shown that a language switch between lists

(e.g., list 1 consisting of English words and list 2 consisting of Spanish) can reduce proactive interference and improve memory for list 2 (Goggin and Wickens, 1971; Dillon et al., 1973; Francis, 1999). The current study extends this work on verbal memory to visual memory, by examining whether a change in language context between events can enhance a bilingual's visual memory (the second hypothesis of the study).

As an initial exploration of potential language context effects on visual memory in bilinguals, the current study asked English-Spanish bilinguals and controls (i.e., participants who did not know both English and Spanish) to encode three lists of novel shapes, with a free recall drawing test following each list. During the visual encoding of the shapes, a language context was present (i.e., each new shape was introduced with the phrase “this drawing looks like this” in either English or Spanish). The language context was entirely in English for one list of shapes (i.e., the English-Only Context), entirely in Spanish for a second list (i.e., the Spanish-Only Context), and partly in English and partly in Spanish for a third list (i.e., the English-Spanish Context), with the order of these lists counterbalanced across participants.

The first hypothesis—i.e., that a change in language context *within* an event might boost a bilingual's visual memory—predicted better recall performance for the English-Spanish bilinguals in the English-Spanish Context (as this context involved a within-event language change from English to Spanish to English) relative to the English-Only and Spanish-Only Contexts (as these contexts did not involve a within-event language change) and relative to the control participants (who had a less-comprehensible change in language context). The second hypothesis—i.e., that a change in language context *between* visual events might boost a bilingual's visual memory—predicted reduced proactive interference from the first list of shapes to the subsequent lists of shapes (i.e., lists 2 and 3) for the English-Spanish bilinguals, because the paradigm entailed a change in language context between each list (i.e., from list 1 to list 2 to list 3). This hypothesis might thus result in the English-Spanish bilinguals having a non-significant decline in recall performance from list 1 to subsequent lists (lists 2 and 3), as well as better performance on lists 2 and 3 relative to control participants (who had a less-comprehensible change in language context). The control participants may also benefit from a change in language context within and between events, despite not knowing one of the languages, though this benefit may be smaller.

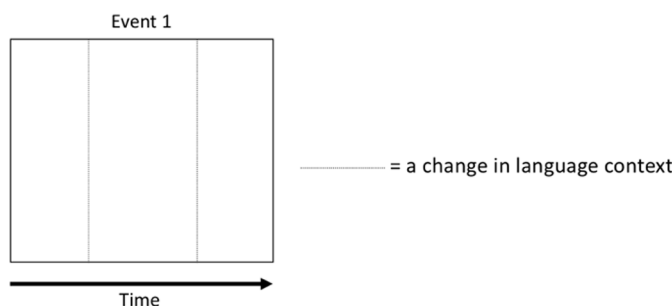
## MATERIALS AND METHODS

### Participants

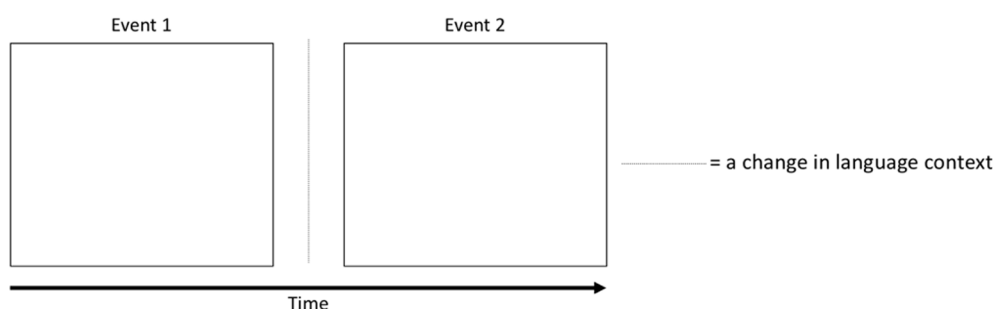
Seventy young adults (mean age = 20.41 years; gender = 41 female, 29 male) were included. Participants consented to participate in the experimental protocol, and the protocol was approved by the ethics board at Hofstra University. Participants were categorized as English-Spanish Bilinguals ( $N = 38$ ) or Controls ( $N = 32$ ) based on a post-experiment questionnaire completed by participants. If participants listed both English and Spanish as languages they have knowledge



### A Change in Language Context *WITHIN* an Event



### A Change in Language Context *BETWEEN* Events



**FIGURE 1 |** How a change in language context could affect event segmentation. This image illustrates scenarios in which a change in language context could facilitate the parsing of an experience into events and subevents. The top panel (“A Change in Language Context *WITHIN* an Event”) depicts a situation in which there are two changes in language context during an event (e.g., from English to Spanish and then from Spanish to English), thereby creating three subevents (i.e., a beginning, middle, and end) within the larger event. The bottom panel (“A Change in Language Context *BETWEEN* Events”) depicts a situation in which there is a change in language context in between two larger events, such that one language context was present during Event 1 and a different language context was present during Event 2 (e.g., English in Event 1 and Spanish in Event 2).

of, then they were placed into the English-Spanish Bilingual group; if they did not list both English and Spanish, then they were placed into the Control group (this atypical group assignment process, in which participants were *post hoc* assigned to groups rather than in the recruitment phase, was chosen because it reduced the likelihood that participants and experimenters would guess their group and the purpose of the experiment).

Demographic information on the English-Spanish Bilinguals and Controls is provided in **Table 1**. Specifically, **Table 1** provides descriptive statistics for Age, Gender, English Proficiency (self-rated receptive proficiency in English on a 0-low to 10-perfect scale), English AoA (age at which participant first started learning English), English Use (% of time participant currently uses English when they are using a spoken language), Spanish Proficiency (self-rated receptive proficiency in Spanish on a 0-low to 10-perfect scale), Spanish AoA (age at which participant first started learning Spanish), Spanish Use (% of time participant currently uses Spanish when they are using a spoken language), and L2 Proficiency (self-rated receptive proficiency on a 0-low to 10-perfect scale in their second most proficient language). Note that L2 Proficiency was set to 0 in participants who did not list a second language.

**TABLE 1 |** Participant demographic information.

	English-Spanish Bilinguals ( <i>N</i> = 38) Mean (SD)	Controls ( <i>N</i> = 32) Mean (SD)	Comparison
Age	20.45 (1.25)	20.38 (1.43)	$t_{(68)} = 0.22, p = 0.82$
Gender	25 Female, 13 Male	16 Female, 16 Male	$\chi^2_{(1)} = 1.79, p = 0.18$
English proficiency	9.50 (0.69)	9.58 (0.72)*	$t_{(67)} = 0.47, p = 0.64$
English AoA	1.55 (2.39)	2.11 (3.78)*	$t_{(67)} = 0.75, p = 0.46$
English use	87.29 (15.98)	87.74 (18.27)*	$t_{(67)} = 0.11, p = 0.91$
Spanish proficiency	4.57 (2.75)	—	—
Spanish AoA	10.74 (4.93)	—	—
Spanish use	5.74 (9.53)	—	—
L2 proficiency	6.16 (2.72)	5.40 (3.38)*	$t_{(67)} = 1.03, p = 0.31$

\*Note that Language Proficiency, AoA, and Use data are missing for one Control participant.

For the 38 English-Spanish Bilinguals, English was the L1 for 33 of the participants, the L2 for 4, and the L3+ for 1, whereas Spanish was the L1 for 2 of the participants, the L2 for 19, and the L3+ for 17. Note that L1, L2, and L3+ designations were determined by proficiency. In the case of a tie in proficiency, AoA was used to break the tie; if the tie remained, English was

given priority because of the English-dominant context in which the participants resided. Of the 38 English-Spanish Bilinguals, 13 only listed English and Spanish as languages they know, whereas 25 listed more than English and Spanish.

Among the 32 Controls, five only listed English, whereas 27 listed more languages than English. English was the L1 for 28 of the Controls, the L2 for 3, and the L3+ for 1. The Controls listed a wide variety of languages, including Romance languages (i.e., a language that derived from Latin, such as French, Portuguese, Italian, Romanian, and Catalan). Nineteen of the Controls listed knowledge of a Romance language. Possibly, knowledge of a Romance language could lead to partial understanding of the Spanish that was heard in the experiment, thereby affecting recall performance. However, the Romance Controls and Non-Romance Controls patterned similarly on the key finding reported in the Results section below. That is, both groups showed a clear proactive interference pattern (recall percentage for Romance Controls: List 1 = 53%, List 2 = 48%, List 3 = 41%; recall percentage for Non-Romance Controls: List 1 = 49%, List 2 = 45%, List 3 = 42%).

Note that the use of the term “bilingualism” in the current study is based on the inclusive and minimalist definition by Mackey (1962): a bilingual is a person who has “the ability to use more than one language.” However, given that bilingualism is a label with “open-ended semantics” (Baetens Beardsmore, 1982) and many definitions, some readers of the current study might have a stricter definition of bilingualism; in such a case, “second language learners” would be a more suitable label for the English-Spanish speakers.

## Procedure

The memory task involved encoding three lists of novel shapes, with a drawing-based free recall test following each list. Each of the three lists had a different language context: (1) English-Only Context; (2) Spanish-Only Context; and (3) English-Spanish Context. The order of the three language contexts was counterbalanced across participants. Participants were told not to be concerned with the language context (rather, they should be concerned with memorizing the novel shapes). The specifics of encoding and retrieval are provided below.

## Encoding

In each of the three encoding lists, participants were presented with 12 drawings of novel shapes. Three sets of 12 drawings were created for the experiment, and each set appeared with each of the three lists (List 1, List 2, List 3) and with each of the three language contexts (English-Only Context, Spanish-Only Context, English-Spanish Context) a similar number of times.

The language context was created by introducing each novel shape with the phrase: “This drawing looks like this” (the phrase in Spanish is “Este dibujo se ve así.”). Both the English and Spanish phrase were recorded in the same voice, by an English-Spanish bilingual speaker. In the English-Only Context, the English phrase was used to introduce each (and every) shape, and likewise, in the Spanish-Only Context, the Spanish phrase was used to introduce each (and every) shape. In the English-Spanish Context, the English phrase was used to introduce shapes

1 through 4, the Spanish phrase was used to introduce shapes 5 through 8, and the English phrase was used to introduce shapes 9 through 12. There were thus two language switches rather than one, because previous work indicated that two segmenting cues (and thus three subevents) led to better memory than one segmenting cue (and thus two subevents; Pettijohn et al., 2016).

The phrase “This drawing looks like this” was heard during the 4,000 ms interstimulus interval before each shape appeared. After the 4,000 ms interstimulus interval, the shape appeared in the center of the screen for 3,000 ms. **Figure 2**, on the left side, provides a visual depiction of the encoding process.

Before each encoding list, participants were instructed to learn the novel shapes for a memory test but not to be concerned with the language context. Specifically, the instructions were: “You are going to see a string of abstract images. Try to remember as many images as you can; you will be asked to draw them after the recording ends. You will also hear someone talking in English or Spanish; you can ignore the person talking, as you will not have to remember what they said or what language they were using.”

## Retrieval

Immediately after each encoding list, participants were asked to perform a free recall test in which they were to draw all of the shapes they remembered from the preceding list of 12 shapes. The recall sheet that was used can be seen on the right side of **Figure 2**. The recall sheet reads: “Please draw all of the images that you remember. You will not use all of the boxes.” Participants were given an unlimited amount of time to complete the free recall test.

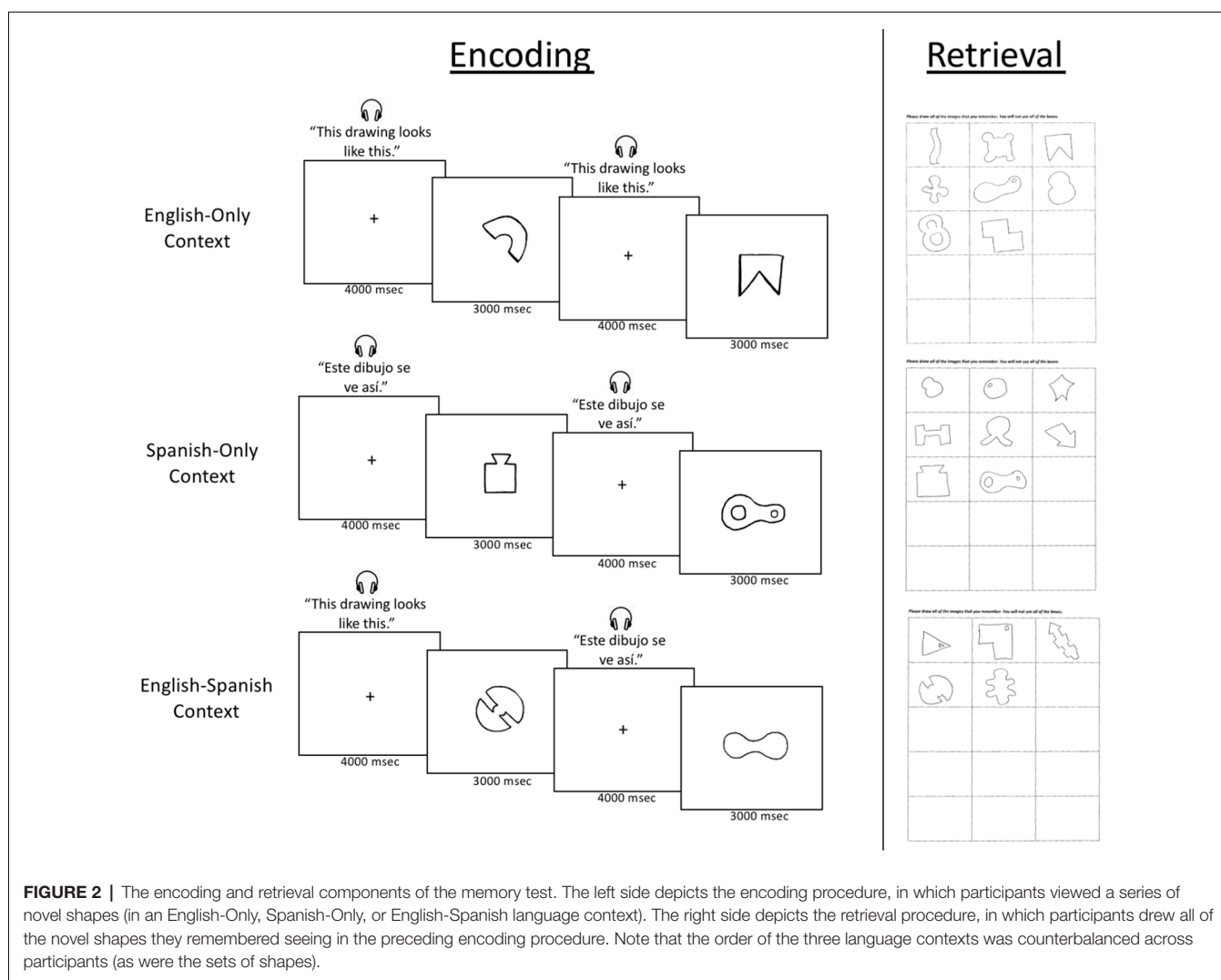
After completing the memory task, participants filled out the post-experiment language and demographic questionnaire. The memory task data were coded by a research assistant who was blind to the participant’s group and to the purpose of the experiment. The coding entailed matching each of the participant’s drawings to one of the drawings in the target list. If the drawing could be exclusively matched to a single drawing in the target list (even if some minor details were missing), the participants received credit for that drawing; however, if the drawing could not be matched to any drawings, could be matched to multiple drawings, or could be matched to a drawing in a non-target list, the participant did not receive credit.

Consistent with the values of open science, the raw visual memory data, the recall scores, and the relevant questionnaire data are freely available to the public through the Open Science Framework at the following web address: <https://osf.io/f5wrh/>. All other information and stimuli will be willingly provided by the author.

## RESULTS

### A Change in Language Context Within an Event

To assess the *first hypothesis* (i.e., that a change in language context *within* an event might help visual memory in bilinguals), we can compare recall when there was a change in language context within an event (i.e., the English-Spanish Context) to recall when there was not (i.e., the English-Only Context and



the Spanish-Only Context) in the English-Spanish Bilinguals and the Controls. These data are displayed in **Figure 3** below. As can be seen in **Figure 3**, the mean percentage of shapes recalled for the English-Spanish Bilinguals was *not* higher when there was a change in language context within an event (i.e., the English-Spanish Context) vs. when there was not; in fact, the opposite pattern was seen. Furthermore, English-Spanish Bilinguals' recall when there was a within-event language context change (i.e., the English-Spanish Context) appears to be very similar to that of Controls' recall. Thus, by visual inspection, the data clearly do not support the hypothesis that a change in language context within an event helps visual memory in bilinguals.

Two statistical tests were conducted to assess the first hypothesis: a traditional ANOVA and a generalized linear mixed-effects model.

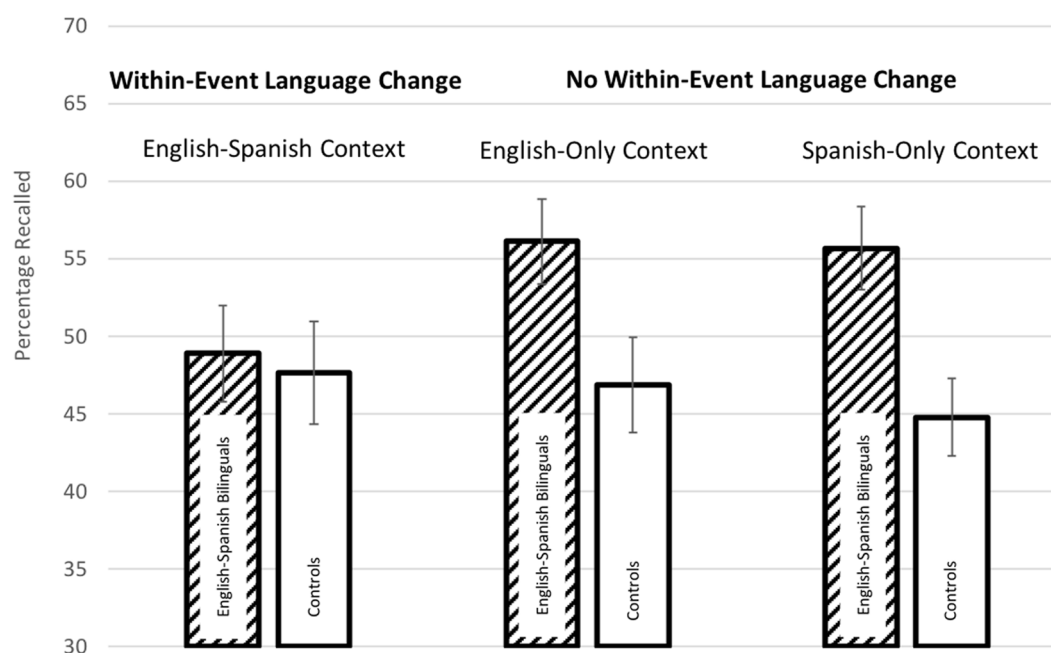
### ANOVA

The ANOVA was a  $3 \times 2$ , with Language Context (English-Spanish Context vs. English-Only Context vs. Spanish-Only Context) as a within-subjects independent variable, Group

(English-Spanish Bilinguals vs. Controls) as a between-subjects independent variable, and mean percentage of shapes recalled as the dependent variable. The ANOVA yielded a non-significant main effect of Language Context,  $F_{(2,136)} = 1.06$ ,  $p = 0.35$ , partial  $\eta^2 = 0.02$ , a significant main effect of Group,  $F_{(1,68)} = 4.59$ ,  $p = 0.04$ , partial  $\eta^2 = 0.06$ , and a non-significant interaction between Language Context and Group,  $F_{(2,136)} = 2.13$ ,  $p = 0.12$ , partial  $\eta^2 = 0.03$ . The significant main effect of Group reflects an advantage for the English-Spanish Bilinguals, but this advantage seems to be due mostly to enhanced performance when there was *no* change in language context within an event, which goes *against* the hypothesis.

### Mixed-Effects Model

A generalized linear mixed-effects model yielded similar results. The model consisted of Group and Language Context as fixed effects and Participant as a random effect (on the intercept). The model was computed using the glmer function in R, with the fixed effects sum coded, and with significance assessed through an Analysis of Deviance Table (Type III Wald chi-square



**FIGURE 3 |** Visual memory performance when there was (and was not) a change in language context within an event. This image displays visual memory performance (percentage of shapes recalled) for English-Spanish Bilinguals and Controls when there was a change in language context within an event (English-Spanish Context) and when there was no change in language context within an event (English-Only Context and Spanish-Only Context).

tests). There was a non-significant main effect of Language Context,  $\chi^2_{(2)} = 1.73$ ,  $p = 0.42$ , a significant main effect of Group,  $\chi^2_{(1)} = 4.62$ ,  $p = 0.03$ , and a trending interaction between Language Context and Group,  $\chi^2_{(2)} = 5.39$ ,  $p = 0.07$ .

## A Change in Language Context Between Events

To assess the *second hypothesis* (i.e., that a change in language context *between* events might help visual memory in bilinguals), we can compare recall when there was a change in language context between events (i.e., Lists 2 and 3) to recall when there was not (i.e., List 1) in the English-Spanish Bilinguals and the Controls. These data are shown in **Figure 4** below. A visual inspection of **Figure 4** reveals a consistent decline in recall for the Controls from List 1 to List 2 to List 3, i.e., a proactive interference effect. For the English-Spanish Bilinguals, however, the decline is less consistent, with List 3 showing the opposite of proactive interference and resulting in a noticeable difference between the English-Spanish Bilinguals and Controls. These visual impressions are partially consistent with the second hypothesis.

As with the first hypothesis, two statistical tests (i.e., a traditional ANOVA and generalized linear mixed-effects model) were used to assess the second hypothesis.

## ANOVA

The  $3 \times 2$  ANOVA had List Number (List 1 vs. List 2 vs. List 3) as a within-subjects independent variable, Group (English-Spanish Bilinguals vs. Controls) as a between-subjects

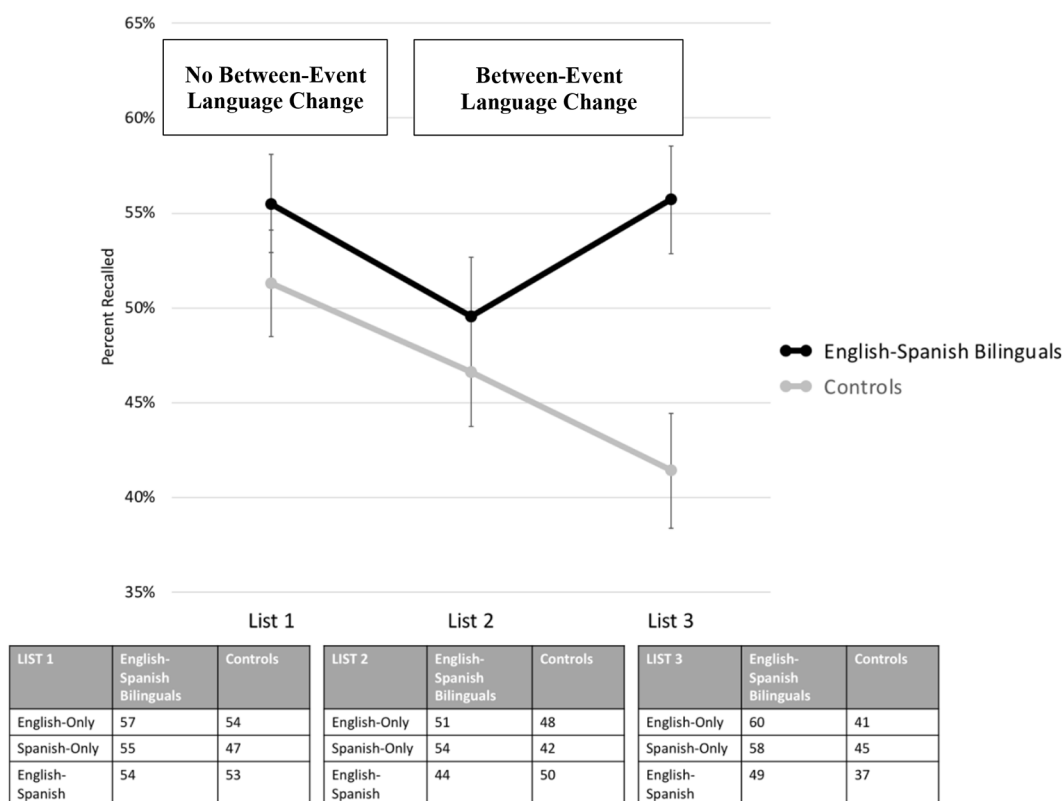
independent variable, and mean percentage of shapes recalled as the dependent variable. The ANOVA revealed a significant main effect of List Number,  $F_{(2,136)} = 3.37$ ,  $p = 0.04$ , partial  $\eta^2 = 0.05$ , reflecting a proactive interference effect from List 1 to List 2 to List 3, and a significant main effect of Group,  $F_{(1,68)} = 5.13$ ,  $p = 0.03$ , partial  $\eta^2 = 0.07$ , reflecting that English-Spanish Bilinguals had better recall overall than Controls. Crucially, there was also a significant interaction between List Number and Group,  $F_{(2,136)} = 3.79$ ,  $p = 0.03$ , partial  $\eta^2 = 0.05$ .

To follow up the interaction, Bonferroni-corrected *t*-test comparisons among lists (i.e., List 1 vs. List 2, List 1 vs. List 3, and List 2 vs. List 3) were conducted for both English-Spanish Bilinguals and Controls. The only comparison that survived correction for multiple comparisons was List 1 vs. List 3 in Controls ( $p = 0.01$ ), reflecting a significant decline in performance from List 1 to List 3 (i.e., a proactive interference effect) for Controls (but not for English-Spanish Bilinguals). The decline for the Controls in List 3, in conjunction with a reversal pattern for English-Spanish Bilinguals, appeared to create a sizable difference between groups in List 3 (but not Lists 1 and 2). To assess statistical significance, Bonferroni-corrected *t*-tests compared groups on each of the 3 lists, with the only significant difference emerging on List 3 ( $p = 0.003$ ).

## Mixed-Effects Model

Next, a generalized linear mixed-effects model was conducted, with Group and List Number as fixed effects and Participant as a random effect (on the intercept). The generalized linear mixed-effects model allows us to determine if the crucial interaction





**FIGURE 4 |** Visual memory performance when there was (and was not) a change in language context between events. This image displays visual memory performance for English-Spanish Bilinguals (top line) and Controls (bottom line) when there was no between-event language context change (i.e., List 1) and when there was a between-event language context change (i.e., Lists 2 and 3). The mean percentage of shapes recalled for the three language contexts (English-Only, Spanish-Only, and English-Spanish) are displayed at the bottom.

between Group and List Number could be replicated with a different type of analysis and with an analysis that accounts for the random effect of participants. The model was computed using the glmer function in R. The fixed effects were sum coded, and statistical significance was determined through an Analysis of Deviance Table (Type III Wald chi-square tests). The analysis yielded a trending main effect of List Number,  $\chi^2_{(2)} = 5.01$ ,  $p = 0.08$ , a significant main effect of Group,  $\chi^2_{(1)} = 4.60$ ,  $p = 0.03$ , and, critically, a significant interaction between List Number and Group,  $\chi^2_{(2)} = 6.54$ ,  $p = 0.04$ .

## Additional Analyses

Analyses were then conducted in order to rule out alternative explanations for the finding of superior recall for English-Spanish Bilinguals relative to Controls on List 3 (resulting from a lack of proactive interference). It seemed possible that the high recall was due, not to the English-Spanish bilingualism *per se*, but either to: (1) bilingualism more generally; or (2) by chance to the order in which groups completed the lists (given the atypical group assignment process). This first alternative explanation was excluded as a likely possibility because English-Spanish Bilinguals and Controls did not differ in their second language proficiency,  $t_{(67)} = 1.03$ ,  $p = 0.31$ , and second

language proficiency did not correlate with recall on List 3 ( $r = 0.03$ ,  $p = 0.81$ ). The second alternative explanation was also excluded as a likely possibility, as a log-linear analysis of a 3-way contingency table of Group (English-Spanish Bilinguals vs. Controls), Language Context (English-Only Context vs. Spanish-Only Context vs. English-Spanish Context), and List Number (List 1 vs. List 2 vs. List 3) revealed no significant or near-significant interaction between Group, Language Context, and List Order,  $G^2 = 2.04$ ,  $df = 12$ ,  $p = 0.99$  (this contingency table is represented in **Table 2** below). In other words, despite the atypical group assignment process, the two groups were exposed to the language contexts in a similar order.

In a final, exploratory analysis, a potential effect of the initial language context (i.e., English vs. Spanish) on subsequent memory performance was assessed. That is, did starting in English or in Spanish on List 1 affect subsequent recall for the English-Spanish Bilinguals? To assess this question, English-Spanish Bilinguals who started on the English-Only Context (i.e., English-starters;  $N = 14$ ) and English-Spanish Bilinguals who started on the Spanish-Only Context (i.e., Spanish-starters;  $N = 13$ ) were compared in their performance on the subsequent single-language context (i.e., Spanish for the English-starters and English for Spanish-starters) and the English-Spanish Context

**TABLE 2 |** Order of conditions by group.

English-Spanish bilinguals (N = 38)	List 1	List 2	List 3
English-only context	14	12	12
Spanish-only context	13	13	12
English-Spanish context	11	13	14
Controls (N = 32)	List 1	List 2	List 3
English-only context	8	11	12
Spanish-only context	11	10	11
English-Spanish context	12	11	9

Note. Each cell contains the number of participants who were exposed to a given language context in a given list order. For example, 14 of the 38 English-Spanish Bilinguals completed the English-Only Context in the first list.

(Eleven started on the English-Spanish Context and were thus not included in this analysis). The English-starters had a mean recall percentage of 53.60% (SD = 12.87%) on the single-language context and 41.07% (SD = 12.85%) on the English-Spanish Context, whereas the Spanish-starters had a mean recall percentage of 53.85% (SD = 20.30%) on the single-language context and 53.21% (SD = 25.58%) on the English-Spanish Context. Thus, numerically, the Spanish-starters performed better than the English-starters on the English-Spanish context. However, the interaction between Group (English-starters vs. Spanish-starters) and Language Context (single-language context vs. English-Spanish Context) did not reach significance in either an ANOVA,  $F_{(1,25)} = 2.49$ ,  $p = 0.13$ , partial  $\eta^2 = 0.09$ , or a generalized linear mixed-effects model with Participant as a random effect (same model details as the above mixed-effects models),  $\chi^2_{(1)} = 2.32$ ,  $p = 0.13$ .

## DISCUSSION

With research on event processing as a guiding theoretical framework, the current study served as a preliminary examination into how changes in the ambient linguistic environment might influence visual memory in bilinguals. Specifically, the study assessed whether a shift in language context within an event (hypothesis 1) or between events (hypothesis 2) enhances a bilingual's visual memory, with the results providing partial initial empirical support for hypothesis 2 (but not hypothesis 1). In partial support of hypothesis 2, the control participants had a consistent downward recall trajectory from the first list to the second list to the third list (i.e., a proactive interference effect), whereas the English-Spanish bilinguals did not have a decline from the second list to the third list (resulting in a recall advantage on the third list for the English-Spanish bilinguals relative to the controls), presumably because of the change in language context between lists. Thus, while merely preliminary, the results suggest that the ambient linguistic background may in some circumstances boost a bilingual's non-linguistic memory performance.

Although the results are consistent with hypothesis 2 (i.e., that a change in language context between events helps memory), they are only partially so, because the memory benefit emerged on the third list but not the second list. Why did the benefit emerge only on the third list? A plausible explanation is that

by the onset of the third list, participants had been exposed to two *different* language contexts, thereby making it clear that the language context changes from list to list and could thus be used to differentiate lists. At the onset of the second list, with exposure to only one list-wide language context, participants did not know that language context would be varied across lists and that it could be used as a distinguishing element to reduce interference.

Notably, the memory benefit on the third list for the English-Spanish bilinguals appears to have been driven more by the single-language contexts (i.e., the English-Only and Spanish-Only Contexts) than the dual-language context (i.e., English-Spanish Context; see bottom of **Figure 4**). Why is this the case? It could be due to the single-language contexts being more distinct from the previous lists. The single-language contexts only share contextual commonalities with one of the previous lists, whereas the dual-language context shares contextual commonalities with both of the previous lists. With more dissimilarity, there is potentially less competition and better memory. A related explanation invokes language-dependent memory (Marian and Neisser, 2000; Marian and Fausey, 2006; Marian and Kaushanskaya, 2007), a phenomenon whereby a language context evokes memories that were encoded in that same language context. In the current paradigm, the single-language contexts would only cue memories of a subset of the previously encoded shapes, whereas a dual-language context would cue memories of all previously encoded shapes, potentially creating more interference.

While there was support for the second hypothesis, there was no support for the first hypothesis. Why did the results fail to support the first hypothesis? That is, why did a within-event language context change not increase memory performance? One possibility is that this benefit is restricted to high proficiency (and high use) bilinguals. Spanish proficiency was low (as was current Spanish use) for many of the English-Spanish bilinguals in the current study, as can be gleaned from the Spanish proficiency (and Spanish use) mean and standard deviation in **Table 1**. However, an exploratory correlation analysis reveals no link between Spanish proficiency and recall performance in the English-Spanish Context (i.e., the within-event context change condition) for the English-Spanish bilinguals (Pearson's  $r = 0.02$ ), suggesting that the low proficiency of many of the bilinguals did not prevent support for hypothesis 1 from emerging. Nevertheless, a follow-up study with high proficiency (and high use) bilinguals is warranted. A second possibility is that, while proficiency may not be especially relevant, code-switching behavior may be, and the potential memory enhancement from a within-event language switch may be restricted to bilinguals who code-switch frequently. A third possibility is that bilinguals incurred a cognitive processing cost when a within-event language switch occurred; that is, bilinguals may have deployed cognitive control resources to suppress the previous language (Philipp and Huestegge, 2015; Olson, 2017; but see Declerck et al., 2019), resulting in fewer resources available for memory encoding. A fourth possibility relates to the strength of the language context; perhaps the ambient linguistic context needs to be stronger and may even need to include expressive language in addition to receptive language. A fifth and final

possibility is that there is a benefit to a within-event language switch, but that it was masked by a potential benefit of the single-language English-Only and Spanish-Only contexts. In other words, a consistent and meaningful context in the form of a single-language context may have aided encoding, which concealed a benefit that may also be derived from a language switch. As this list shows, there are many possibilities for why a within-event language context change effect did not manifest in the current paradigm, rendering this study preliminary and warranting additional studies.

Given that this study served as merely an initial foray into this research topic, there were several limitations (four of which are noted here) that should be addressed with future research. One shortcoming is the limited data on the linguistic backgrounds of the bilinguals (such as whether they code-switch often and are objectively proficient in the two languages). A second shortcoming is the absence of repeated language contexts (such as an English-Only Context followed by an English-Only Context) and a no-language context. A third shortcoming is that cognitive abilities, such as IQ and working memory, were not measured and thus may have differed between groups. A fourth shortcoming is that the retrieval task of drawing shapes was not completely language-free, as instructions were provided in English.

Despite these limitations, the current work provides initial data suggesting that a bilingual's non-linguistic memory can be influenced (and even boosted) by a subtle and task-irrelevant

linguistic context. On a practical level, these data imply a possible way to enhance memory, such as when studying for tests. Potentially, studying for a course's first exam in one language context and for the second exam in a different language context could prove beneficial. On a theoretical level, these data provide further evidence that memory is influenced both by context (Smith and Vela, 2001) and by multi-sensory audio-visual interactions (Thelen et al., 2015). More broadly, the current data underline the tight link between two of our most cherished mental abilities—language and memory.

## DATA AVAILABILITY STATEMENT

The data are available through the Open Science Framework at the following web address: <https://osf.io/f5wrh/>.

## ETHICS STATEMENT

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

## AUTHOR CONTRIBUTIONS

SS conceived the study, designed the study, analyzed the data, wrote and revised the manuscript.

## REFERENCES

- Anderson, M. C., and Neely, J. H. (1996). Interference and inhibition in memory retrieval. in *Memory* eds E. L. Bjork and R. A. Bjork (Cambridge, MA: Academic Press), 237–313. doi: 10.1016/b978-012102570-0/50010-0
- Baetens Beardsmore, H. (1982). *Bilingualism: Basic Principles*. Clevedon: Multilingual Matters.
- Basnight-Brown, D. M., and Altarriba, J. (2016). Multiple translations in bilingual memory: Processing differences across concrete, abstract and emotion words. *J. Psycholinguist. Res.* 45, 1219–1245. doi: 10.1007/s10936-015-9400-4
- Bäuml, K.-H. T., and Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *J. Mem. Lang.* 68, 39–53. doi: 10.1016/j.jml.2012.07.006
- Declerck, M., Koch, I., Duñabeitia, J. A., Grainger, J., and Stephan, D. N. (2019). What absent switch costs and mixing costs during bilingual language comprehension can tell us about language control. *J. Exp. Psychol. Hum. Percept. Perform.* 45, 771–789. doi: 10.1037/xhp0000627
- Dillon, R. F., McCormack, P. D., Petrusic, W. M., Cook, G. M., and Lafleur, L. (1973). Release from proactive interference in compound and coordinate bilinguals. *Bull. Psychon. Soc.* 2, 293–294. doi: 10.3758/bf03329277
- 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
- Goggin, J., and Wickens, D. D. (1971). Proactive interference and language change in short-term memory. *J. Verbal Learning Verbal Behav.* 10, 453–458. doi: 10.1016/S0022-5371(71)80046-4
- Grosjean, F. (2015). Bicultural bilinguals. *Int. J. Biling.* 19, 572–586. doi: 10.1177/1367006914526297
- Hart, B., and Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *Am. Educ.* 27, 4–9.
- Heredia, R. R., and Cieřlicka, A. B. (2019). Bilingual episodic memory: encoding and storage. *Tesol Encyclopedia Eng. Lang. Teach.* 1–7. doi: 10.1002/9781118784235.eelt0958
- Kane, M. J., and Engle, R. W. (2000). Working-memory capacity, proactive interference and divided attention: Limits on long-term memory retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 336–358. doi: 10.1037/0278-7393.26.2.336
- Kurby, C. A., and Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends Cogn. Sci.* 12, 72–79. doi: 10.1016/j.tics.2007.11.004
- Mackey, W. F. (1962). The description of bilingualism. *Can. J. Linguist. Revue* 7, 51–85. doi: 10.1017/S0008413100019393
- Marian, V., and Fausey, C. M. (2006). Language-dependent memory in bilingual learning. *Appl. Cogn. Psychol. Off. J. Soc. Appl. Res. Mem. Cogn.* 20, 1025–1047. doi: 10.1002/acp.1242
- Marian, V., and Kaushanskaya, M. (2007). Language context guides memory content. *Psychon. Bull. Rev.* 14, 925–933. doi: 10.3758/bf03194123
- Marian, V., and Neisser, U. (2000). Language-dependent recall of autobiographical memories. *J. Exp. Psychol. Gen.* 129, 361–368. doi: 10.1037/0096-3445.129.3.361
- Olson, D. J. (2017). Bilingual language switching costs in auditory comprehension. *Lang. Cogn. Neurosci.* 32, 494–513. doi: 10.1080/23273798.2016.1250927
- Pastötter, B., and Bäuml, K.-H. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 977–982. doi: 10.1037/0278-7393.33.5.977
- Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., and Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition* 148, 136–144. doi: 10.1016/j.cognition.2015.12.013
- Philipp, A. M., and Huestegge, L. (2015). Language switching between sentences in reading: exogenous and endogenous effects on eye movements and comprehension. *Biling. Lang. Cogn.* 18, 614–625. doi: 10.1017/s1366728914000753
- Postman, L., Stark, K., and Fraser, J. (1968). Temporal changes in interference. *J. Verbal Learning Verbal Behav.* 7, 672–694. doi: 10.1016/S0022-5371(68)80124-0
- Pu, X., and Tse, C.-S. (2014). “The revised hierarchical model: Explicit and implicit memory,” in *Foundations of Bilingual Memory*

- (New York, NY: Springer), 147–184. doi: 10.1007/978-1-4614-9218-4\_8
- Radvansky, G. A. (2012). Across the event horizon. *Curr. Dir. Psychol. Sci.* 21, 269–272. doi: 10.1177/0963721412451274
- Radvansky, G. A., and Zacks, J. M. (2017). Event boundaries in memory and cognition. *Curr. Opin. Behav. Sci.* 17, 133–140. doi: 10.1016/j.cobeha.2017.08.006
- Sahakyan, L., and Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 1064–1072. doi: 10.1037/0278-7393.28.6.1064
- Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., et al. (2013). Event segmentation ability uniquely predicts event memory. *Cognition* 129, 241–255. doi: 10.1016/j.cognition.2013.07.002
- Smith, S. M., and Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychon. Bull. Rev.* 8, 203–220. doi: 10.3758/bf03196157
- Swallow, K. M., Zacks, J. M., and Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *J. Exp. Psychol. Gen.* 138, 236–257. doi: 10.1037/a0015631
- Thelen, A., Talsma, D., and Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition* 138, 148–160. doi: 10.1016/j.cognition.2015.02.003
- Zacks, J. M., and Swallow, K. M. (2007). Event segmentation. *Curr. Dir. Psychol. Sci.* 16, 80–84. doi: 10.1111/j.1467-8721.2007.00480.x

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

Copyright © 2019 Schroeder. 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.





# Bridging the Gap Between Second Language Acquisition Research and Memory Science: The Case of Foreign Language Attrition

Anne Mickan<sup>1,2\*</sup>, James M. McQueen<sup>1,3</sup> and Kristin Lemhöfer<sup>1</sup>

<sup>1</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, <sup>2</sup>International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, <sup>3</sup>Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Gulsen Yilmaz,  
Humboldt University of Berlin,  
Germany  
Claudia Peñaloza,  
Boston University, United States

### \*Correspondence:

Anne Mickan  
a.mickan@donders.ru.nl

### Specialty section:

This article was submitted to Speech  
and Language, a section of the  
journal *Frontiers in Human  
Neuroscience*

**Received:** 27 June 2019

**Accepted:** 23 October 2019

**Published:** 13 November 2019

### Citation:

Mickan A, McQueen JM and  
Lemhöfer K (2019) Bridging the Gap  
Between Second Language  
Acquisition Research and Memory  
Science: The Case of Foreign  
Language Attrition.  
*Front. Hum. Neurosci.* 13:397.  
doi: 10.3389/fnhum.2019.00397

The field of second language acquisition (SLA) is by nature of its subject a highly interdisciplinary area of research. Learning a (foreign) language, for example, involves encoding new words, consolidating and committing them to long-term memory, and later retrieving them. All of these processes have direct parallels in the domain of human memory and have been thoroughly studied by researchers in that field. Yet, despite these clear links, the two fields have largely developed in parallel and in isolation from one another. The present article aims to promote more cross-talk between SLA and memory science. We focus on foreign language (FL) attrition as an example of a research topic in SLA where the parallels with memory science are especially apparent. We discuss evidence that suggests that competition between languages is one of the mechanisms of FL attrition, paralleling the interference process thought to underlie forgetting in other domains of human memory. Backed up by concrete suggestions, we advocate the use of paradigms from the memory literature to study these interference effects in the language domain. In doing so, we hope to facilitate future cross-talk between the two fields and to further our understanding of FL attrition as a memory phenomenon.

**Keywords:** foreign language attrition, forgetting, retrieval-induced forgetting, interference, competition, second language acquisition, memory

## INTRODUCTION

In 2016, more than 60% of adult European citizens were able to speak *at least one* foreign language (FL; European Commission—Eurostat, 2016). With multilingualism on the rise, learning foreign languages (FLs) is so common these days, it is often taken for granted. Yet, regardless of how ordinary it might seem, mastering a new language is and always will be an immensely complex task. Being able to formulate sentences in any language requires knowledge of its words and grammatical structures, all of which have to first be encoded, and then consolidated and integrated into long-term memory. All of these processes are common to other types of learning as well and are ultimately underpinned by the same fundamental memory processes. Surprisingly, despite the obvious overlap between second language processing and memory function, the empirical investigations of the two have often gone on in parallel; and so for a long time the fields of second language acquisition (SLA) and memory

science developed in isolation from each other<sup>1</sup>. This is also true for the study of FL attrition, which investigates the phenomenon of forgetting a previously mastered FL. While it is not new to apply memory theories of forgetting and their corresponding paradigms to (FL) attrition, they have, as we shall argue, not been used to their full potential. Taking FL attrition as an example, we will posit that both SLA and human memory could benefit from more cross-talk. In doing so, we encourage future research exploiting parallels between the two fields.

## Previous Research on Foreign Language (FL) Attrition

In the past 50 years, researchers have gone to great lengths to document language forgetting. Most of this research has been directed towards first language (L1) attrition in migrants (for reviews on L1 attrition, see Köpke and Schmid, 2004; Schmid, 2016). Much less work has been dedicated to FL attrition, the forgetting of a language learned later in life (for an overview, see Schmid and Mehotcheva, 2012). The present article will focus on this latter type of attrition, partly because it is a less well studied and hence less well-understood type of attrition; but also because, as will become apparent later on, the approach we are advocating in this article is most directly applicable to the FL attrition context.

FL forgetting often first manifests in a decrease in fluency and lexical diversity in the FL (Bardovi-Harlig and Stringer, 2010). Consequently, the majority of studies on FL attrition have focused on the lexicon (i.e., vocabulary), leaving (morpho-) syntax and phonology aside (but see for example, Berman and Olshtain, 1983; Hedgcock, 1991; Dugas, 1999; Tomiyama, 2008). Given this bias in the existing literature, most examples given below will pertain specifically to *lexical* FL attrition, though we will speculate about the applicability of our proposed approach to other types of attrition as well (see “Going Beyond Foreign Language Vocabulary Attrition” section).

Based on the existing FL attrition literature, forgetting seems to set in very quickly after one stops using a FL; yet it then gradually levels off, with the most basic vocabulary apparently preserved in so-called “permastore” (Bahrick, 1984). It furthermore appears that productive skills deteriorate faster than receptive skills (e.g., Bahrick, 1984; De Groot and Keijzer, 2000), and that one tends to first lose the information learned last (or possibly what has been consolidated or practiced the least; i.e., the “regression hypothesis (RH)”); e.g., Cohen, 1986; Olshtain, 1989; Kuhberg, 1992; see “Discussion and Directions for Future Research” section for a more in-depth discussion of this hypothesis and its alternative formulations). Next to those commonalities, attrition differs heavily from person to person. Exactly how severe and fast the attrition process is depends on a variety of factors. Bahrick’s (1984) study on school-learned Spanish, for example, showed that those individuals with the highest Spanish proficiency before attrition onset were

least affected by forgetting (see also Weltens, 1988; Murtagh, 2003; Mehotcheva, 2010). Age at attrition onset (Olshtain, 1989; Bardovi-Harlig and Stringer, 2010) as well as language usage and exposure patterns (e.g., Mehotcheva, 2010) are also believed to play a role in determining the course of FL attrition: the earlier one stops learning the FL and the less exposure one has to it afterward, the more likely one is to suffer from attrition in that language. Finally, one’s attitude towards the target language as well as motivation to maintain it is also believed to influence its attrition rate (e.g., Mehotcheva, 2010). Research on especially the latter two variables’ role in attrition, however, has often yielded contradictory results (see Schmid and Mehotcheva, 2012), thus calling for more research into the determinants of attrition.

Interestingly, in most cases FL attrition (just as in fact any type of attrition) appears to be temporary and reversible: while one may not consciously recall the words learned in French class in high school, studies have demonstrated that relearning seemingly lost FL vocabulary is much easier and faster than learning new FL vocabulary from scratch (e.g., Hansen et al., 2002; de Bot et al., 2004). Such relearning advantages indicate residual storage of the purportedly “forgotten” words and thus speak against complete loss of memory traces. Attrition, like any other kind of forgetting, is thus best understood as a performance problem, characterized by accessibility difficulties rather than actual loss (Sharwood Smith, 1989).

Relatedly, it should be noted that observed FL attrition rates depend heavily on how the FL knowledge is tested: while attriters may be unable to freely recall and produce a word, they might still be able to recognize the word in a lexical decision task or other recognition-based tests. In other words, as mentioned earlier, productive recall failure appears to precede recognition inability. In earlier studies on FL attrition, the focus was often on receptive vocabulary knowledge, as this was thought to give the clearest picture of a person’s existing FL knowledge, but those studies often reported very little to no attrition even after years of no exposure (e.g., Weltens et al., 1989; Grendel, 1993). These null results stand in stark contrast to studies reporting significant attrition in productive recall tasks already within the first year of disuse (e.g., Bahrick, 1984; Mehotcheva, 2010). This distinction needs to be kept in mind in interpreting differences in attrition rates between populations and studies.

## Forgetting Due to Competition and Interference: A Domain-General Perspective

The above studies have made important contributions to our understanding of language attrition. It remains unclear, however, what exactly causes language forgetting, and thus which cognitive mechanisms underlie it. Forgetting is by no means limited to language though, and is, in fact, a rather pervasive phenomenon: we forget where we park our car, or what that distant friend’s name was. Research on forgetting from a more domain-general perspective dates back to the 19th century and Ebbinghaus’ research on the ease of learning and relearning nonsense-syllable sequences (Ebbinghaus, 1885, 1913). Ebbinghaus discovered that memory loss was not linear over time, but logarithmic

<sup>1</sup>We use the terms “second” and “foreign” language interchangeably in this article. Both refer to any language other than one’s mother tongue (L1); that is any language learned later in life, be it a second (L2), third (L3) or even fourth foreign language. In using both terms we stick to common terminology in both “second language acquisition” as well as “foreign language attrition” research.

instead: most forgetting happens over the first minutes to hours and then gradually levels off; note that this resembles what Bahrick (1984) observed for the retention of school-learned Spanish. Ebbinghaus' work inspired many theories about the possible mechanisms behind forgetting (for an overview, see Ecke, 2004; Anderson, 2015). The probably most influential of these is interference theory. Rather than assuming that forgetting is a by-product of time (see decay theory, Thorndike, 1914), interference theory attributes forgetting to interference from related, competing memories. Essentially, it relies on the fact that memories that share a common retrieval cue (e.g., your car being the shared cue for its location today vs. yesterday) compete with one another for selection upon presentation of that cue, thus hindering future retrieval.

One example of forgetting by competition is retrieval-induced forgetting (RIF; Anderson et al., 1994). In a typical RIF study, participants first study a number of category-exemplar pairs (e.g., FRUIT-apple, FRUIT-banana, FURNITURE-table). This phase is followed by selective retrieval practice of some exemplars from some of the categories (FRUIT-banana, but not FRUIT-apple or FURNITURE-table). Finally, recall is tested for all originally studied pairs. Of course, recall is best for the practiced pairs (FRUIT-banana), but interestingly, it is worse for unpracticed exemplars from practiced categories (FRUIT-apple) compared to recall for unpracticed exemplars from unpracticed categories (FURNITURE-table). The mere act of retrieving information can thus hamper access to information related to the practiced material. RIF is typically attributed to executive control processes in the form of inhibition applied to competitors during the retrieval practice phase (e.g., to apple), making these suppressed competitors harder to retrieve at final test (e.g., Anderson, 2003; Bäuml et al., 2005; Román et al., 2009; though see Williams and Zacks, 2001; Raaijmakers and Jakab, 2013; for alternative explanations). RIF effects have been demonstrated with a wide variety of stimulus materials. It thus appears to be a generalizable phenomenon (for a review, see Storm et al., 2015).

## EXPLORING PARALLELS BETWEEN MEMORY AND SLA RESEARCH

Competition does not only exist between exemplars of a semantic category, but also between translation equivalents in different languages that share a common concept. When a speaker of English and Spanish wants to refer to a "table," both Spanish "mesa" and English "table" will be activated and compete for selection. This between-language competition is well-known to affect (online) word production in bilinguals (e.g., Hermans et al., 1998; Colomé, 2001; for an overview, see Kroll et al., 2008). As in RIF, bilingual lexical access is often seen as a matter of executive control: it is assumed that in order to avoid unwanted language selection/production errors, speakers need to inhibit the non-target language during speaking. This, however, can, as an undesirable side effect, lead to later retrieval difficulties in the inhibited language (Green, 1998). In terms of competition for retrieval, translation equivalents (sharing the same concept) are thus similar to pairs of exemplars sharing one semantic category cue. Given this parallel, the question

arises whether between-language competition is also a driving mechanism behind language forgetting, and thus whether the between-language competition observed during the short-term, online processing also has long-term ramifications.

## Between-Language Interference as a Mechanism Behind FL Attrition?

The idea that attrition is the result of complex interactions and competition between languages is not new. Sharwood Smith (1989) as well as Seliger and Vago (1991) already noticed how *L1 attrition* is influenced by the newly acquired FL (L2), for instance in the form of code switches to L2 while speaking in L1. This "cross-linguistic influence hypothesis" (Sharwood Smith, 1989) is also central to more recent approaches to attrition. Its ideas have been formally discussed, for example, within the context of Paradis' Activation Threshold Hypothesis (ATH, see Köpke, 2002; Gürel, 2004; Paradis, 2004, 2007). The ATH assumes that all items in the linguistic system, such as words, are interconnected and influence one another. Each item has an activation threshold (AT). Retrieving a word requires that its activation exceed its AT. The AT is lowered after successful retrieval but is increased again either gradually through disuse or through top-down inhibition during access of other, competing words (as is also believed to be the case in RIF; Anderson, 2003). A latter mechanism is a form of cross-linguistic influence since competing items will often be translation equivalents in other languages. Heightened ATs then lead to retrieval difficulties because more activation is needed to pass them. Ultimately, ATs can be so high that a word can no longer be accessed, and hence is "forgotten" until accessed again (e.g., during re-learning). For Paradis (L1 lexical), attrition is thus the result of a lack of stimulation of certain words, combined with more recent and frequent access of competing for translation equivalents<sup>2</sup>.

Applying ideas and frameworks like the ATH to attrition forms part of an increasing effort towards studying the phenomenon from a psycholinguistic perspective (see also Schmid and Köpke, 2017). Surprisingly though, only a handful of psycholinguistically-inspired attrition articles explicitly connect their ideas to memory theories of forgetting. In an effort to encourage more such cross-talk, Ecke (2004) summarized theories of forgetting from the memory literature and stressed their theoretical relevance and explanatory value for attrition (see also Köpke, 2004). While discussing multiple theories of forgetting, he identifies interference processes as the "main contributor to attrition" (Ecke, 2004; p. 337). We aim to build on Ecke's (and Paradis') contribution, but go one crucial step further: next to using memory theories of forgetting as a

<sup>2</sup>Paradis (2004) also distinguishes between declarative (i.e., explicit) and procedural (i.e., implicit) memory. Declarative memory encompasses knowledge of facts and events and is accessed consciously; procedural memory refers to memory for skills and is largely unconscious (Squire, 1992). Paradis (2004) assumes the language faculty is subserved by both these types of memory as well, with any language's vocabulary and the majority of FL grammar being instances of declarative memory, and L1 grammar mostly procedural (also see Ullman, 2001, 2004). For attrition, Paradis further speculates that declarative memory, and hence L1/FL vocabulary and FL grammar, will be more prone to interference and forgetting than aspects of procedural memory. The present article is on *lexical FL attrition* and thus concerns the *declarative* memory system only.

framework within which to think about attrition, we argue that an important added value of these theories are the *experimental paradigms* that have been developed to test them. The application of these paradigms to the study of (at least FL lexical) attrition forms a crucial addition to the available evidence on FL attrition. Via snapshots of FL ability at different time points, existing observational studies document merely the result, but not the process of attrition itself. Observed links between, for example, language use (as typically measured in questionnaires) and attrition are then purely correlational and do not indicate causal relationships. Traditional attrition studies thus cannot be taken as proof that any factor discovered in this manner is a driving force in attrition. The memory approach to the investigation of forgetting is quite different from the traditional attrition approach: experimental paradigms (e.g., the RIF paradigm) aim at inducing and thus simulating forgetting in a tightly controlled setting. To do so, cognitive psychologists typically manipulate the presence or absence of a presumed cause of forgetting (e.g., interference) while keeping all other potentially relevant factors (e.g., amount of studied materials) constant. By determining the conditions that do and do not lead to forgetting, this procedure allows for *causal* inference.

This experimental approach can be applied to language attrition. Retrieval of words in one language should induce retrieval difficulties, and thus ultimately forgetting of translation equivalents in another language. A handful of studies have put this hypothesis to test, among those a recent study by Mickan et al. (submitted). Participants first learned a set of new L3 Spanish words. A day later, interference was introduced: participants were asked to retrieve half of these newly learned words in either L1 Dutch or L2 English; the other half were not interfered with. Finally, after a 20-min delay, all originally learned words were tested again in Spanish. In line with predictions based on interference theory, participants were worse at recalling interfered words compared to not-interfered words. That is, retrieving translation equivalents in other languages made them forget (some of) the recently learned Spanish words. This interference effect was also visible in naming speed for the words that were still successfully remembered: it took participants longer to retrieve interfered compared to not-interfered words. In reaction times, this effect persisted in a follow-up Spanish test a week later, thus showing that interference has true long-term effects and establishing it as a plausible mechanism behind FL attrition.

Similar observations had previously been made for L1 attrition, for which Levy et al. (2007) had shown that retrieval practice in L2 Spanish impaired subsequent recall for L1 translation equivalents. For FL attrition, there are two corroborating studies with similar results. Bailey and Newman (submitted) tested L1 English learners of L2 Welsh and showed that participants were slower to recall newly learned Welsh words when these words had intermittently been retrieved in L1. Likewise, Isurin and McDonald (2001) found that memory for newly learned L2 Russian words was worse if participants had in between learned the same words in L3 Hebrew as compared to when no extra learning had taken place. Note though that neither Bailey and Newman (submitted) nor Isurin

and McDonald (2001) allowed for consolidation of the newly learned Welsh or Russian L2 words: finding interference in these cases is less surprising given that the newly learned material had no chance to consolidate. Moreover, neither of these studies had a delay between “interference” and final recall, thus not providing evidence for long-term interference effects (for which a delay of at least 20 min is called for, following standard memory procedures; Anderson et al., 1994). Both these studies are thus somewhat more removed from real-world attrition scenarios and less convincingly link interference to long-term forgetting than Mickan et al. (submitted).

Delays of up to 1 week, as tested in Mickan et al. (submitted), might seem minuscule compared to the time delays of multiple months or even years that are typically reported in observational attrition studies. In experimental terms, however, it is quite remarkable for effects to persist for an entire week. While looking at longer time delays would be theoretically interesting for future studies, doing so only makes sense if: (1) it can be guaranteed that the participants are not re-exposed to the target language within that time; and (2) only if additional interference can be reliably quantified. If these two conditions are not met, the experimenter would no longer have the experimental control that makes the simulation approach so useful. What is more, it would be difficult to interpret the outcome of a longer time delay: additional interference through the intermittent use of other languages would happen equally often for items in the interference and no interference conditions, and so would wash the interference effect out. The experimentally induced interference effect might thus disappear with time, however, *not* necessarily because it is not long-lasting, but instead because of additional interference, the very mechanism that caused the effect in the first place. There is thus a logical limit to the length of the delays one can sensibly look at while maintaining experimental and explanatory control; and 1 week is arguably already stretching this limit.

Overall, while there is clearly a need for more studies to establish interference-based forgetting as a robust phenomenon in the language domain, the above-cited studies illustrate that using paradigms from the memory literature complements more traditional approaches to attrition, allows for causal rather than just correlational inferences and thereby advances our understanding of why we forget languages.

## DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

We have argued that the use of memory paradigms for the study of attrition has a number of advantages over more traditional approaches. Simulating rather than observing attrition in real life makes it possible to control, isolate and manipulate possible determinants of the attrition process and assess their effect on retention rates. Most obviously, this includes manipulations of the interference phase as the phase during which attrition is (presumably) taking place. As a secondary question, Mickan et al. (submitted), for example, asked whether the *type* of interference matters: manipulating interference language between experimental groups, they found that another FL (L2 English) interfered more with L3 Spanish



than the participants' dominant L1. Along similar lines, one might ask what the role of typological proximity or language distance is in driving attrition. Do two related languages interfere more with one another than two distant ones? For vocabulary, this question translates to a comparison of cognates (i.e., words that share meaning and form across languages) and non-cognates. Cognates have often, not surprisingly, been reported to be less affected by attrition than non-cognates (e.g., Weltens, 1988). This assumption is in line with interference theory: since cognates share form and meaning, there is no need to suppress the translation equivalent when retrieving a cognate in the target FL. One might even expect a boost for identical cognates given the form overlap. Often, however, words in typologically similar languages overlap only partially (e.g., "table" in English and "Tafel" in Dutch). It is unclear whether such non-identical cognates interfere more or less with each other than two non-cognate translation equivalents. This could be addressed in an experiment that compares interference effects for the two types of nouns. Given that, by definition, two typologically similar languages will have relatively more (non-identical) cognates than two distant languages (Chiswick and Miller, 2005), stronger (or weaker) interference effects between non-identical cognates as compared to non-cognates would be evidence in favor of (or against) the hypothesis that two typologically close languages interfere more with one another than more distant languages.

Likewise, it would be interesting to test whether active use of other languages is necessary to induce forgetting, or whether mere passive exposure to other languages is enough. Evidence from memory studies seems to suggest that active retrieval and response generation (though not necessarily successful retrieval, Hellstedt and Johansson, 2016) is necessary to induce RIF; passive exposure or even reading out loud of exemplar-pairs does not induce forgetting of related items (Anderson et al., 2000; Bäuml, 2002). It remains to be seen whether these findings generalize to the attrition context.

Memory paradigms might also help answer long-standing open questions regarding FL attrition. The RH is a case in point: in its original formulation, RH posits that we tend to forget first the information (e.g., words) we learned last (Jakobson, 1941). However, the order of acquisition itself might not actually matter as much as the degree of learning of a given word (i.e., "best learned = last forgotten"; Hedgecock, 1991). In the real world, these two theories are almost impossible to tease apart: with more time for rehearsal and repetition, remotely learned words will be better encoded than recently learned words. To worsen matters further, the first words one learns in a new language tend to be the most frequent; later learned words or structures instead are usually less frequent, harder to learn, and possibly more vulnerable to forgetting because of their difficulty rather than order of acquisition. A lab study could disentangle these options by manipulating the acquisition order during the initial learning phase while keeping the amount of exposure (and thus the degree of learning) for each word—as well as subsequent interference—equal.

Another option would be to compare receptive and productive recall abilities. As mentioned in "Previous Research

on Foreign Language Attrition" sections, it is generally assumed that productive loss precedes receptive loss. Support for this claim often comes from comparisons across studies, that is between different groups of participants and even different language combinations (though see Bahrick, 1984; De Groot and Keijzer, 2000; for exceptions). The above-reported studies (all testing productive recall with the exception of Bailey and Newman, submitted) could easily be adjusted to test both productive *and* receptive recall at final test (e.g., *via* a lexical decision test in addition to a picture-naming test) and hence could be used to *directly* compare the two. Finding that words that are already "forgotten" in productive tasks are still available to the participant in receptive tasks would be much more convincing proof of the claim that receptive knowledge outlasts productive recall ability than the cross-experiment comparisons that have often been the basis for this claim in the past.

Yet another possible future research avenue concerns manipulations of the level of FL proficiency reached prior to attrition onset. Higher ultimate attainment, as it is often called, has consistently been linked to better retention (see Schmid and Mehotcheva, 2012). Yet, it is unclear whether this means that highly proficient FL learners really attrite less, or whether they, in fact, attrite equally much in absolute terms, but are left with a larger vocabulary because they had a bigger lexicon to begin with (Bahrick, 1984, supports the latter). Most studies to date cannot disentangle these two options because they often do not have the necessary baseline measurement (with the exception of longitudinal studies, yet those tend to be underpowered). A simulation study might again help to disentangle the two: one could test two groups of participants, one as described above, and one with an extended initial learning session (possibly spread over multiple days and/or with an increase in the number of words to be learned) to simulate a higher FL proficiency level, while keeping the amount of interference constant. It should get harder to simulate attrition in the lab as FL proficiency increases if higher ultimate attainment really leads to less attrition. Crucially though, comparing the low and high attainment groups will reveal whether forgetting rates are comparable or actually different across different levels of ultimate FL attainment.

Finally, lab studies could be used to investigate individual differences in attrition. With a large enough pool of participants, there is bound to be variability in forgetting rates, even in an otherwise tightly controlled lab study. It would be interesting to test whether factors known to modulate RIF and interference resolution in bilingual processing also play a role in determining the rate of interference-induced FL attrition. An interesting case in point is executive/cognitive control ability, which has been found to be implicated both in bilingual processing (e.g., Linck et al., 2008) and RIF (e.g., Mall and Morey, 2013; though not always to same extent or even in the same manner, see Aslan and Bäuml, 2011). Traditional attrition studies have, to our knowledge, paid little attention to cognitive control ability. Should it turn out to be a reliable predictor of forgetting rates in the lab, it would merit investigation in large scale studies with real attriters; and might explain

some of the residual variances that remain unexplained by the otherwise mostly socio-linguistic variables assessed in previous studies.

As these examples show, the possibilities using variations of memory paradigms are manifold. Of course, as with any approach, there are also downsides: using tightly controlled experiments clearly comes at the cost of ecological validity. Attrition is a complex, multi-faceted phenomenon, which is simplified in the above experiments. There are undoubtedly questions that our approach will not be able to answer (e.g., motivational/attitudinal aspects). Before looking at large complexities, however, one needs to understand the basic mechanisms. The power of lab studies lies in making exactly that possible: by simplifying matters and isolating individual factors, they can substantially contribute to the development of a cognitive theory of attrition, uncontaminated by the noise that is inevitable in observational studies. Of course, the mechanisms unraveled in the lab will then need to be verified by large-scale longitudinal studies with real attriters. Such studies need to keep detailed records of all possible determiners of attrition, like their participants' language usage patterns, for example, *via* questionnaires or more formal tasks at regular and *short* intervals (monthly at least). While it will remain challenging to recruit a large enough, homogenous sample of "natural" attriters, we would like to highlight the possibility of online testing nowadays. There are many types of measures that can be taken online (both receptive and productive tasks are possible), and one can much more easily reach a large number of people, which would ultimately allow for firmer conclusions if patterns emerge reliably. Hence, we do not propose that experimental studies should replace traditional ones. Instead, we see the two approaches as complementary and believe there should be a healthy balance between them.

## Going Beyond Foreign Language Vocabulary Attrition

### Syntax

The present article focuses on FL lexical attrition. Yet, speaking a language requires much more than just mastery of its words. Similarly, attrition is by no means limited to the forgetting of vocabulary; grammatical structures have, for instance, also been found to attrite (e.g., Hansen, 1999; Tomiyama, 2008). It will be crucial to extend the current line of research to cover these other types of attrition as well. As a first step, one could look at grammatical gender, for which negative transfer and interference effects are well documented in online processing (e.g., Lemhöfer et al., 2008). One might ask whether retrieving L1 gender for a set of nouns (e.g., "der Mond," masculine in German) interferes with and makes people forget just recently learned, but incompatible FL gender assignments (e.g., "la luna," female in Spanish).

For more rule-governed aspects of grammar, the design might need some adjustments: the learning phase, for example, will most likely need to be longer, and include tasks other than just

picture naming for participants to learn the rule. Moreover, the control (i.e., no-interference) condition will need to be carefully chosen: if the syntactic property is not item-specific (unlike grammatical gender), one would need to find a syntactic rule that is comparable in complexity, yet not in conflict with and thus not prone to interference from L1. This might prove challenging for some aspects of grammar. In such cases, one might need to resort to between-subject designs and compare a group that learns a conflicting rule with a group that learns the same rule but in a language which implements this rule similarly to their L1. Though somewhat more challenging, we think that extending our approach to syntax would be a very interesting line of research.

### First Language Attrition

While almost all the above studies are concerned with FL attrition, Levy et al.'s (2007) research suggests that L1 attrition can also be experimentally induced. The design of a study on L1 attrition differs slightly from the FL attrition studies reported above though: there is no need for an initial learning phase in L1; one instead would start with a baseline L1 picture naming test. This baseline speed and accuracy measurement would then be followed by an interference phase that could consist of learning of some of the same words in a new FL. Finally, retrieval speed and accuracy would be measured again for all words in L1. While such a study is perfectly conceivable, it remains unclear how easily deeply engrained L1 knowledge can be interfered with. Even though Levy et al. (2007) observed worse L1 English recall rates after L2 Spanish retrieval practice, Runnqvist and Costa (2012) were later unable to replicate this finding. What is more, Levy et al. (2007) used a rather indirect measure of recall ability (rhyme-generation rather than picture naming), which possibly underestimated L1 productive knowledge. For successful L1 attrition induction, as measured in a final picture naming task, the interference phase might need to be longer, or spaced out over multiple days. Even if L1 attrition proves to be inducible in the lab though, it should be mentioned that the L1 words under investigation will have been learned in the wild and *not* under controlled circumstances. Hence, there will be limitations to the types of simulations one can run; some of the questions addressed in "Going Beyond Foreign Language Vocabulary Attrition" section will be impossible to implement for L1 attrition (e.g., disentangling the effects of order of acquisition vs. degree of learning of L1 words on L1 attrition rates).

## Bridging Between Memory and SLA—a Two-Way Street

Finally, we would also like to emphasize that the benefit of applying psychological theories of forgetting to the language attrition context is not a one-way street. Traditional memory paradigms often make use of artificial learning materials that are hardly representative of what people learn outside the laboratory (e.g., word lists, association pairs, or visual patterns). FL learning and forgetting offers a more realistic scenario to memory researchers to test their theories on. This advantage should not be overlooked

given that the FL scenario is arguably as close to real-life as memory studies on the topic can get, while maintaining tight experimental control.

## Concluding Remarks

We hope to have shown that the approach of using experimental paradigms from human memory research to study FL attrition is a promising avenue for future research. It provides a fresh look at FL attrition that allows for very different types of inferences than those supported by traditional observational studies. We believe that a sound mixture of both approaches is needed if we are to understand what it means to forget a FL. The field of FL attrition is only one example out of many in SLA research for which such interdisciplinary cross-talk is relevant (e.g., effects of testing, spacing and later

consolidation for FL vocabulary learning). We hope to have contributed our share to a productive bridging between SLA and memory science.

## AUTHOR CONTRIBUTIONS

AM wrote and edited the manuscript. KL and JM provided feedback and edited the manuscript. All authors read and approved the submitted version.

## FUNDING

This work was supported by an International Max-Planck Research School for Language Sciences PhD Fellowship awarded to the first author AM (Grant period: 2016–2020).

## REFERENCES

- Anderson, M. C. (2003). Rethinking interference theory: executive control and the mechanisms of forgetting. *J. Mem. Lang.* 49, 415–445. doi: 10.1016/j.jml.2003.08.006
- Anderson, M. C. (2015). “Incidental forgetting,” in *Memory*, 2nd Edn. eds A. Baddeley, M. W. Eysenck and M. C. Anderson (London, New York, NY: Psychology Press), 231–263.
- Anderson, M. C., Bjork, R. A., and Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1063–1087. doi: 10.1037/0278-7393.20.5.1063
- Anderson, M. C., Bjork, E. L., and Bjork, R. A. (2000). Retrieval-induced forgetting: evidence for a recall-specific mechanism. *Psychon. Bull. Rev.* 7, 522–530. doi: 10.3758/bf03214366
- Aslan, A., and Bäuml, K.-H. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 264–269. doi: 10.1037/a0021324
- Bahrack, H. P. (1984). Semantic memory content in permastore: fifty years of memory for Spanish learned in school. *J. Exp. Psychol. Gen.* 113, 1–29. doi: 10.1037/0096-3445.113.1.1
- Bardovi-Harlig, K., and Stringer, D. (2010). Variables in second language attrition. *Stud. Second Lang. Acquis.* 32, 1–45. doi: 10.1017/s0272263109990246
- Bäuml, K.-H. (2002). Semantic generation can cause episodic forgetting. *Psychol. Sci.* 13, 356–360. doi: 10.1111/1467-9280.00464
- Bäuml, K.-H., Zellner, M., and Vilimek, R. (2005). When remembering causes forgetting: retrieval-induced forgetting as recovery failure. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1221–1234. doi: 10.1037/0278-7393.31.6.1221
- Berman, R. A., and Olshtain, E. (1983). Features of first language transfer in second language attrition. *Appl. Linguist.* 4, 222–234. doi: 10.1093/applin/4.3.222
- Chiswick, B. R., and Miller, P. W. (2005). Linguistic distance: a quantitative measure of the distance between English and other language. *J. Multiling. Multicult. Dev.* 26, 1–11. doi: 10.1080/14790710508668395
- Cohen, A. D. (1986). “Forgetting foreign-language vocabulary,” in *Language Attrition in Progress*, eds B. Weltens, K. De Bot and T. J. M. Van Els (Dordrecht: Foris Publications), 143–158.
- Colomé, à. (2001). Lexical activation in bilinguals’ speech production: language-specific or language-independent? *J. Mem. Lang.* 45, 721–736. doi: 10.1006/jmla.2001.2793
- de Bot, K., Martens, V., and Stoessel, S. (2004). Finding residual lexical knowledge: the “Savings” approach to testing vocabulary. *Int. J. Biling.* 8, 373–382. doi: 10.1177/13670069040080031101
- De Groot, A. M., and Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness, cognate status and word frequency in foreign-language vocabulary learning and forgetting. *Lang. Learn.* 50, 1–56. doi: 10.1111/0023-8333.00110
- Dugas, L. G. (1999). *Attrition of Pronunciation Accuracy Among Advanced American Learners of French*. Bloomington: Indiana University. Unpublished Doctoral Dissertation. Available online at: <https://elibrary.ru/item.asp?id=5360458>.
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur Experimentellen Psychologie*. Leipzig: Duncker and Humblot.
- Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology* (H. A. Ruger and C. E. Busenius, Trans.), New York, NY: Teachers College, Columbia University.
- Ecke, P. (2004). Language attrition and theories of forgetting: a cross-disciplinary review. *Int. J. Biling.* 8, 321–354. doi: 10.1177/13670069040080030901
- European Commission—Eurostat. (2016). *Foreign Language Skills Statistics. [Online Dataset]*. Available online at: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign\\_language\\_skills\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_skills_statistics). Accessed October 1, 2019.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Biling. Lang. Cogn.* 1, 67–81. doi: 10.1017/s1366728998000133
- Grendel, M. (1993). *Verlies en Herstel van Lexicale Kennis*. Nijmegen: Radboud University. Unpublished Doctoral Dissertation. Available online at: [https://repository.ubn.ru.nl/bitstream/handle/2066/120134/mmubn000001\\_161400183.pdf?sequence=1](https://repository.ubn.ru.nl/bitstream/handle/2066/120134/mmubn000001_161400183.pdf?sequence=1). Accessed January 26, 2019.
- Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *J. Neurolinguistics* 17, 53–78. doi: 10.1016/s0911-6044(03)00054-x
- Hansen, L. (1999). “Not a total loss: the attrition of Japanese negation over three decades,” in *Second Language Attrition in Japanese Contexts*, ed. L. Hansen (Oxford: Oxford University Press), 142–153.
- Hansen, L., Umeda, Y., and McKinney, M. (2002). Savings in the relearning of second language vocabulary. *Lang. Learn.* 52, 653–678. doi: 10.1111/1467-9922.00200
- Hedgcock, J. (1991). Foreign language retention and attrition: a study of regression models. *Foreign Lang. Ann.* 24, 43–55. doi: 10.1111/j.1944-9720.1991.tb00440.x
- Hellerstedt, R., and Johansson, M. (2016). Competitive semantic memory retrieval: temporal dynamics revealed by event-related potentials. *PLoS One* 11:e0150091. doi: 10.1371/journal.pone.0150091
- Hermans, D., Bongaerts, T., De Bot, K., and Schreuder, R. (1998). Producing words in a foreign language: can speakers prevent interference from their first language? *Biling. Lang. Cogn.* 1, 213–229. doi: 10.1017/s1366728998000364
- Isurin, L., and McDonald, J. L. (2001). Retroactive interference from translation equivalents: implications for first language forgetting. *Mem. Cognit.* 29, 312–319. doi: 10.3758/bf03194925
- Jakobson, R. (1941). *Kindersprache, Aphasie und Allgemeine Lautgesetze*. Uppsala: Almqvist.
- Köpke, B. (2002). “Activation thresholds and non-pathological first language attrition,” in *Advances in the Neurolinguistics of Bilingualism: Essays in Honor of Michel Paradis*, ed. F. Fabbro (Udine: Forum), 119–142.
- Köpke, B. (2004). Neurolinguistic aspects of attrition. *J. Neurolinguistics* 17, 3–30. doi: 10.1016/s0911-6044(03)00051-4

- Köpke, B., and Schmid, M. S. (2004). "Language attrition: the next phase," in *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, eds M. S. Schmid, B. Köpcke, M. Keijzer and L. Weilemar (Amsterdam/Philadelphia: John Benjamins), 1–43.
- Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychol.* 128, 416–430. doi: 10.1016/j.actpsy.2008.02.001
- Kuhberg, H. (1992). Longitudinal L2-attrition versus L2-acquisition, in three Turkish children—empirical findings. *Interl. Stud. Bull.* 8, 138–154. doi: 10.1177/026765839200800203
- Lemhöfer, K., Spalek, K., and Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *J. Mem. Lang.* 59, 312–330. doi: 10.1016/j.jml.2008.06.005
- Levy, B. J., McVeigh, N. D., Marful, A., and Anderson, M. C. (2007). Inhibiting your native language: the role of retrieval-induced forgetting during second language acquisition. *Psychol. Sci.* 18, 29–34. doi: 10.1111/j.1467-9280.2007.01844.x
- Linck, J. A., Hoshino, N., and Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *Mental Lex.* 3, 349–374. doi: 10.1075/ml.3.3.06lin
- Mall, J. T., and Morey, C. C. (2013). High working memory capacity predicts less retrieval induced forgetting. *PLoS One* 8:e52806. doi: 10.1371/journal.pone.0052806
- Mehotcheva, T. H. (2010). *After the Fiesta is Over: Foreign Language Attrition of Spanish in Dutch and German Erasmus Students*. Barcelona, Spain: Universitat Pompeu Fabra. Doctoral Dissertation. Available online at: <https://www.tdx.cat/bitstream/handle/10803/37468/ttm.pdf?sequence>. Accessed May 16, 2019.
- Murtagh, L. (2003). *Retention and Attrition of Irish As a Second Language*. Groningen: University of Groningen. Doctoral Dissertation. Available online at: <https://www.rug.nl/research/portal/files/2999446/thesis.pdf>.
- Olshtain, E. (1989). Is second language attrition the reversal of second language acquisition? *Stud. Second Lang. Acquis.* 11, 151–165. doi: 10.1017/s0272263100000589
- Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam/Philadelphia: John Benjamins.
- Paradis, M. (2007). "L1 attrition features predicted by a neurolinguistic theory of bilingualism," in *Language Attrition: Theoretical Perspectives*, eds B. Köpcke, M. S. Schmid, M. Keijzer and S. Dostert (Amsterdam/Philadelphia: John Benjamins), 121–133.
- Raaijmakers, J. G. W., and Jakab, E. (2013). Rethinking inhibition theory: on the problematic status of the inhibition theory for forgetting. *J. Mem. Lang.* 68, 98–122. doi: 10.1016/j.jml.2012.10.002
- Román, P., Soriano, M. F., Gómez-Ariza, C. J., and Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychol. Sci.* 20, 1053–1058. doi: 10.1111/j.1467-9280.2009.02415.x
- Runnqvist, E., and Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Biling. Lang. Cogn.* 15, 365–377. doi: 10.1017/s1366728911000034
- Schmid, M. S. (2016). First language attrition. *Lang. Teach.* 49, 186–212. doi: 10.1017/S0261444815000476
- Schmid, M. S., and Köpcke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguist. Approaches Biling.* 7, 637–667. doi: 10.1075/lab.17058.sch
- Schmid, M. S., and Mehotcheva, T. H. (2012). Foreign language attrition. *Dutch J. Appl. Linguist.* 1, 102–124. doi: 10.1075/dujal.1.1.08sch
- Seliger, H. W., and Vago, R. M. (1991). "The study of first language attrition: an overview," in *First Language Attrition*, eds H. W. Seliger and R. M. Vago (Cambridge, UK: Cambridge University Press), 3–15.
- Sharwood Smith, M. (1989). "Crosslinguistic influence in language loss," in *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity, and Loss*, eds K. Hyltenstam and L. K. Obler (Cambridge, UK: Cambridge University Press), 185–201.
- Squire, L. R. (1992). Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. *J. Cogn. Neurosci.* 4, 232–243. doi: 10.1162/jocn.1992.4.3.232
- Storm, B. C., Angello, G., Buchli, D. R., Koppel, R. H., Little, J. L., and Nestojko, J. F. (2015). "A review of retrieval-induced forgetting in the contexts of learning, eyewitness memory, social cognition, autobiographical memory, and creative cognition," in *The Psychology of Learning and Motivation*, (Vol. 62) ed. B. H. Moss (Waltham, MA: Elsevier Ltd.), 141–194.
- Thorndike, E. L. (1914). *The Psychology of Learning*. New York, NY: Teacher College.
- Tomiyama, M. (2008). Age and proficiency in L2 attrition: data from two siblings. *Appl. Linguist.* 30, 253–275. doi: 10.1093/applin/amn038
- Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nat. Rev. Neurosci.* 2, 717–726. doi: 10.1038/35094573
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008
- Weltens, B. (1988). *The Attrition of French as a Foreign Language*. Nijmegen: Katholieke Universiteit Nijmegen. Doctoral Dissertation. Available online at: [https://repository.ubn.ru.nl/bitstream/handle/2066/113589/mmubn000001\\_071056505.pdf?sequence=-1](https://repository.ubn.ru.nl/bitstream/handle/2066/113589/mmubn000001_071056505.pdf?sequence=-1). Accessed May 16, 2019.
- Weltens, B., Van Els, T. J. M., and Schils, E. (1989). The long-term retention of French by Dutch students. *Stud. Second Lang. Acquis.* 11, 205–216. doi: 10.1017/s0272263100000619
- Williams, C. C., and Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *Am. J. Psychol.* 114, 329–354. doi: 10.2307/1423685

**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 © 2019 Mickan, McQueen and Lemhöfer. 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.





# Distinguishing Old From New Referents During Discourse Comprehension: Evidence From ERPs and Oscillations

Mante S. Nieuwland<sup>1,2\*</sup>, Cas W. Coopmans<sup>1,3</sup> and Rowan P. Sommers<sup>1</sup>

<sup>1</sup> Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, <sup>2</sup> Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands, <sup>3</sup> Centre for Language Studies, Radboud University, Nijmegen, Netherlands

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Cybelle Marguerite Smith,  
University of Pennsylvania,  
United States

Heather Dee Lucas,  
Louisiana State University,  
United States

### \*Correspondence:

Mante S. Nieuwland  
mante.nieuwland@mpi.nl

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 17 July 2019

**Accepted:** 23 October 2019

**Published:** 14 November 2019

### Citation:

Nieuwland MS, Coopmans CW  
and Sommers RP (2019)  
Distinguishing Old From New  
Referents During Discourse  
Comprehension: Evidence From  
ERPs and Oscillations.  
*Front. Hum. Neurosci.* 13:398.  
doi: 10.3389/fnhum.2019.00398

In this EEG study, we used pre-registered and exploratory ERP and time-frequency analyses to investigate the resolution of anaphoric and non-anaphoric noun phrases during discourse comprehension. Participants listened to story contexts that described two antecedents, and subsequently read a target sentence with a critical noun phrase that lexically matched one antecedent ('old'), matched two antecedents ('ambiguous'), partially matched one antecedent in terms of semantic features ('partial-match'), or introduced another referent (non-anaphoric, 'new'). After each target sentence, participants judged whether the noun referred back to an antecedent (i.e., an 'old/new' judgment), which was easiest for ambiguous nouns and hardest for partially matching nouns. The noun-elicited N400 ERP component demonstrated initial sensitivity to repetition and semantic overlap, corresponding to repetition and semantic priming effects, respectively. New and partially matching nouns both elicited a subsequent frontal positivity, which suggested that partially matching anaphors may have been processed as new nouns temporarily. ERPs in an even later time window and ERPs time-locked to sentence-final words suggested that new and partially matching nouns had different effects on comprehension, with partially matching nouns incurring additional processing costs up to the end of the sentence. In contrast to the ERP results, the time-frequency results primarily demonstrated sensitivity to noun repetition, and did not differentiate partially matching anaphors from new nouns. In sum, our results show the ERP and time-frequency effects of referent repetition during discourse comprehension, and demonstrate the potentially demanding nature of establishing the anaphoric meaning of a novel noun.

**Keywords:** anaphora and coreference resolution, EEG and ERP, time-frequency analysis, N400 and P600, gamma and theta activity, beta activity, old/new effect, lexical repetition

## INTRODUCTION

All nouns have a general meaning, maybe even multiple general meanings, but they acquire a particular, referential meaning when used to refer to someone or something in the world. This flexible use of language and memory yields incredible expressive power for communicating information about the world (e.g., Clark and Murphy, 1982; Martinich, 1985;

Gibson and Pearlmutter, 2011), but also harbors a potential mapping problem for language comprehenders: different words like ‘martian’ and ‘alien’ can have the same referent, and the same word can have different potential referents, such as ‘the alien’ when there are multiple aliens in the context. To examine how people solve such mapping problems, we compared electrophysiological brain responses [event-related potentials (ERPs) and oscillatory activity] to referring expressions that have either one, two or no suitable referent in the linguistic context and that may differ in form (and general meaning) from their referent.

Our study investigates the comprehension of expressions that refer to a previously mentioned referent in the discourse context, i.e., *anaphoric reference* to a linguistic antecedent (e.g., Garnham, 2001; Almor and Nair, 2007). Psycholinguistic theories stipulate the importance of general memory representations and processes during anaphor resolution (e.g., Garrod and Sanford, 1977; Gernsbacher, 1989; McKoon and Ratcliff, 1998; Myers and O’Brien, 1998). Such theories often distinguish an initial activation phase, wherein anaphors are thought to reactivate antecedents from a memory representation of the context (including the described referents), and a subsequent integration phase wherein the reactivated representation is integrated with the unfolding representation of the narrated event. Our main interest in this paper is antecedent activation, which is viewed as a memory-based process in which semantic and syntactic content of an anaphor serves as a memory cue to the antecedent. This process entails the recognition of the anaphor as an instantiation of the antecedent – even when they differ in linguistic form – through the computation of a similarity/identity relation between the two words. This computation gives the language system both great flexibility and speed, by enabling efficient reactivation of semantically complex concepts (e.g., ‘Boris Johnson’), either by other complex concepts (‘blonde haired Brexiteer’) or by minimal-content pronouns (‘he’). The ease with which people understand noun phrase anaphors depends on content overlap of the anaphor with the intended referent relative to other antecedents (e.g., Garrod and Sanford, 1977, 1982; Krahmer and Deemter, 1998; Almor, 1999; Van Gompel et al., 2004; Pyke, 2007). Repeated noun phrase anaphors are easier to resolve than anaphors that only partially match an antecedent (e.g., McKoon and Ratcliff, 1980; Tyler, 1983; Walker and Yekovich, 1987), e.g., ‘the alien’ referring to an alien/a martian<sup>1</sup>. An anaphor whose semantic content does not distinguish between antecedents, e.g., ‘the alien’ in a story about two aliens, is referentially ambiguous. A preceding determiner may already hint at whether the upcoming noun is anaphoric (e.g., Garrod and Sanford, 1977; Clark and Sengul, 1979; Garnham, 1989), with the definite determiners ‘the’ heralding an anaphoric noun phrase and the indefinite determiner ‘a’ heralding a novel, non-anaphoric noun phrase. However, definite noun phrases sometimes introduce a new referent (e.g., Heim, 1982; Fraurud, 1990; Garrod et al., 1994; Poesio and Vieira, 1998; Gundel et al., 2001; Pyke, 2007; Pyke et al., 2007a,b), and people

can use the semantic content of a definite noun as a basis to introduce a novel referent when required, e.g., ‘the alien’ when the context only mentioned astronauts. This process is sometimes referred to as discourse updating (e.g., Burkhardt, 2006), which is related to, yet distinct from the integration process by which people process discourse-level meaning (e.g., Coopmans and Nieuwland, 2019). In other words, processes involved in noun phrase anaphor resolution must distinguish old from new referents, and may do so partly relying on *memory* processes (for a review and computational account, see Pyke, 2007). To address this issue, the current study investigates whether old and new noun phrase referents elicit distinct neural responses, as measured with ERPs and time-frequency analysis.

## Noun Phrase Anaphors and ERPs

Noun phrase anaphors have been associated with several distinct ERP effects, in particular with modulations of the N400, the Late Positive Component (LPC), and the Nref effect. The N400 component is a negative ERP deflection that peaks approximately 400 ms after word onset and is maximal at centroparietal electrodes (Kutas and Hillyard, 1980). The N400 reflects semantic processing and its amplitude is modulated by the relationship between the meaning of a word and its context (Kutas and Hillyard, 1980, 1984; for review, see Kutas and Federmeier, 2011). Words whose meaning is easier to access based on the context typically elicit reduced N400 amplitude compared to words whose meaning is unrelated to the context (Kutas and Federmeier, 2011). Compatible with such findings, noun phrase anaphors that are either repeated from the context or that are contextually implied (‘the conductor’ in a context describing an orchestra) elicit reduced N400 amplitude compared to novel, unrelated noun phrases (e.g., Burkhardt, 2006, 2007). Such N400 modulations may reflect the ease with which the meaning of the anaphor is activated as a function of the context (e.g., Kutas and Federmeier, 2011), and need not reflect higher-level processes such as discourse updating or integration. While recent studies suggest that N400 activity can arise from a cascade of processes that activate and integrate word meaning with context into a sentence-level meaning (e.g., Baggio and Hagoort, 2011; Baggio, 2019; Nieuwland et al., 2019), some studies have failed to observe updating- or integration-related effects on the N400 and found them on a later positive-going ERP component, the LPC (e.g., Burkhardt, 2006, 2007; Delogu et al., 2019). For example, Burkhardt (2006) reported that contextually implied and novel definite referents (‘the conductor’ when the context does or does not describe an orchestra, respectively) elicit a similar post-N400, LPC when compared to a repeated noun phrase anaphor. Burkhardt concluded that the LPC effect reflected the costs of updating a discourse representation with an additional referent (for such costs observed in behavioral studies, see, for example, Murphy, 1984; cf., Pyke, 2007). Subsequent studies found compatible results with related manipulations (Burkhardt, 2007; Schumacher and Hung, 2012). However, the nature and generalizability of this reference-related LPC effect remains to be established. One study with a similar manipulation did not report any LPC modulation (Yang et al., 2007). And while one recent study with repeated proper name anaphors also reported

<sup>1</sup> Partially matching anaphors are particularly taxing to comprehension when they are semantically more specific than the antecedent, like ‘the martian’ referring back to an alien, or when they are atypical of a semantic category rather than typical (e.g., Almor, 1999; Van Gompel et al., 2004).

enhanced LPC for new names (Coopmans and Nieuwland, 2019), two other studies with proper names reported a reverse LPC pattern (Van Petten et al., 1991; Swaab et al., 2004). For example, in a study on natural text comprehension, Van Petten et al. (1991) reported enhanced LPC amplitude for repeated proper names compared to novel names, and suggested that these effects reflect the retrieval of semantic information associated with known names<sup>2</sup>.

Whereas the semantic relationship between an anaphor and its context can modulate the N400 (and LPC), the referential relationship between an anaphor and its context can elicit an LPC effect or yet another ERP effect. Referentially ambiguous anaphors, like 'the alien' when two different aliens were mentioned in the context, or the pronoun 'he' without a male antecedent in the sentence, elicit a sustained, frontal negativity compared to non-ambiguous anaphors (the Nref effect; for reviews, see Van Berkum et al., 2007; Nieuwland and Van Berkum, 2008b). The Nref effect can start at about 200–300 ms after word onset (not unlike an N400 effect, at least for written language comprehension), and has been obtained with noun phrases (e.g., Van Berkum et al., 1999a, 2003; Nieuwland et al., 2007; Nieuwland and Van Berkum, 2008a), pronouns (e.g., Nieuwland and Van Berkum, 2006; Nieuwland, 2014; Karimi et al., 2018), noun phrase ellipsis (e.g., Martin et al., 2012), and proper names (e.g., Coopmans and Nieuwland, 2019). While the onset latency of the Nref suggests that it indexes processes that rapidly link expressions to potential referents, the sustained nature of this effect suggests that inability to resolve reference may have a prolonged impact on comprehension (see Nieuwland et al., 2007; Nieuwland and Martin, 2017).

## Anaphora and Neural Oscillations

ERPs are the most common dependent measure in electrophysiological research on language comprehension, but some studies have instead or additionally examined neural oscillatory responses, measured with time-frequency analysis. Oscillatory activity reflects the synchronization and desynchronization of neural populations, i.e., the transient coupling or uncoupling of functional cell assemblies (e.g., Engel et al., 2001; Buzsáki and Draguhn, 2004). ERPs and oscillatory responses are complementary electrophysiological measures, because whereas ERP analysis can only detect activity that is both time- and phase-locked to stimulus onset, time-frequency analysis can detect activity that is time-locked only<sup>3</sup>. To date, only a handful of studies have applied time-frequency analysis to examine reference processing (Van Berkum et al., 2004;

Heine et al., 2006; Boudewyn et al., 2015<sup>4</sup>; Meyer et al., 2015; Nieuwland and Martin, 2017; Coopmans and Nieuwland, 2019).

Heine et al. (2006) reported that pronouns with low-frequency antecedent nouns elicit reduced power in the theta (4–7 Hz) range compared to pronouns with high-frequency antecedents. They argued that pronoun resolution is relatively easy for low-frequency words because they capture elevated attention. Consistent with a role for memory processes in pronoun resolution, source analyses (albeit based on low resolution, 27-channel EEG data) suggested a contribution from the parahippocampal gyrus to the observed theta effect.

Meyer et al. (2015) reported that pronouns with antecedents that were embedded in a subordinate clause elicit enhanced theta power compared to pronouns referring to non-embedded antecedents, and source analysis suggested contributions from left-frontal, left-parietal, and bilateral-inferior-temporal cortices (based on 64-channel data). Meyer and colleagues argued that embedded antecedents were harder to retrieve from verbal working memory compared to non-embedded antecedents.

In other words, both Heine et al. (2006) and Meyer et al. (2015) took enhanced theta power to index difficulty with reactivating or retrieving an antecedent from memory, in line with the literature on theta effects and verbal and non-verbal working memory retrieval (e.g., Bastiaansen and Hagoort, 2003; Jacobs et al., 2006). However, it is unclear whether the reported theta effects were truly oscillatory in nature and distinct from phase-locked activity that also yields an associated ERP effect.

Two other studies report effects of reference processing in the gamma (>30 Hz) frequency range but not in the theta range. An unpublished study by Van Berkum et al. (2004) reported increased gamma power (40–55 Hz) range for pronouns with a single matching antecedent (e.g., 'she' in a sentence with one male and one female antecedent) compared to pronouns with two or zero matching antecedents ('she' in a sentence with either two female or two male antecedents, respectively). A study by Nieuwland and Martin (2017) re-analyzed four EEG datasets that had initially been collected for ERP analysis (Nieuwland and Van Berkum, 2006; Nieuwland et al., 2007; Martin et al., 2012; Nieuwland, 2014). In each dataset they observed increased gamma power for referentially successful expressions (pronouns, noun phrases, ellipsis that matched a single antecedent) compared to referentially problematic expressions (with either two matching antecedents or no matching antecedent). In one of those four studies, they compared the oscillatory response to a matching pronoun with that to a mismatching, ambiguous pronoun (e.g., "The boy said that he/she would win the race"). They found a brief gamma power increase in the 35–45 Hz range between 400 and 600 ms after pronoun onset. Beamformer source analysis (64-channel data) suggest contributions from left posterior parietal cortex, a brain region that is thought to be involved in recognition memory (Cabeza et al., 2008). They also observed a more extended gamma power increase in

<sup>2</sup>In studies on recognition memory, correctly recognized items are associated with enhanced parietal LPC responses compared to correctly rejected items, which is referred to as the parietal old/new LPC effect (e.g., Van Petten and Senkfor, 1996; Rugg and Curran, 2007; Voss and Paller, 2009). It is unknown whether such LPC effects are related to LPC effects associated with anaphoric processing.

<sup>3</sup>The brain continuously generates neural oscillations at a wide range of frequencies and the phase of these frequencies may differ at stimulus onset. By averaging over trials, ERP analysis cancels out activity that differs in phase over trials. However, a stimulus may impact the activity in a specific frequency band without changing its phase (e.g., Bastiaansen et al., 2013; Lewis et al., 2015). This impact cannot be detected in an ERP analysis, but can be detected in time-frequency analysis of spectral power.

<sup>4</sup>Boudewyn et al. (2015) investigated correlations between antecedent-elicited spectral power and ERP activity associated with noun phrase anaphora, but did not investigate spectral power changes associated with anaphors themselves and is therefore not discussed in this section.

the 60–80 Hz range between 500 and 1000 ms after pronoun onset, with source analysis suggesting a contribution from left inferior frontal gyrus, and brain region that is thought to be involved in sentence-level unification/integration processes (e.g., Hagoort, 2005; Hagoort and Indefrey, 2014). Based on these findings, Nieuwland and Martin (2017) argued that the observed gamma-band power increases reflect successful referential binding and resolution, which links incoming information to antecedents through an interaction between the brain's recognition memory networks and fronto-temporal language network.

In a recent study on comprehension of proper name anaphors, Coopmans and Nieuwland (2019) observed effects in both the theta and gamma frequency range. Their participants read story contexts that described characteristics of two people (e.g., “John and Peter are the best players in the football team”), followed by a target sentence containing a repeated or novel proper name that was either congruent or incongruent with the discourse context (e.g., “The top scorer of the team was John with thirty goals in total”). Repeated names elicited increased theta power compared to new names, which may have originated from anterior temporal regions (based on beamformer source analysis of 64-channel data), and a weak effect in the 40–55 Hz gamma range (see also Van Berkum et al., 2004). Discourse-congruent names elicited increased gamma power (60–80 Hz) compared to incongruent names in the 500–1000 ms time window, with source analysis suggesting a contribution from left frontal cortex.

In sum, reference processing thus far has been associated with modulations of theta and gamma activity. However, the available studies report mixed results, which may have to do with differences in type of linguistic expression (pronoun, noun phrase, proper name) and experimental manipulation (difficulty with retrieving an antecedent, referential ambiguity, comparing old, anaphoric names with new names). Heine et al. (2006) and Meyer et al. (2015) investigated pronouns that had uniquely identifiable antecedents but differed in the extent to which the antecedent was easily retrieved from memory, whereas Nieuwland and Martin (2017) compared ambiguous to unambiguous anaphors, and Coopmans and Nieuwland (2019)

compared anaphoric to non-anaphoric proper names that were coherent or incoherent with the preceding discourse. The type of linguistic expression may matter in particular for modulations of theta activity, because theta activity can be modulated by a word's semantic meaning (e.g., Bastiaansen et al., 2005, 2008).

## The Present Study

The present EEG study investigated how people establish anaphoric meaning for noun phrases, which contain more semantic content than pronouns and proper names and therefore allow an investigation of how people can use semantic memory representations (i.e., word meaning) to resolve anaphoric reference (e.g., Garrod and Sanford, 1977; Garnham, 1989). This semantic richness raises the question of whether or to what extent anaphoric noun phrases are resolved through similar processes as other types of anaphors. Our participants listened to two-sentence story contexts followed by a written sentence that contained a target noun. These stories appeared in one of four conditions that only differed in the two antecedents described in the first sentence (see **Table 1**). Due to these differences, the target noun was either a given or ‘old’ anaphor (lexically identical to one of the two antecedents), an ‘ambiguous’ anaphor (lexically identical to both antecedents), a ‘partial-match’ anaphor (lexically different from both antecedents but close enough in meaning to one of the antecedents to allow an anaphoric interpretation, as indicated in a norming pre-test), or a ‘new’ noun (lexically and semantically different enough from both antecedents such that a novel referent must be introduced). After each story, the participants used a button press to indicate whether the target sentence contained an anaphoric noun phrase or not (old/new judgment). While this task requires meta-linguistic judgments and is therefore not representative for naturalistic comprehension, we included it in order to separate trials in which participants arrived at the intended interpretation from trials where they did not (as is also done in studies on recognition memory).

For this experimental design, we derived hypotheses from memory-based theories of anaphor resolution (e.g., Myers and O'Brien, 1998), which distinguish an early phase of memory

**TABLE 1** | Example stimulus item in Dutch, containing all four conditions.

Condition	First spoken context sentence	Second spoken context sentence	Written target sentence
Old	Een oude receptioniste en een jonge sollicitant plannen een nieuwe afspraak. <i>An old receptionist and a young applicant are planning a new appointment.</i>	De afspraak vindt in mei plaats. <i>The appointment will take place in May.</i>	Na het plannen schrijft de <b>receptioniste</b> direct de datum op. <i>After planning, the receptionist immediately writes down the date.</i>
Ambiguous	Een oude receptioniste en een jonge receptioniste plannen een nieuwe afspraak. <i>An old receptionist and a young receptionist are planning a new appointment.</i>		
Partial	Een oude balie medewerker en een jonge sollicitant plannen een nieuwe afspraak. <i>An old desk clerk and a young applicant are planning a new appointment.</i>		
New	Een oude sollicitant en een jonge sollicitant plannen een nieuwe afspraak. <i>An old applicant and a young applicant are planning a new appointment.</i>		

Approximate English translation is provided below each sentence. The critical word is printed in bold for presentation purposes only. All stimuli available via our OSF page <https://osf.io/uak8g>.



activation from subsequent discourse updating and integration. We hypothesized that activity in the early phase primarily depends on the ease with which word meaning can be activated, which is easiest for repeated nouns. For the ERP analysis, we expected to observe this phase in N400 activity (e.g., Kutas and Federmeier, 2011), with smaller (less negative) N400 ERPs for old and ambiguous anaphors compared to new nouns and partial-match anaphors (i.e., a lexical repetition effect on the N400, e.g., Van Petten et al., 1991; Besson et al., 1992; Swaab et al., 2004). We also expected smaller N400s for partial-match anaphors compared to novel nouns, because the semantic meaning of partial-match anaphors is more strongly related to the context and therefore more easily activated than that of novel nouns (Kutas and Federmeier, 2000). In our time-frequency analysis, we tested for complementary effects in the theta- and gamma-band, which are strongly associated with memory processes. We expected to observe enhanced theta (and low gamma) power for anaphoric nouns compared to new nouns (see Nieuwland and Martin, 2017, for discussion). Such a pattern would be compatible with the proper name effects recently observed by Coopmans and Nieuwland (2019), and consistent with theta and gamma band effects associated with successful recognition in memory research. However, this hypothesis disregards the association between theta activity and activation of semantic representations (e.g., Bastiaansen et al., 2005, 2008; Piai et al., 2016), which is why we also considered an alternative possibility: if theta power tracks the amount of semantic activation (e.g., Bastiaansen et al., 2005), new nouns could elicit enhanced theta power compared to old nouns.

Activity in the later, post-N400 time-window may be associated with either repetition or with discourse-level processes<sup>5</sup>. For example, we considered the possibility that anaphoric nouns would elicit larger LPCs than novel nouns (Van Petten et al., 1991; Swaab et al., 2004), although such a pattern for repeated referents has not yet been found for noun phrases. We also considered an alternative possibility, namely that new nouns would elicit larger LPCs than anaphors (which would suggest that this component indexes updating of the discourse representation to include a new referent; Burkhardt, 2006; Coopmans and Nieuwland, 2019). Furthermore, we expected ambiguous anaphors to elicit an Nref effect compared to non-ambiguous anaphors (Van Berkum et al., 1999a; Nieuwland et al., 2007; Nieuwland and Van Berkum, 2008a,b). For the time-frequency analysis, we expected enhanced high gamma (60–80 Hz) activity for anaphors compared to new nouns, possibly related to updating or integration processes (e.g., Nieuwland and Martin, 2017).

Of specific interest were the processes involved in resolving partially matching anaphors, which differ in form and meaning from the antecedent (e.g., *balimedewerker-receptioniste*, *desk clerk-receptionist*, in **Table 1**). Previous literature suggests that such anaphors may be relatively difficult to resolve because they unexpectedly introduce new information

(Garrod and Sanford, 1977; Garnham et al., 1997), which is atypical for anaphors. This violation of pragmatic principles may cause people to consider the possibility that a new referent is being introduced, and the resulting situation can only be resolved through an elaborative, anaphoric inference based on the semantic similarity of anaphor and antecedent. In such an account, old, new, and partially matching anaphors may elicit a difference in measures that index semantic activation (N400, possibly theta), but later measures could indicate whether the partially matching noun is temporarily processed as a new noun, by comparing the associated neural responses to responses elicited by new or old nouns, respectively. Alternatively, ambiguity regarding the anaphoric nature of partially matching nouns could lead to the type of Nref effect we expected for ambiguous nouns (Nieuwland, 2014).

## MATERIALS AND METHODS

We pre-registered the number of participants and crucial elements of data processing and analysis on AsPredicted.org, available through the OSF pre-registration portal<sup>6</sup>. Procedures and analyses that were not pre-registered are designated as exploratory.

### Participants

We invited 41 participants (right-handed native-Dutch speakers who were free from known learning or language disorders) from the MPI participant pool (34 females, average age = 23.3 years, range = 19–32 years). All participants gave informed written consent to take part in the experiment, which was approved by the Ethics Committee for Behavioural Research of the Social Sciences Faculty at Radboud University Nijmegen in compliance with the Declaration of Helsinki. They received 18 euros for their participation. One participant did not finish the experiment and was replaced. For the ERP analysis, we excluded three participants due to low trial numbers (on average across conditions < 35 artifact-free trials with correct responses). For the time-frequency analysis, we excluded five participants due to low trial numbers.

### Stimuli

The entire set of stimuli consisted of 200 experimental and 50 filler mini stories in Dutch. Each mini story consisted of three sentences, of which the first sentence introduced two antecedents (persons or objects), and the third sentence contained a critical noun phrase that also denoted a person or object (see **Table 1**). The antecedents appeared in an indefinite conjoined noun phrase that included two prenominal adjectives and that either repeated the same noun (ambiguous and new condition) or contained different nouns (old and partial-match condition). The critical word (CW) in the third sentence was always a definite noun phrase without a prenominal adjective, was never the first or second word of the sentence, and was followed by exactly four additional words in the sentence.

<sup>5</sup>Because we did not manipulate the ease with which old or new referents could be integrated (i.e., whether they were semantically coherent with the preceding discourse), our hypotheses primarily focused on the discourse updating processes associated with a new referent.

<sup>6</sup><https://osf.io/7pkc5>

Both the second context sentence and the target sentence were identical across conditions. The four conditions differed only in the two antecedents described in the first sentence, which determined the available co-referential relationships between the critical word and the antecedents. The critical word in the old condition was a repeated name anaphor, which was identical to and co-referential with one antecedent (*receptionist-receptionist*). The ambiguous anaphor was identical to both antecedents. The partially matching anaphor was semantically overlapping or synonymous with only one of the antecedents (*desk clerk-receptionist*, we report semantic similarity values below), which were chosen such that the critical word would be a reasonably plausible anaphor for one antecedent. In the new condition, the critical word did not appear elsewhere in the context, and it had little semantic overlap with either antecedent to the extent that it would not be a plausible anaphor. We tried to write stories wherein the partially matching anaphor was related in meaning to the story context and to the antecedent and plausibly co-referential with the first antecedent, and wherein the novel noun was at least somewhat related in meaning to the story context but not plausibly co-referential and would therefore be interpreted as introducing a new referent. In both the given and the partial-match condition, the anaphor always referred to the first antecedent in the context sentence.

In an effort to optimize our stimulus set for these constraints, we performed a behavioral norming study on an initial set of 240 items. Twenty-four participants, who did not take part in the EEG experiment, each read 240 stories in the New, Old or Partial-Match condition, with conditions counterbalanced over three stimulus lists such that each participant saw the same number of items per condition and each item was seen in each condition equally often across participants. The participants read each story presented as a whole on the screen with the target word in boldface, and judged whether each target word referred back to someone or something in the story ('old') or whether it referred to someone or something new ('new'). Based on the results, we selected the best 200 items, that is, items receiving responses most in line with our design (partial-matching and old anaphors considered 'old' and novel nouns considered 'new'). Because we made further changes to the selected materials after the norming study, and because we also collected old/new judgments during the main EEG experiment (which are the most relevant behavioral data), the results of the stimulus norming test are not discussed here, but they can be found on our OSF page<sup>7</sup>.

For the final set of items, we confirmed that partially matching nouns were more semantically similar to the corresponding first antecedent than new nouns. We used semantic similarity scores obtained from 'snaut' (Mandera et al., 2017)<sup>8</sup>, using a word2vec-compatible 'continuous bag of words' (CBOW) model for Dutch lemmas, trained on the SONAR-500 corpus and an additional subtitle corpus. With the caveat that not all our words found a match in the corpus (155 partially matching nouns and 149 new nouns), partially matching nouns and their antecedents had a smaller semantic distance (i.e., were more

semantically similar) than new nouns and their antecedents (0.57 versus 0.70, two-sample  $t$ -test  $t = 8.47$ ,  $p < 0.001$ ).

For the EEG experiment, we added 50 filler items to the final set of 200 experimental items. Three fillers served as practice items (one item corresponding to the New, Old and Partial-Match condition each). The other 47 fillers had the same format as the New condition, which was done to increase the percentage of stories without an anaphor. Roughly 60% of the items in each stimulus list contained an anaphor, while 40% of all items contained a new noun.

We followed previous studies on discourse comprehension (Van Berkum et al., 1999a,b; Nieuwland and Van Berkum, 2006, 2008a) by using a mixed-modality design where the context sentences were spoken and the target sentence was written. We created audio-recordings (44.1 kHz sampling) for the four different story contexts. All recordings were performed by the same native-Dutch, female speaker in a sound-shielded booth. This speaker recorded both context sentences for the old condition. For the other three conditions, only the first context sentence was recorded, which was then paired with the second sentence recorded for the Old condition. Because the speaking rate for the recordings was considered slightly too fast for the experiment, the recordings were lengthened by 15% using the Praat software (Boersma and Weenink, 2013). This yielded a speaker rate that was comfortable for listening without being unnaturally slow (as evaluated by two native speakers of Dutch) and without compromising sound quality.

As there were four conditions, we created four stimulus lists. Each list contained 50 items of each condition and 50 filler items. The lists were created such that they never contained multiple conditions of the same item. Next, the four lists were distributed equally among the participants. For each participant, the items in the list were pseudorandomized, such that there were no consecutive trials of the same condition.

## Procedure

After participants had given written informed consent, they were tested in a sound-shielded booth. They were told that the experiment was about understanding mini stories. They were also told that the last sentence of each mini story was about a specific person or object, and that they had to indicate after each trial whether this person or object had been referred to before ('old') or not ('new'). To discourage participants from using a strategy based on noun repetition alone, and to encourage them to establish co-referential relationships between anaphor and antecedent whenever plausible, we told them that anaphors did not have to be exactly the same as antecedent and could be a different word.

Each trial started with a fixation cross. When participants pressed a button, the two spoken context sentences were presented over loudspeakers located on the desk in front of the participant. Then, 700 ms after the end of the audio recording, the third sentence was presented visually, one word at a time, in black letters (font Lucidia Console, size 20) on the center of a computer screen, which had a light gray background. Each word was presented for 300 ms, with an inter-stimulus-interval of 300 ms. Sentence-final words were presented for 550 ms and

<sup>7</sup><https://osf.io/uak8g/>

<sup>8</sup><http://meshugga.ugent.be/snaut>

followed by a blank screen for 300 ms. Subsequently, the old-new question was presented, which could be answered by a button press (left button for “new,” right button for “old”). Participants were asked to minimize eye blinks and body movements during the word-by-word presentation of the third sentence.

The experiment started with three practice trials, after which the experimental trials would be presented. These were presented in five blocks of 50 items. Participants were allowed to take short breaks between blocks. In total, the experiment lasted approximately 80 min.

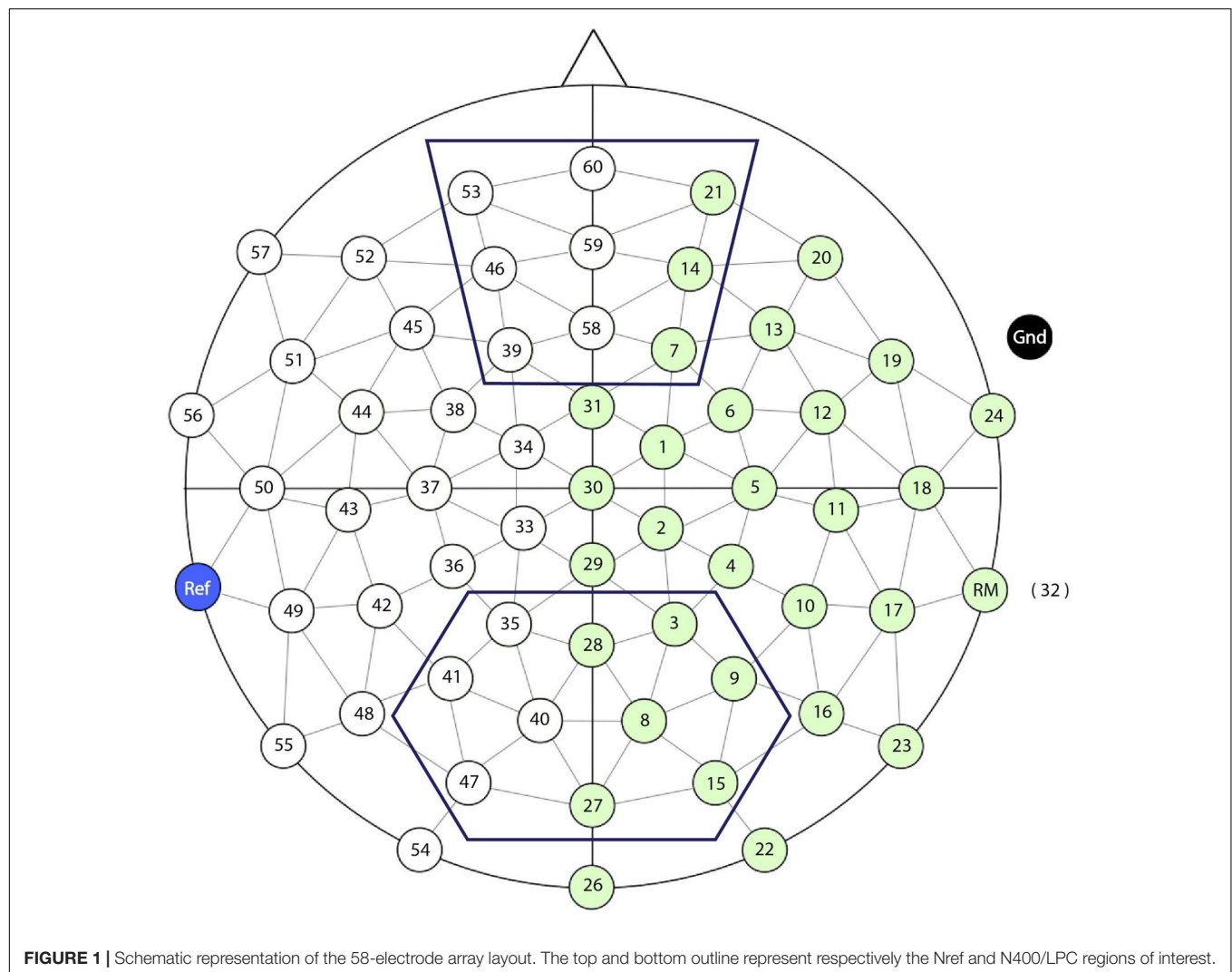
## EEG Recording

The electroencephalogram (EEG) was recorded using an MPI custom actiCAP 64-electrode montage (Brain Products, Munich, Germany), of which 58 electrodes were mounted in the electrode cap (see **Figure 1**). We recorded horizontal EOG with one electrode placed on the outer canthus of the right eye, and vertical EOG with two electrodes placed below both eyes. One electrode was placed on the right mastoid, the reference electrode was placed on the left mastoid, and the ground was placed on

the forehead. The EEG signal was amplified through BrainAmp DC amplifiers, referenced online to the left mastoid, sampled at 500 Hz and filtered with a passband of 0.016–249 Hz. Pre-processing was performed in BrainVision Analyzer 2.1 (Brain Products, Munich, Germany).

## ERP Pre-processing and Analysis

We first visually inspected the raw data and interpolated bad channels if they contained strong 50 Hz line noise or indicated broken electrodes. The data was then band-pass filtered at 0.03–40 Hz (24 dB/oct) and re-referenced to the average of the left and right mastoid. Segments were extracted ranging from –500 to 1500 ms relative to CW onset, and segments in which an incorrect response had been given (‘new’ response to old, partial-match or ambiguous; ‘old’ response to new) were rejected. Based on visual inspection, we then removed bad segments containing large eye movements, muscle activity, or amplifier blocking. Subsequently, we removed blinks, eye-movements and steady muscle activity using Independent Component Analysis (ICA; Jung et al., 2000), using ICA weights from a 1 Hz high-pass filtered version



of the data. We then performed baseline correction using a 250 ms pre-CW baseline interval, and then automatically rejected segments that contained voltage values exceeding  $\pm 90 \mu\text{V}$ . We excluded three participants who retained fewer than 140 trials in total (35 per condition, on average). In the final set of trials for the ERP analysis, participants had on average 45.3 trials for ambiguous nouns, 42.8 for old nouns, 43.7 for new nouns, and 35.4 for partially matching nouns.

For analysis of the behavioral responses, we performed mixed effects logistic regression (Baayen et al., 2008) in the R software (R Core Team, 2018)<sup>9</sup>, with correction for multiple comparisons using the Holm method (Holm, 1979, implemented in the `p.adjust` function). For the ERP analysis, we performed a linear mixed-effects analysis (Baayen et al., 2008). The ERP analyses were done separately for three dependent variables corresponding to a specific region of interest (ROI): N400, LPC and Nref.

For the N400, we calculated the average voltage across the centroparietal electrodes 35, 28, 3, 41, 40, 8, 9, 47, 27, 15 in a 300–500 ms window after CW onset, for each trial and each participant (see **Figure 1**). For the LPC, we calculated the average voltage across these same centroparietal electrodes but in a 500–1000 ms window after CW onset. For the Nref, we calculated the average voltage across the frontal electrodes 53, 60, 21, 46, 59, 14, 39, 58, 7 in a 300–1500 ms window after CW onset.

The variable ‘condition’ had four levels: old, ambiguous, new, and partial, which were deviation coded. The models had subject and item as random effects, and initially included a by-subject and by-item random slope for ‘condition’ (Barr et al., 2013) but these slopes were removed due to convergence issues. We compared models with a chi-square test using R’s `anova()` function, and treated  $p$ -values below  $\alpha = 0.05$  as statistically significant. For the N400 and LPC, we performed all (Holm-corrected<sup>10</sup>) pairwise comparisons between given anaphors, partially matching anaphor and novel nouns, but not ambiguous anaphors. For the Nref, we specifically tested whether ERPs elicited by ambiguous anaphors were more negative than the mean ERP values across the other three conditions.

## Oscillatory Pre-processing and Analysis

After interpolation of bad channels, we band-pass filtered the data at 0.1–100 Hz (24 db/oct), re-referenced the data to the average of the left and right mastoid, and segmented the data into epochs ranging from –1000 to 2500 ms relative to CW onset. After this, we used the same procedure as for the ERP analysis to reject trials with incorrect responses or artifacts and to perform ICA-based correction for blinks, eye movements and steady muscle activity. The resulting dataset for each participant contained many artifact-free trials with voltage values exceeding  $\pm 100 \mu\text{V}$ . We therefore considered the preregistered  $\pm 100 \mu\text{V}$  amplitude criterion to be too conservative, excluding on average

50.9 trials per participant ( $SD = 38.6$ ). We chose to use a more liberal difference criterion, which excluded segments for which the difference between the maximum and minimum voltage exceeded  $200 \mu\text{V}$  (see Coopmans and Nieuwland, 2019). We excluded four participants who retained fewer than 140 trials in total. In the final set of trials for the time-frequency analysis, participants had on average 46.5 trials for ambiguous nouns, 45.2 for old nouns, 43.2 for new nouns, and 37 for partially matching nouns.

Time-frequency analysis was performed using the Fieldtrip toolbox (Oostenveld et al., 2011). We performed time-frequency analysis in two different, but partially overlapping frequency ranges. For the low (2–30 Hz) range, we used a 400-ms Hanning window to compute power changes in frequency steps of 1 Hz and time steps of 10 ms. For the high (25–90 Hz) frequency range, we computed power changes with a multitaper approach (Mittra and Pesaran, 1999) based on Slepian sequences as tapers, with a 400-ms time-smoothing and a  $\pm 5$  Hz spectral-smoothing window, in frequency steps of 2.5 Hz and time steps of 10 ms. Then, for each trial, we computed power in the post-stimulus interval as a relative change from a baseline interval spanning from –500 to –250 ms relative to CW onset. Average power changes per subject were computed for each condition separately.

For the statistical analysis, we pre-registered three ROIs: theta (4–7 Hz) activity in the 0–1000 ms interval after critical word onset, averaged over frequency but not over time; low gamma (35–45 Hz) in the 400–600 ms interval, average over both frequency and time; high gamma (60–80 Hz) in the 500–1000 ms interval, average over both frequency and time. In addition to these ROIs, we also pre-registered an analysis of the 200–1500 ms time window that did not average activity over time or frequency.

We used cluster-based random permutation tests (Maris and Oostenveld, 2007) to compare differences in oscillatory power across conditions. In brief, this statistical test works as follows: first, by means of a two-sided dependent samples  $t$ -test we performed all pairwise comparisons between the four conditions on the three dependent variables described above, which yielded uncorrected  $p$ -values. Neighboring data triplets of electrode, time and frequency-band that exceeded a critical  $\alpha$ -level of 0.05 were clustered. Clusters of activity were evaluated by comparing their cluster-level test statistic (sum of individual  $t$ -values) to a permutation distribution that was created by computing the largest cluster-level  $t$ -value on 1000 permutations of the same dataset. Clusters falling in the highest or lowest 2.5th percentile were considered statistically significant. We used the correct-tail option that corrects  $p$ -values for doing a two-sided test, which allowed us to evaluate  $p$ -values at  $\alpha = 0.05$ .

## RESULTS

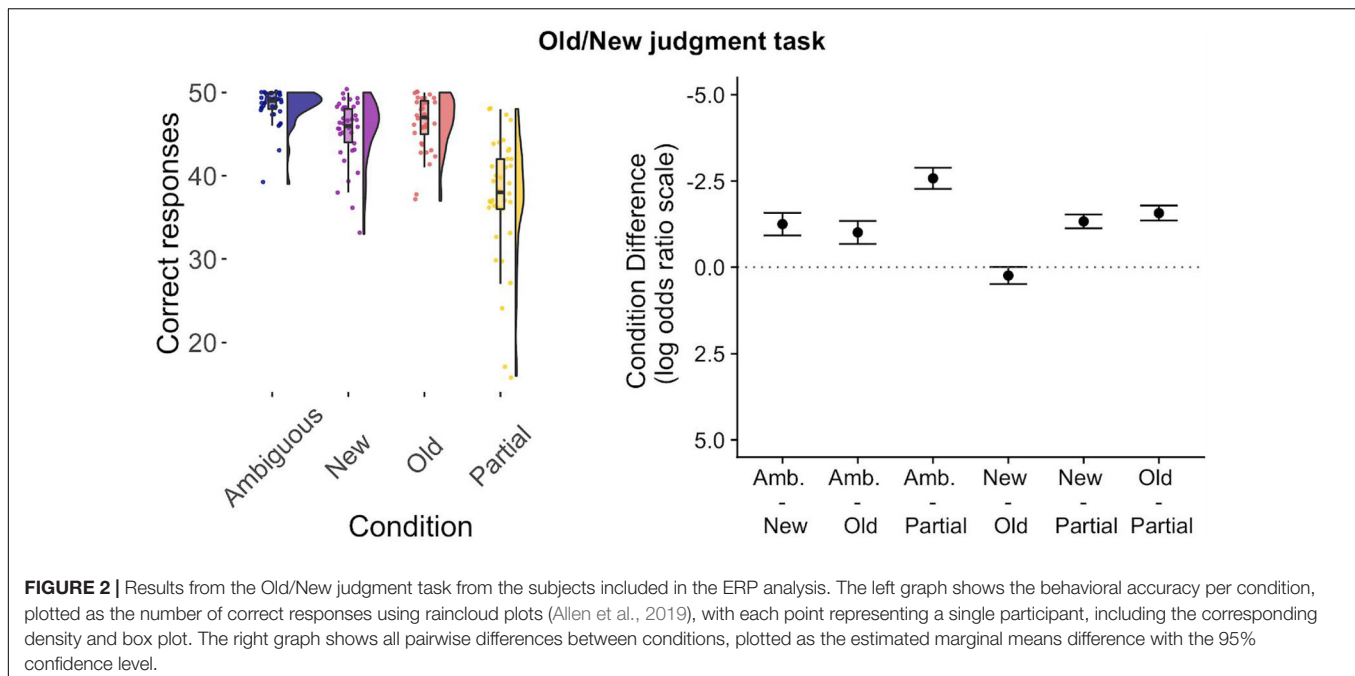
### Old/New Judgments

Participants responded most accurately to ambiguous nouns, then to old nouns, new nouns and partially matching nouns (**Figure 2**; this figure and the analysis only includes participants used in the ERP analysis, average number of trials per conditions is  $M = 48.3, 46.3, 45.4$ , and  $37.5$ , respectively). Our analysis

<sup>9</sup>For data manipulation, analysis and visualization, we used the following packages: `dplyr` (Wickham et al., 2019), `gdata` (Warnes et al., 2017), `tidyverse` (Wickham, 2017), `tidyr` (Wickham and Henry, 2019), `Rmisc` (Hope, 2013), `ggplot2` (Wickham, 2016), `cowplot` (Wilke, 2019), `lme4` (Bates et al., 2015), `lmerTest` (Kuznetsova et al., 2017), `emmeans` (Lenth, 2019).

<sup>10</sup>This correction was not pre-registered but requested by a reviewer.





revealed a strong effect of condition ( $\chi^2 = 517.06$ ,  $p < 0.001$ ) and differences between all pairs of conditions, with the strongest effects seen in comparison to the partially matching condition.

### Pre-registered ERP Analyses

#### N400 (300–500 ms)

Our experimental manipulation was associated with modulations of activity in the N400 region of interest ( $\chi^2 = 196.18$ ,  $p < 0.001$ ), with most negative amplitude elicited by new nouns, followed by partially matching, old and ambiguous nouns in that order (Figure 3; ERP waveforms at all individual channels are shown in Supplementary Figure 1). Pairwise follow-up tests revealed reliable differences between all conditions (Figure 4).

#### LPC (500–1000 ms)

Our experimental manipulation was also associated with modulations of activity in the subsequent LPC time window ( $\chi^2 = 13.311$ ,  $p = 0.004$ ; Figures 3, 4 and Supplementary Figure 1). This effect mostly reflected a carry-over effect from the enhanced N400 to new nouns, as the pairwise follow-ups showed that while new nouns elicited reliably more negative voltage than the other three conditions (although for partially matching nouns, this difference was not statistically significant after multiple comparisons correction), these other conditions did not reliably differ from each other.

#### Nref (300–1500 ms)

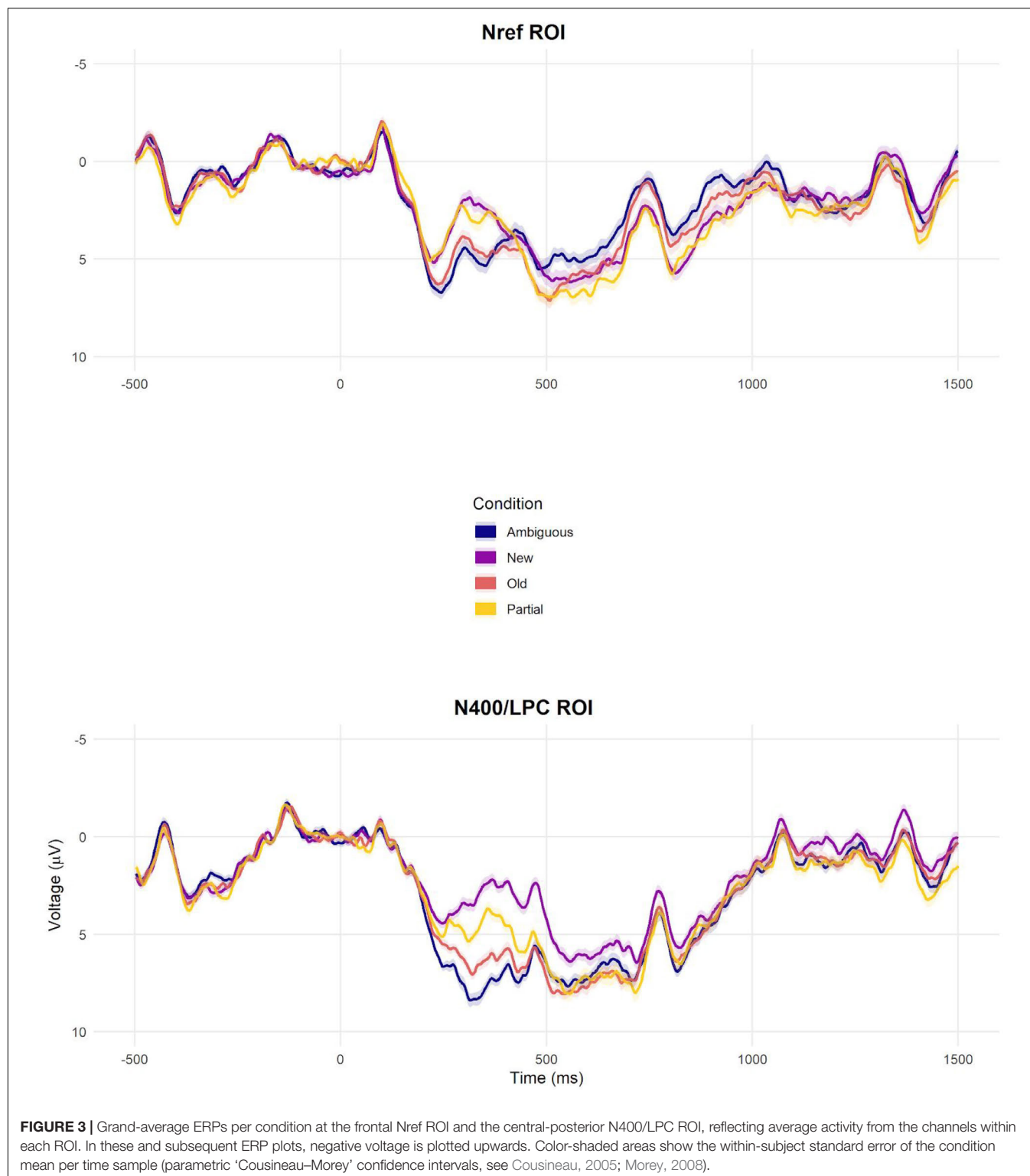
At the frontal ROI, ambiguous nouns elicit more negative voltage compared to the other conditions ( $M = -0.32$ ,  $S.E. = 0.28$ ; Figures 3, 4 and Supplementary Figure 1), compatible with an Nref effect, but this contrast did not reach the conventional  $\alpha = 0.05$  criterion ( $\chi^2 = 1.3$ ,  $p = 0.25$ ).

### Exploratory ERP Analyses

Our pre-registered ERP analyses showed that EEG activity was most sensitive to whether or not the critical noun had featured in the spoken story context, but did not differentiate anaphoric nouns and new nouns. Although amplitude in the N400 ROI differentiated between all four conditions, this pattern could merely reflect the relative ease of accessing the meaning of a noun that is more strongly related to context words, in other words, it need not reflect the process of anaphor resolution. Moreover, the smallest N400 was obtained for the ambiguous condition, wherein anaphor resolution was not straightforward. Likewise, we did not obtain a clear pattern of correlation between anaphor resolution and modulation of the LPC in the pre-registered ROI. We offer further discussion of these results in the Section “Discussion.”

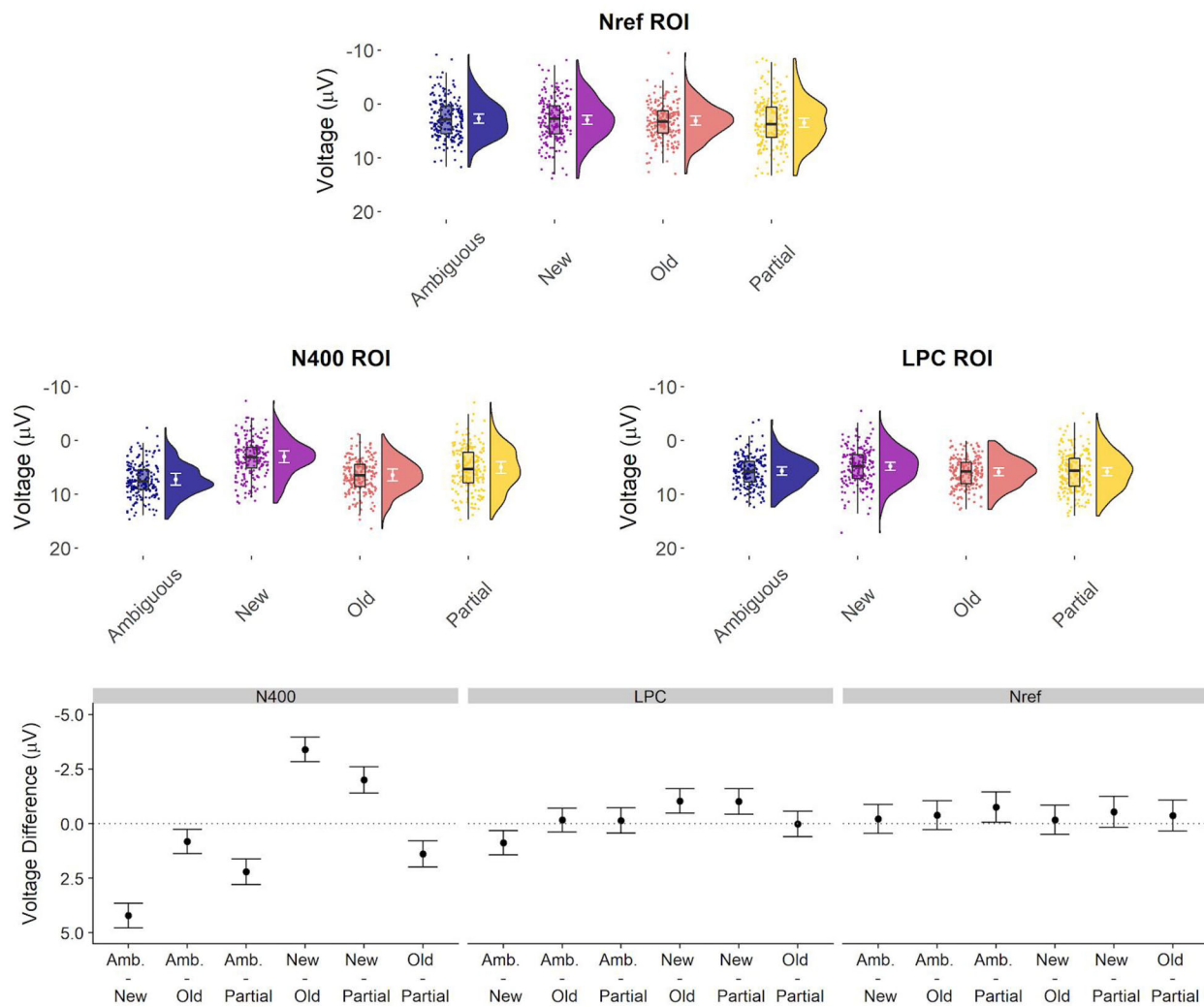
We considered the possibility that our participants used a strategy whereby they based their initial interpretation on whether the noun had been heard before (old/ambiguous versus new/partial), and subsequently changed this initial interpretation if the new noun could plausibly refer back to an antecedent (partial versus new). Such a strategy could be associated with an ERP effect of partial-matching nouns in a different ROI than the one we pre-registered. We tested for such an effect in two exploratory ERP analyses.

Our first exploratory analysis employed a mass regression approach (e.g., Groppe et al., 2011; Nieuwland et al., 2019) to test for later effects in the data segments from the pre-registered analysis. We down-sampled the data to 100 Hz and then ran a mixed-effects model analysis to test the contrast partial-match against the mean of the other three conditions at each electrode channel and at each data point between  $-500$  ms before to  $1500$  ms after noun onset. This yielded an effect estimate and standard error for each timepoint and channel. The associated



$p$ -values in the post-N400 window (from 500 to 1500 ms after noun onset) were corrected for multiple comparisons using the Benjamini and Hochberg method to control the false discovery rate (Benjamini and Hochberg, 1995). The resulting estimates are plotted as ERPs along with the corrected  $p$ -values

(Figure 5 for an ROI-based plot, and Supplementary Figure 2 for a plot of activity at all individual channels and highlighting of statistically significant samples after multiple comparison correction), revealing that partially matching nouns elicited more positive voltage than the other three conditions, particularly at



**FIGURE 4 |** Voltage per condition in the pre-registered ROIs and pairwise comparisons. **Upper graphs:** observations for each condition at the three ROIs using raincloud plots (Allen et al., 2019), with each point representing mean voltage associated with one item (averaged over subjects), the corresponding density and box plot, and the estimated marginal means per condition with the 95% confidence level (plotted in white on top of the density plot). **Lower graphs:** pairwise contrasts for each ROI with 95% confidence level.

middle-frontal, right-frontal and right-central channels in the post-N400 time window. Of note, the ROIs in **Figure 5** contain different numbers of channels.

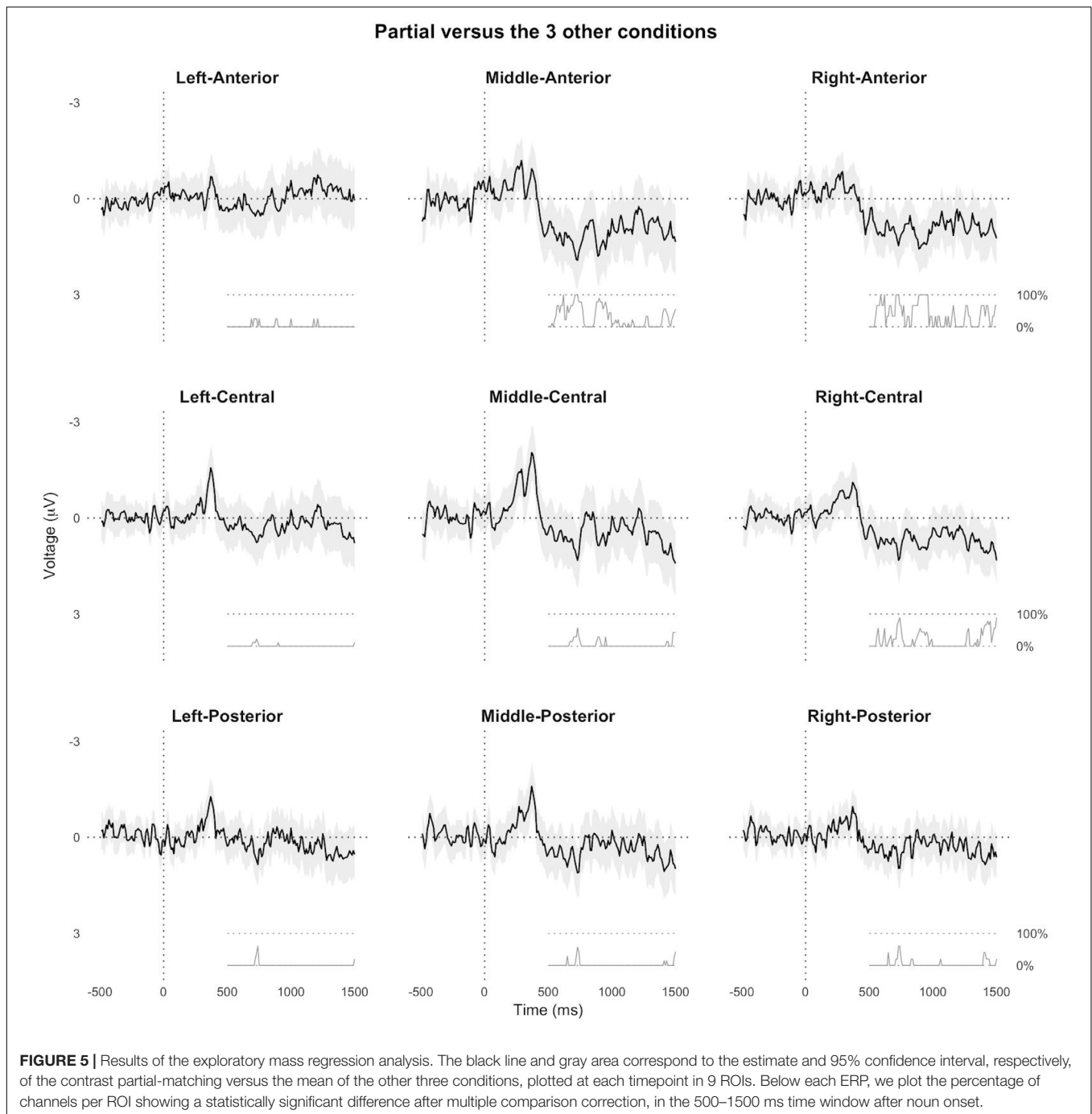
We performed similar analyses that directly compared partially matching nouns to only new or old nouns, and new nouns to old nouns (**Figure 6** and **Supplementary Figures 3–5**). These results suggest that the processing consequences of the partial match condition extended beyond the pre-registered ROI, and that partially matching nouns and new nouns both elicited a frontal positive ERP effect compared to old nouns in the post-N400 window around 500–1000 ms.

Our second exploratory analysis involved activity elicited by sentence-final words, to which we applied the same pre-processing steps as to the critical nouns (except that we segmented epochs of shorter duration, until 800 ms after word onset). As shown in **Figure 7** (and corresponding

**Supplementary Figure 6** showing ERP waveforms at all individual channels), partially matching nouns elicited more negative voltage than the other conditions. Using the N400/LPC spatial ROI, a contrast-based analysis showed more negative voltage for the partially matching nouns when compared to the mean of voltage for the other nouns ( $M = -0.48$ ,  $S.E. = 0.24$ ,  $t = 2.01$ ,  $p = 0.044$ ). This pattern is compatible with a sentence-final N400 effect, which extended beyond 500 ms after word onset (see also Nieuwland, 2014). In sum, both our exploratory analyses suggested enhanced processing difficulty associated with partial-matching nouns that extended up to the end of the sentence.

### Pre-registered Time-Frequency Analyses

As shown in **Figure 8**, all the conditions elicited a visually salient, relative power increase in the theta band in the first 500 ms



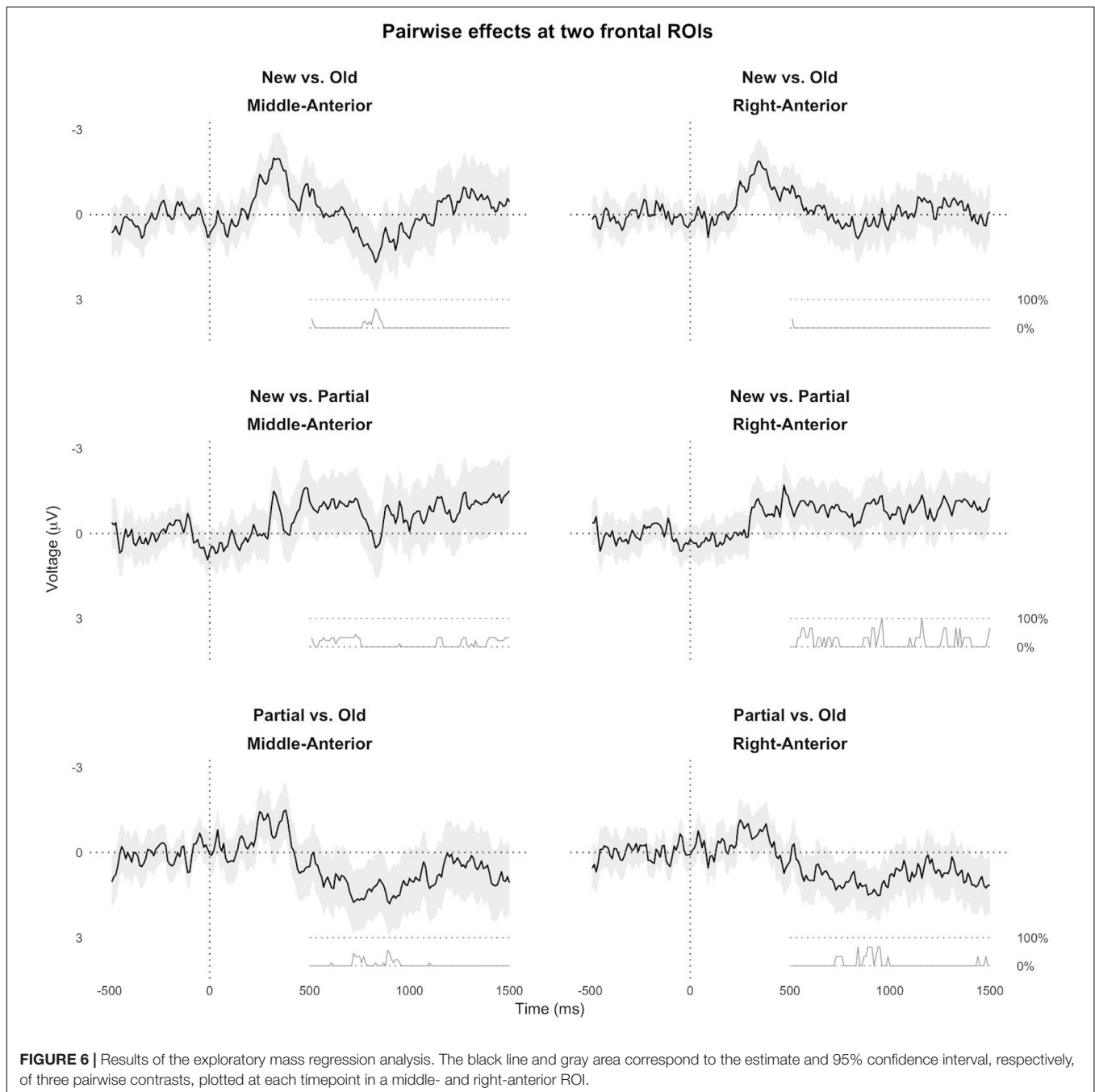
after noun onset, and a subsequent power decrease in the beta (10–15 Hz) band that extended until approximately 1300 ms after nouns onset. Patterns in the high frequency range were less pronounced.

As shown in **Figure 9**, the pairwise contrasts showed activity differences in the pre-registered ROIs but also in the beta range. In the theta (4–7 Hz) ROI, the contrasts Old-New, Old-Partial, Ambiguous-New, and Ambiguous-Partial showed significant differences (**Table 2**): new and partially matching nouns elicited greater theta power increases than old and

ambiguous nouns. Ambiguous nouns also elicited greater theta power than old nouns, suggested by a smaller yet sizeable cluster, although this contrast did not reach the  $\alpha = 0.05$  criterion. The results suggested no clear difference between partially matching and new nouns.

In the low gamma (35–45 Hz) ROI, new nouns elicited greater power than old nouns in the 400–600 ms time window after critical word onset (**Table 3**). Partially matching and ambiguous nouns also elicited greater low gamma power than old nouns, although these clusters did not reach the  $\alpha = 0.05$  threshold.



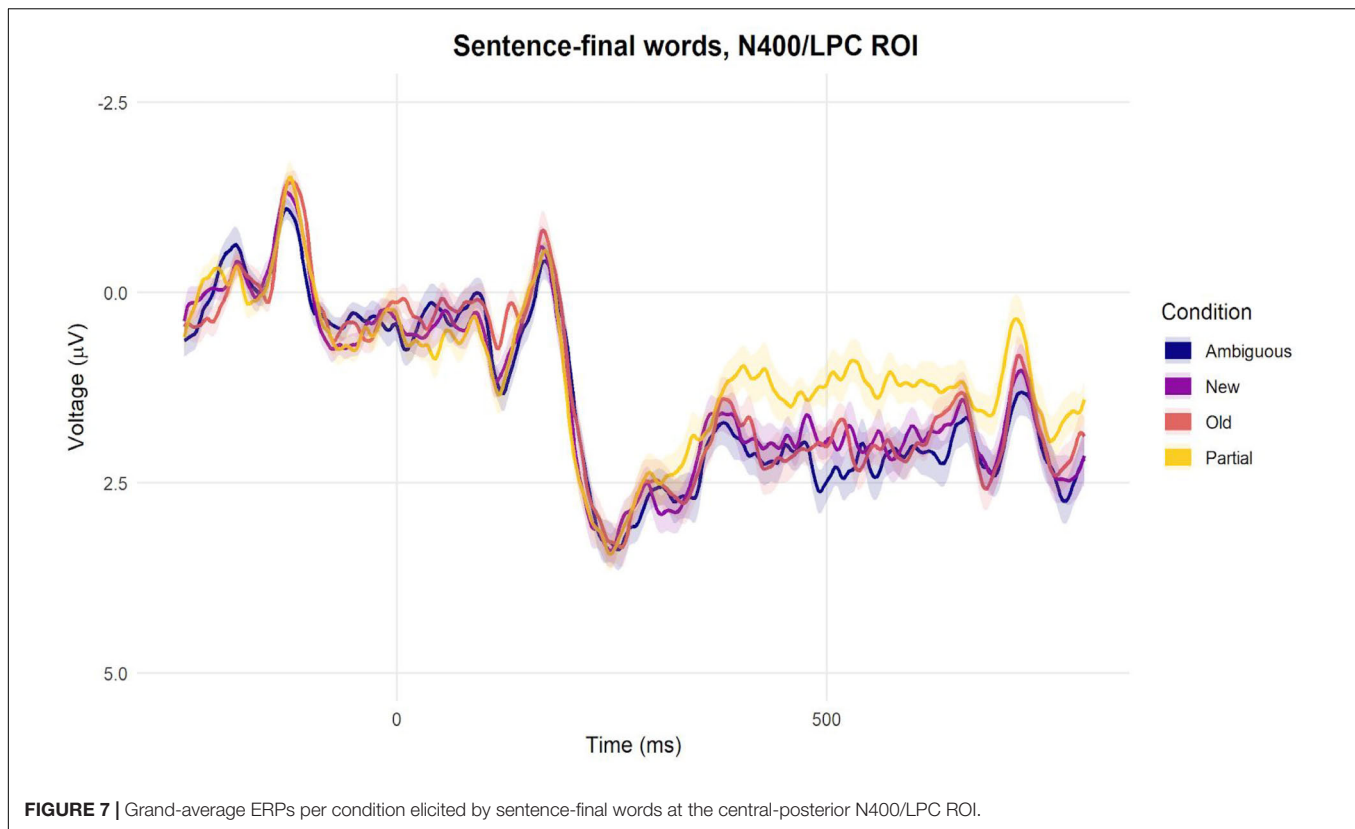


In the high gamma (60–80 Hz) ROI, there were no significant differences in the 500–1000 ms time interval after critical word onset (**Table 4**), although a sizeable cluster that did not reach the conventional threshold suggested more power for partially matching nouns compared to old nouns.

Our pre-registration also included additional analyses of a more exploratory nature that tested for effects in the 200–1500 ms time window after noun onset without averaging over time or frequency, for lower (2–30 Hz) and higher (30–90 Hz) frequencies separately. This analysis revealed six significant clusters (**Supplementary Table 1**), all of

which were in the low (2–30 Hz) frequencies. However, some of the effects in this analysis were composed of seemingly unrelated clusters. For this reason, based on visual inspection, we performed an extra (exploratory) analysis which averaged over the beta (10–15 Hz) frequency range within the 0–1500 ms time window after critical word onset.

This analysis revealed four clusters with greater power for old and ambiguous nouns compared to new and partially matching nouns (**Table 5**). Visual inspection (**Figure 9**) indicates that these clusters were most prominent around 1000 ms after noun onset.



## Exploratory Time-Frequency Analyses

We performed two types of exploratory analysis. First, we tried to localize the sources of the obtained time-frequency effects using beamformer analysis (Groß et al., 2001; for a detailed description of the method as applied to similar data sets, see Nieuwland and Martin, 2017; Coopmans and Nieuwland, 2019<sup>11</sup>). For the theta effects, which were focused on the 350–850 ms interval after critical word onset, this analysis did not reveal any statistically significant clusters. For the beta effects, the analysis was focused on a 700–1200 ms time window after critical word onset. This suggested a distributed source ranging from (pre)frontal to temporal areas (see **Figure 10**), with a slight left hemispherical focus.

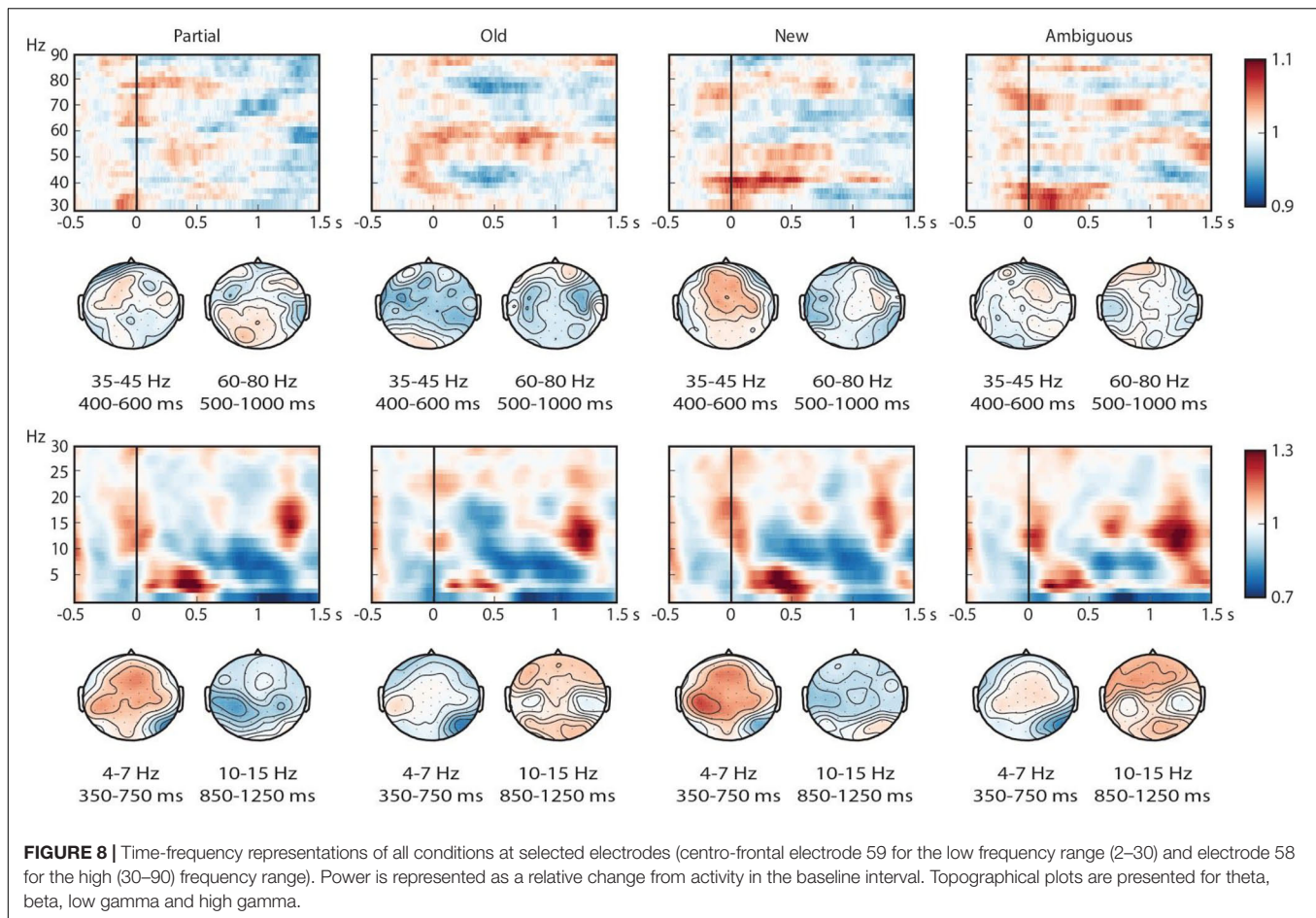
To ensure that the reported time-frequency effects in the 2–30 Hz frequency band provide information over and above the information found in the ERPs, we performed a second exploratory analysis. Similar to Bastiaansen et al. (2008), we tested whether the reported time-frequency effects could also be obtained from phase-locked activity alone by performing the same analysis on averaged ERPs per condition per subject (see Cohen, 2014, for limitations of this method). When a cluster is present in our pre-registered analysis, but absent in this phase-locked time-frequency analysis, we have greater evidence

that the observed effects are independent of the ERP effects. We found two 4–7 Hz theta-band effects (**Figure 11**), one in the Old-New contrast ( $p = 0.016$ ), and one in the Ambiguous-New contrast ( $p = 0.036$ ). Both of these clusters are in the same negative direction and in roughly the same time windows (around 400 ms after critical word onset) as the pre-registered theta effects. This means that for these contrasts, part of our effect in the theta-band is phase-locked. However, visual inspection of the time-frequency representations (**Figures 8, 11**) leads us to believe that not everything in the pre-registered theta cluster can be explained by the phase-locked information alone (i.e., the pre-registered theta clusters cover higher frequencies). The fact that the phase-locked effects are only present in 2 out of the 4 contrasts in which we found a significant cluster in the pre-registered analysis corroborates this line of reasoning.

## DISCUSSION

In this EEG study, we used ERP and time-frequency analyses to investigate the resolution of anaphoric noun phrases during discourse comprehension. We had a particular interest in how people resolve anaphors that are semantically related but different in form from the antecedent (e.g., *martian-alien*). Participants listened to story contexts that described two antecedents, and subsequently read a target sentence with a critical noun phrase. Depending on the story context, the critical noun phrase lexically matched one antecedent ('old'),

<sup>11</sup>The effects in the beta range covered a large number of areas. In order to identify where the effect was strongest, we adopted a conservative cut-off of  $\alpha = 0.005$  for data points to be subjected to the permutation analysis. All other settings were identical to those reported in Nieuwland and Martin (2017) and Coopmans and Nieuwland (2019).



matched two antecedents ('ambiguous'), partially matched one antecedent in terms of semantic features ('partial-match'), or introduced another referent (non-anaphoric, 'new'). After each story, participants judged whether the noun referred back to an antecedent (an 'old/new' judgment), and we used these responses to select trials in which participants arrived at the 'intended' interpretation ('old'/anaphoric for old, ambiguous and partially matching nouns, 'new'/non-anaphoric for new nouns) for further analyses.

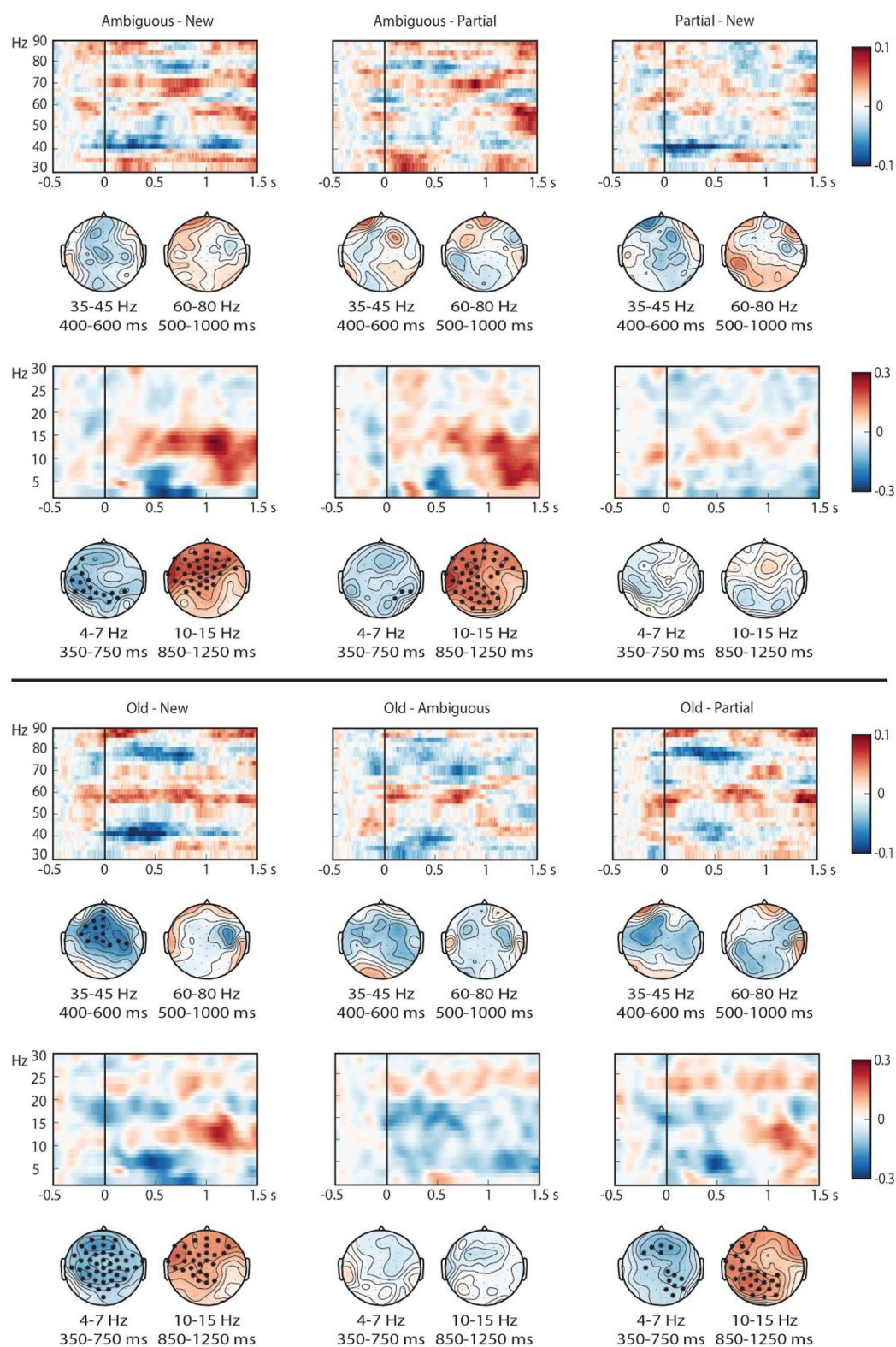
Pre-registered ERP analyses revealed modulation of the N400 ERP component by the status of the critical noun. We observed a stepwise decrease (becoming less negative) in N400 amplitude: the new condition had the highest N400 amplitude, then partially matching, old and finally, ambiguous nouns showed the lowest amplitude. We take this to reflect the context-based facilitation of access to the semantic meaning of the noun (e.g., Kutas and Federmeier, 2000; Burkhardt, 2006, 2007; Nieuwland and Van Berkum, 2008a; Lau et al., 2009). In addition, although we did not find an Nref effect that was statistically significant at the conventional  $\alpha = 0.05$  threshold, ambiguous nouns did elicit a sustained, frontal negativity compared to the other nouns, which is compatible with previous effects of referential processing difficulty (Van Berkum et al., 1999a, 2003; Nieuwland et al., 2007). Finally, additional exploratory ERP analyses revealed that

partially matching nouns and new nouns had similar positive ERP components in the early part of the post-N400 window, but that they diverged later on in the sentence and in response to sentence-final words.

Pre-registered time-frequency analyses were performed in theta, low gamma and high gamma ROIs. Theta effects were most pronounced and sensitive to whether or not the noun had been heard in the context, and did not differentiate partially matching nouns and new nouns. These theta effects could not entirely be explained as a time-frequency effect of the phase-locked ERP effects (see also Bastiaansen et al., 2005, 2008). Gamma effects were weak but suggested a power decrease for old nouns in the lower gamma frequency band (35–45 Hz). Exploratory time-frequency analyses further revealed strong differences between conditions in the beta (10–15 Hz) frequency range, primarily demonstrating sensitivity to whether or not the noun had occurred before. The time-frequency patterns therefore did not reveal a clear difference between partially matching and new nouns, as would be indicative of anaphor resolution.

The combination of our behavioral, ERP and time-frequency results suggests the cognitively demanding nature of resolving the anaphoric meaning of partially matching nouns. In the sections below, we will unpack this conclusion for both ERP and time-frequency results separately.





**FIGURE 9 |** Time-frequency representations of all pairwise contrasts at selected electrodes (centro-frontal electrode 59 for the low frequency range (2–30) and electrode 58 for the high (30–90) frequency range). Power is represented as a relative change from activity in the baseline interval. Scalp topographies for low and high gamma represent the activity within the preregistered time windows, and those for theta and beta reflect the time windows in which effects were most pronounced. Electrodes with significant differences in more than 60% of the attested time points are indicated with an \*.



**TABLE 2 |** Time-frequency effects in the theta range (4–7 Hz) occurring in the 0–1000 ms time window after noun onset.

	Cluster <i>t</i> -value	Cluster size	<i>p</i> -value
Old – New	–6002	1823	0.002/0.012
Old – Partial	–3127	1112	0.008/0.032
Old – Ambiguous	–1287	506	0.066/0.132
Ambiguous – New	–3562	1150	0.006/0.030
Ambiguous – Partial	–2552	877	0.010/0.032
Partial – New	–108	45	0.745/0.745

*In this and all following tables, the values correspond to the largest cluster that was found for each comparison. We report uncorrected/corrected *p*-values for each pairwise comparison. In this and following tables, for each test *df* = 34.*

**TABLE 3 |** Time-frequency effects in the lower gamma range (35–45 Hz) occurring in the 400–600 ms time window after noun onset.

	Cluster <i>t</i> -value	Cluster size	<i>p</i> -value
Old – New	–846	347	0.038/0.228
Old – Partial	–332	135	0.090/0.370
Old – Ambiguous	–354	149	0.074/0.370
Ambiguous – New	No cluster	No cluster	No cluster
Ambiguous – Partial	No cluster	No cluster	No cluster
Partial – New	No cluster	No cluster	No cluster

**TABLE 4 |** Time-frequency effects in the higher gamma range (60–80 Hz) occurring in the 500–1000 ms time window after noun onset.

	Cluster <i>t</i> -value	Cluster size	<i>p</i> -value
Old – New	–144	56	0.302/1.00
Old – Partial	–697	271	0.070/0.42
Old – Ambiguous	No cluster	No cluster	No cluster
Ambiguous – New	60	27	0.356/1.00
Ambiguous – Partial	No cluster	No cluster	No cluster
Partial – New	No cluster	No cluster	No cluster

**TABLE 5 |** Time-frequency effects in the 10–15 Hz time-frequency analysis of the 0–1500 ms time window after critical noun onset.

	Cluster <i>t</i> -value	Cluster size	<i>p</i> -value
Old – New	3955	1451	0.010/0.032
Old – Partial	5032	1857	0.008/0.032
Old – Ambiguous	–1886	730	0.056/0.112
Ambiguous – New	1077	3751	0.002/0.012
Ambiguous – Partial	8974	3095	0.004/0.020
Partial – New	–247	97	0.599/0.599

## Interpretation of ERP Results

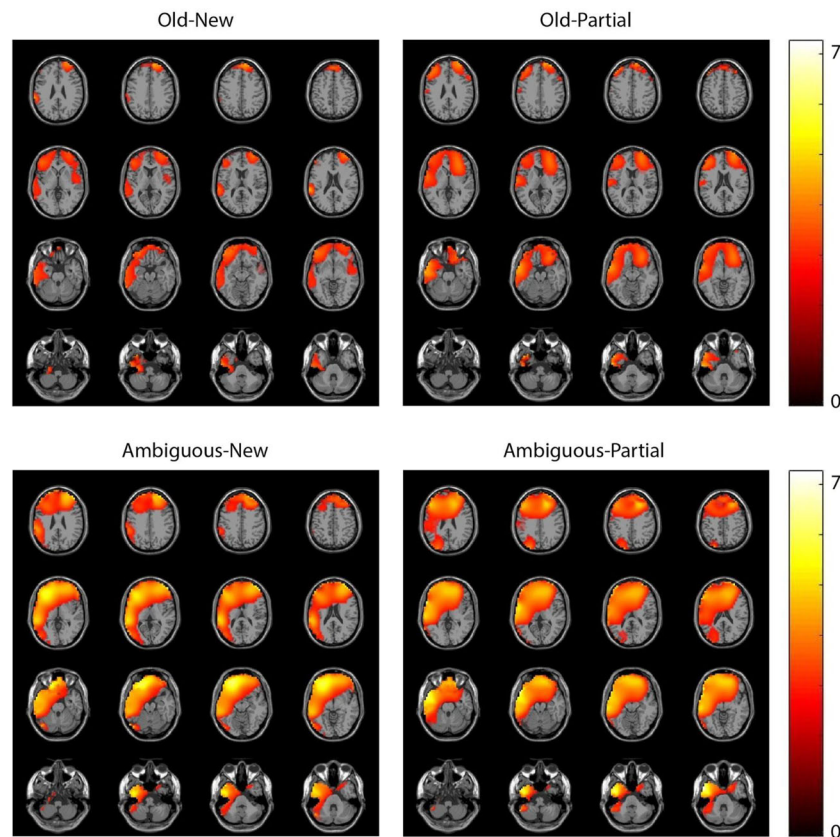
Our N400 results suggest that the semantic meaning of partially matching nouns was easier to access than that of new nouns, but harder to access than that of old or ambiguous nouns. Nevertheless, three distinct results in the later time windows suggest that the referential, anaphoric meaning of partially matching nouns may have been difficult to establish. Firstly, in approximately the 500–1000 ms time window after noun onset, partially matching nouns and new nouns both elicited enhanced positivity compared to ambiguous and old

nouns at the frontal channels (**Figure 6** and **Supplementary Figures 3, 4**), suggesting that partially matching nouns may have been initially considered as new, non-anaphoric nouns<sup>12</sup> (e.g., Burkhardt, 2006, 2007; Brouwer et al., 2012; Wang and Schumacher, 2013). Secondly, in an even later time window, approximately 1000–1500 ms, partially matching nouns elicited more positive voltage compared to old nouns and new nouns, while new nouns elicited more negative voltage than old nouns (**Figure 6** and **Supplementary Figures 3–5**). This late window thus revealed processing difficulty associated with partially matching nouns and with new nouns, but each with a distinct ERP profile (and thus presumably a distinct processing mechanism). Finally, ERPs elicited by sentence-final words suggested downstream processing difficulty for partially matching nouns compared to the other conditions (**Figure 7** and **Supplementary Figure 6**).

We think that the processing difficulty associated with partially matching nouns stems from the combination of the materials and the task. The old/new task might have focused the participant's attention on the lexical form of the words, rather than their referential meaning. For partially matching nouns, participants were required to remember two lexically different antecedents over the course of two spoken sentences, and then establish an anaphoric interpretation on yet another different word. Although the partially matching nouns were related in meaning to and sometimes synonymous with one antecedent, such anaphors (even the synonyms) may have been difficult to immediately recognize as such, especially in an experimental setting where the target noun on many trials introduced a new referent and where the task could have implied focus on lexical form. In comparison, the three other conditions were easier in terms of task demands. For ambiguous and old nouns, the task could be performed based on lexical repetition alone, and for ambiguous nouns participants only needed to remember one antecedent. The latter seemed to matter for the task, as participants were more accurate in recognizing ambiguous nouns than old nouns. For new nouns, participants only needed to remember one antecedent, and they could often rely on coarse semantic cues that ruled out an anaphoric interpretation, such as animacy or biological gender, or on semantic role information (e.g., patient–doctor).

Several patterns in our results suggest that although participants did ultimately establish the anaphoric meaning of partially matching nouns, they may have initially treated them as new, perhaps as part of a strategy that focused first on identifying lexical repetition and subsequently resolving the anaphor based on meaning. For example, new and partially matching nouns elicited a similar frontal, post-N400 positive effect compared to

<sup>12</sup> As pointed out by a reviewer, the observed frontal positive ERP effect may be due to a general unexpectedness of non-repeated nouns (e.g., Van Petten and Luka, 2012), rather than to the introduction of a new referent *per se*. Indeed, a cloze completion test on a subset of 12 items in 12 participants, in which we counted repeated nouns (regardless of a preceding adjective), the expectancy of a repeated noun anaphor was relatively high (72% cloze, range across conditions 69–75%, across items 43–100%, across subjects 50–100%, all cloze data is on our OSF page). In our study, therefore, we cannot distinguish novelty from unexpectedness.



**FIGURE 10 |** Beamformer source localization of the pair-wise beta effects (10–15 Hz). Colorbar represents *t*-values, masked for significance.

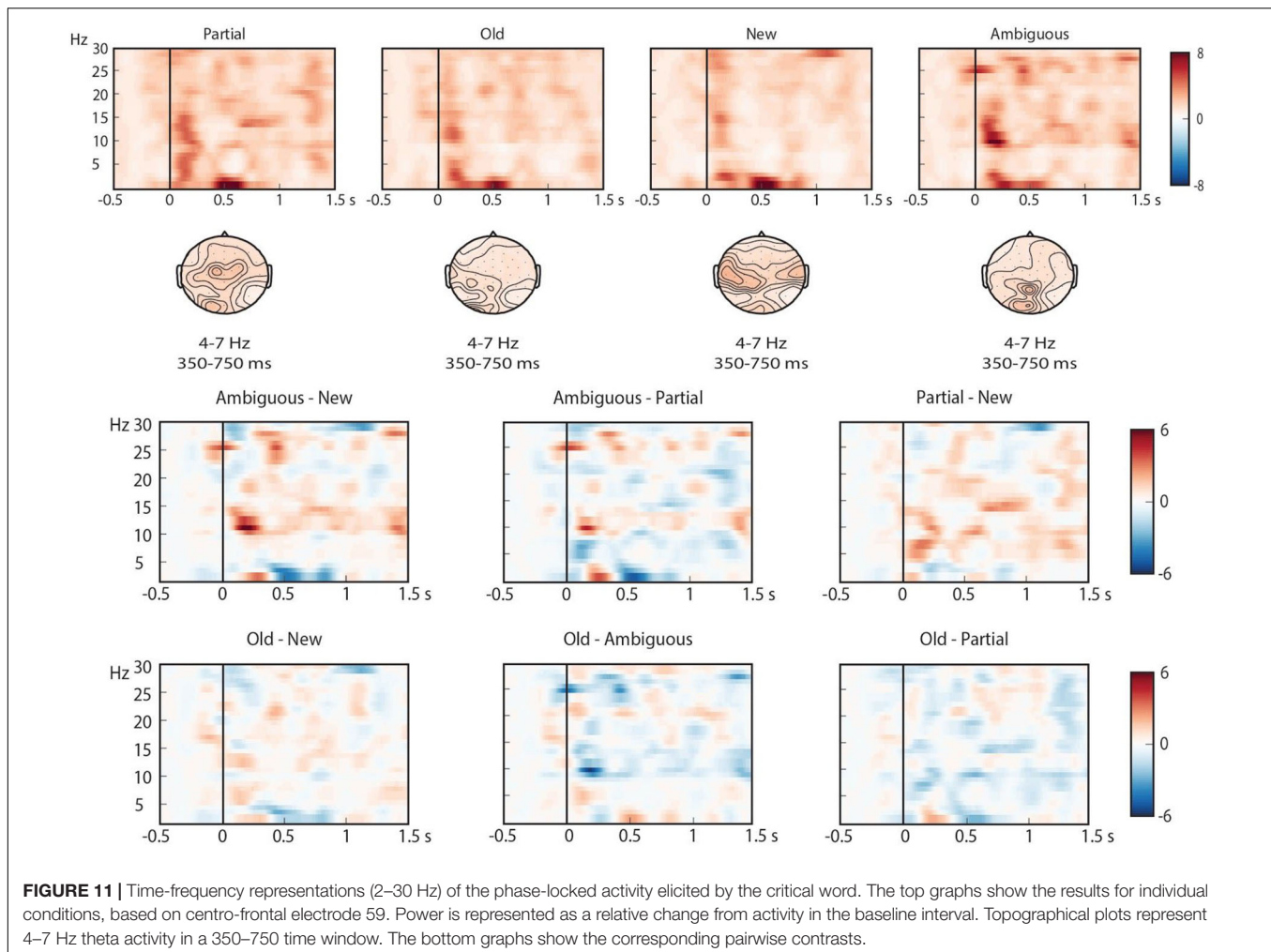
old nouns. This effect could be linked to the introduction of a new referent (discourse updating; Burkhardt, 2006), but, alternatively, may simply be due the unexpectedness of these nouns. Likewise, as discussed in the next section, the time-frequency results did not reveal clear differences between partially matching and new nouns. If participants switched from a non-anaphoric to an anaphoric interpretation (from ‘new’ to ‘old’) later on in the sentence, this could have caused difficulty keeping up with the remainder of the unfolding sentence. Compatible with this idea, sentence-final words following partially matching nouns elicit an N400-like effect compared to the other three conditions. Several studies have reported N400-like negativities for sentence-final words of unexpected or otherwise difficult sentences (Anderson and Holcomb, 2005; Paczynski and Kuperberg, 2012; Nieuwland, 2014; Vega-Mendoza et al., 2018), suggestive of continued sentence comprehension difficulty. Such effects may be more pronounced when participants perform a meta-linguistic judgment task (Nieuwland, 2014; Vega-Mendoza et al., 2018).

We emphasize that although participants in our experiments may have found it cognitively demanding to resolve partially matching anaphors, it is unclear whether this generalizes to regular language settings, where preceding discourse and surrounding visual context often facilitate anaphor resolution, or to a situation where the context only contains a single antecedent

(for discussion, see Dell et al., 1983; O’Brien et al., 1986). Likewise, it is possible that without the explicit task in our experiment to create anaphoric relations, participants would arrive at a non-anaphoric interpretation for partially matching nouns more often or even most of the time (see also O’Brien et al., 1997; Levine et al., 2000; Klin et al., 2004; Klin et al., 2006).

One further aspect of our ERP results is noteworthy, namely that while ambiguous nouns did not elicit robust Nref effects, they elicited less negative voltage in the N400 ROI compared to old nouns. The latter pattern may be caused by the noun repetition in the story context, because two identical context nouns may lead to a stronger repetition priming effect than a single noun (Van Petten et al., 1991). Previous studies did not observe such an effect, perhaps because they did not use identical context nouns (e.g., Van Berkum et al., 1999a, 2003; Nieuwland et al., 2007; Nieuwland and Van Berkum, 2008a), but instead used constructions such as “one alien who...and another one who.” Moreover, as noted earlier, remembering one antecedent was easier than two, as suggested by the recognition task results<sup>13</sup>.

<sup>13</sup>We considered the possibility that the reduced N400 for ambiguous nouns is in fact an enhanced positivity associated with easier task performance, but this pattern is difficult to reconcile with the other N400 patterns, such as the smaller N400 for partially matching nouns compared to new nouns despite the fact that partial-match nouns were more difficult to evaluate.



In sum, our ERP analyses generated a varied range of effects. While our results showed relatively clear effects associated with referent activation, they are somewhat inconclusive in the sense that we could not conclusively tie any single effect specifically to the difference between old or new referents (discourse updating). This may have had to do with the task demands of our experiment, and with the fact that old and partially matching anaphors showed little similarity in brain responses despite being both interpreted as anaphoric.

## Time-Frequency Results

Whereas the ERP results clearly differentiated old from ambiguous nouns, and partially matching from new nouns, the time-frequency results primarily yielded effects of lexical repetition: effects of old/ambiguous versus new/partially matching, with some evidence for a difference between old and ambiguous nouns (which differed in number of repetitions), but no clear difference between new and partially matching nouns (which were both lexically new and thus did not differ in repetition). The observed effects were strong in the theta and beta frequency range, but much less so in the gamma frequency range. The time-frequency analysis alone therefore

did not allow us to identify activity that might be related to resolution of partially matching nouns, and this suggests that ERPs are more sensitive to these processes. However, we emphasize that time-frequency analysis typically requires a larger number of trials than ERP analysis to obtain stable estimates (e.g., Bastiaansen et al., 2013). Our data contained relatively low trial numbers in particular for partially matching nouns, which received the lowest number of correct ‘old’ responses. This will have decreased our ability to pick up on relevant differences.

We found greater theta (and, to a lesser extent, gamma) power for new/partially matching nouns than for old/ambiguous nouns. These patterns clearly differ in their directionality and functionality from recent findings on proper name anaphors (Coopmans and Nieuwland, 2019), which revealed increased theta (and to a lesser extent, low gamma) for old/repeated compared to new proper names. The theta effects in these studies also differ in the frequency range they appear to cover. It is possible that these differences somehow stem from the differences in anaphor type, in particular because proper names (of unfamiliar discourse characters) contain much less semantic content than noun phrases.



One possibility is that theta power correlates with the amount of semantic information that is retrieved from long-term memory (e.g., Bastiaansen et al., 2005, 2008). In Coopmans and Nieuwland (2019), this would not differ between old and new proper names, perhaps because the names themselves contain little semantic content. For new noun phrases in the current study, however, the full meaning of the word will be retrieved, whereas for old noun phrases most of the relevant meaning may already be active due to the first presentation. Another difference was that the stimuli used by Coopmans and Nieuwland were all written, whereas the current study combined spoken with written language. It is possible that theta effects are sensitive not only to lexical repetition but also to repetition of form. Beyond these differences in anaphor type and modality, other differences in terms of task demands may be relevant too. For example, participants in our experiment may have focused strongly on word repetition to perform the task, at the expense of attention to the meaning of the unfolding story. Our time-frequency effects may thus be related to repetition priming effects (e.g., Gruber and Müller, 2004), which could explain why we also obtained power differences between old and ambiguous nouns (which differed in number of repetitions). At any rate, our results demonstrate that theta and gamma effects do not depend on anaphoricity alone. This might make their use to study anaphor comprehension less straightforward than previously suggested (Nieuwland and Martin, 2017; Coopmans and Nieuwland, 2019), although it remains unclear to what extent the observed theta/gamma effects are driven by the task demands. A dedicated follow-up study could shed light on this issue by directly comparing repetition/anaphoricity effects for proper names and noun phrases, or, for instance, by directly manipulating the semantic distance of old and new nouns.

While the effects in the theta frequency band were relatively strong, effects in the gamma range were very weak and inconclusive. One explanation for this lack of results is that there is relatively lower power in the gamma band compared to lower frequency bands, which may make it rather hard to obtain clear gamma effects with a low number of trials, as in the current study. Another explanation could be that gamma activity is primarily sensitive to sentence/discourse-level semantic integration costs (e.g., Peña and Melloni, 2012; Rommers et al., 2013; Fedorenko et al., 2016; Nieuwland and Martin, 2017; Coopmans and Nieuwland, 2019), which was not manipulated in our experiment (in contrast to, for example, a comparison between semantically incongruent and congruent words, see Coopmans and Nieuwland, 2019).

In addition to the effects in the pre-registered theta and gamma ROIs, we found greater beta (~10–15 Hz) power for old/ambiguous nouns than for new/partially matching nouns, and to some extent for ambiguous nouns compared to old nouns. Beamformer source localization suggested a fairly widely distributed, prefrontal/temporal source with a left hemisphere bias. Beta effects have previously been observed in a wide range of language comprehension studies (for a review, see Weiss and Mueller, 2012; Lewis et al., 2016). One proposal is that beta power is related to maintenance/changes in the current mode

of processing and representation of a sentence-level meaning (Lewis et al., 2016), which is based on observed decreases in beta power to unexpected stimuli (e.g., Engel and Fries, 2010). Our results seem compatible with this proposal. Another proposal is that beta synchronization serves to bind distributed sets of neurons into a coherent representation of (memorized) contents during language processing (Weiss and Mueller, 2012).

We refrain from claims about the functional significance of these unanticipated effects. Moreover, we emphasize the fact that, in terms of condition-wise patterns, beta power behaved in largely the same way as theta power, which complicates a functional differentiation of these frequency bands. None of the frequency bands clearly differentiated new from partially matching nouns and could therefore be linked to the difference between anaphoric and non-anaphoric meaning, and all of the frequency bands showed some sensitivity to the difference between old and ambiguous names, suggesting sensitivity to either lexical repetition or to the task demands. What does differ between the frequency bands, however, is the directionality of the effects (increased beta power but decreased theta/gamma power for repeated nouns compared to non-repeated nouns; see Lundqvist et al., 2011, for a similar distinction between these frequency bands in relation to working memory load), the timing of the effects (theta and gamma effects occurred within roughly the first 1000 ms after noun onset, beta effects occurred later), and possibly the underlying neural source of these effects.

In sum, as with the ERP results, our time-frequency results did not allow us to tie one specific effect to anaphoric meaning, and they were chiefly driven by noun repetition. We suspect that the task demands of our experiment were the main driving force behind these effects.

## CONCLUSION

The flexible nature of human language allows people to establish referential relationships between words that differ in meaning. Very little work to date has examined the neural processes that may underlie such anaphoric interpretations. We addressed this issue in an EEG study on discourse comprehension, wherein we investigated the ERP and time-frequency correlates of how people resolve noun phrases, and in particular how they resolve anaphoric nouns that either lexically match or mismatch the intended antecedent. The N400 ERP component demonstrated initial sensitivity to noun repetition and semantic overlap, corresponding to repetition and semantic priming effects, respectively. A subsequent frontal positivity demonstrated sensitivity to whether the noun had been repeated, suggesting that partially matching anaphors may have been processed as new nouns temporarily. ERPs in even later time windows and ERPs time-locked to sentence-final words suggested that partially matching nouns and new nouns had different effects on comprehension. In contrast to the ERP results, the time-frequency results primarily demonstrated sensitivity to noun repetition, and did not differentiate partially matching anaphors from new nouns. In sum, our results show the ERP and time-frequency effects of referent repetition during



discourse comprehension, and demonstrate the potentially demanding nature of establishing the anaphoric meaning of a novel noun.

## DATA AVAILABILITY STATEMENT

In accordance with the Peer Reviewers' Openness Initiative (<https://opennessinitiative.org>, Morey et al., 2016), all materials (data, materials, scripts, figures, and supplementary figures) associated with this manuscript are available on <https://osf.io/uak8g/>.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee for Behavioural Research of the

Social Sciences Faculty at Radboud University Nijmegen. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MN designed the experiment and wrote the manuscript. CC and RS collected the data and provided crucial edits. All authors analyzed the data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00398/full#supplementary-material>

## REFERENCES

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., and Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res.* 4:63. doi: 10.12688/wellcomeopenres.15191.1
- Almor, A. (1999). Noun-phrase anaphora and focus: the informational load hypothesis. *Psychol. Rev.* 106, 748–765. doi: 10.1037//0033-295x.106.4.748
- Almor, A., and Nair, V. A. (2007). The form of referential expressions in discourse. *Lang. Linguist. Compass* 1, 84–99. doi: 10.1111/j.1749-818X.2007.00009.x
- Anderson, J. E., and Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain Lang.* 94, 200–216. doi: 10.1016/j.bandl.2005.01.001
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Baggio, G. (2019). *Meaning in the Brain*. Cambridge, MA: MIT Press.
- Baggio, G., and Hagoort, P. (2011). The balance between memory and unification in semantics: a dynamic account of the N400. *Lang. Cogn. Process.* 26, 1338–1367. doi: 10.1080/01690965.2010.542671
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bastiaansen, M., and Hagoort, P. (2003). Event-induced theta responses as a window on the dynamics of memory. *Cortex* 39, 967–992. doi: 10.1016/s0010-9452(08)70873-6
- Bastiaansen, M., Mazaheri, A., and Jensen, O. (2013). "Beyond ERP's: oscillatory neuronal dynamics," in *The Oxford Handbook of EventRelated Potential Components*, eds S. J. Luck, and E. S. Kappenman (Oxford, UK: Oxford University Press), 31–49.
- Bastiaansen, M. C. M., Linden, M., Van der Keurs, M., Ter Dijkstra, T., and Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *J. Cogn. Neurosci.* 17, 530–541. doi: 10.1162/08989290532792729469
- Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., and Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain Lang.* 106, 15–28. doi: 10.1016/j.bandl.2007.10.006
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). lme4: linear mixed-effects models using Eigen and S4. *R package version 1.1–7*, 2014.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. BMethodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Besson, M., Kutas, M., and Petten, C. V. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *J. Cogn. Neurosci.* 4, 132–149. doi: 10.1162/jocn.1992.4.2.132
- Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer [Computer program]. Version 5.3.51*.
- Boudewyn, M. A., Long, D. L., Traxler, M. J., Lesh, T. A., Dave, S., Mangun, G. R., et al. (2015). Sensitivity to referential ambiguity in discourse: the role of attention, working memory, and verbal ability. *J. Cogn. Neurosci.* 27, 2309–2323. doi: 10.1162/jocn\_a\_00837
- Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055
- Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: evidence from event-related brain potentials. *Brain Lang.* 98, 159–168. doi: 10.1016/j.bandl.2006.04.005
- Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport* 18, 1851–1854. doi: 10.1097/WNR.0b013e328f1a999
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Cabeza, R., Ciaramelli, E., Olson, I. R., and Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nat. Rev. Neurosci.* 9, 613–625. doi: 10.1038/nrn2459
- Clark, H. H., and Murphy, G. L. (1982). Audience design in meaning and reference. *Adv. Psychol.* 9, 287–299. doi: 10.1016/s0166-4115(09)60059-5
- Clark, H. H., and Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Mem. Cogn.* 7, 35–41. doi: 10.3758/bf03196932
- Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. Cambridge, MA: MIT Press.
- Coopmans, C. W., and Nieuwland, M. S. (2019). Dissociating activation and integration of discourse referents: evidence from ERPs and oscillations. *bioRxiv* [Preprint]. doi: 10.1101/671933
- Core Team, R. (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* 1, 42–45. doi: 10.20982/tqmp.01.1.p042
- Dell, G. S., McKoon, G., and Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *J. Verbal Learn. Verbal Behav.* 22, 121–132. doi: 10.1016/s0022-5371(83)80010-3

- Delogu, F., Brouwer, H., and Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain Cognition* 135:103569. doi: 10.1016/j.bandc.2019.05.007
- Engel, A. K., and Fries, P. (2010). Beta-band oscillations — signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565
- Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., et al. (2016). Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6256–E6262.
- Fraurud, K. (1990). Definiteness and the processing of noun phrases in natural discourse. *J. Semant.* 7, 395–433. doi: 10.1093/jos/7.4.395
- Garnham, A. (1989). “Integrating information in text comprehension: the interpretation of anaphoric noun phrases,” in *Linguistic Structure in Language Processing*, eds C. K. Greg, and N. T. Michael (Dordrecht: Springer), 359–399. doi: 10.1007/978-94-009-2729-2\_10
- Garnham, A. (2001). *Mental Models and the Interpretation of Anaphora*. Hove: Psychology Press.
- Garnham, A., Oakhill, J., and Cain, K. (1997). The interpretation of anaphoric noun phrases time course, and effects of overspecificity. *Q. J. Exp. Psychol. A* 50, 149–162. doi: 10.1080/713755687
- Garrod, S., Freudenthal, D., and Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *J. Mem. Lang.* 33, 39–68. doi: 10.1006/jmla.1994.1003
- Garrod, S., and Sanford, A. (1977). Interpreting anaphoric relations: the integration of semantic information while reading. *J. Verbal Learn. Verbal Behav.* 16, 77–90. doi: 10.1016/s0022-5371(77)80009-1
- Garrod, S. C., and Sanford, A. J. (1982). The mental representation of discourse in a focussed memory system: implications for the interpretation of anaphoric noun phrases. *J. Semant.* 1, 21–41. doi: 10.1093/jos/1.1.21
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition* 32, 99–156. doi: 10.1016/0010-0277(89)90001-2
- Gibson, E., and Pearlmuter, N. J. (eds) (2011). *The Processing and Acquisition of Reference*. Cambridge, MA: MIT Press.
- Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x
- Groß, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 98, 694–699. doi: 10.1073/pnas.98.2.694
- Gruber, T., and Müller, M. M. (2004). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb. Cortex* 15, 109–116. doi: 10.1093/cercor/bhh113
- Gundel, J. K., Hedberg, N., and Zacharski, R. (2001). Definite descriptions and cognitive status in English: why accommodation is unnecessary. *English Lang. Linguist.* 5, 273–295. doi: 10.1017/s1360674301000247
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423. doi: 10.1016/j.tics.2005.07.004
- Hagoort, P., and Indefrey, P. (2014). The neurobiology of language beyond single words. *Ann. Rev. Neurosci.* 37, 347–362. doi: 10.1146/annurev-neuro-071013-013847
- Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Doctoral Dissertation, University of Massachusetts, Amherst.
- Heine, A., Tamm, S., Hofmann, M., Bösel, R. M., and Jacobs, A. M. (2006). Event-related theta activity reflects memory processes in pronoun resolution. *Neuroreport* 17, 1835–1839. doi: 10.1097/WNR.0b013e328010a096
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- Hope, R. M. (2013). *Rmisc: Ryan Miscellaneous. R package version 1.5*.
- Jacobs, J., Hwang, G., Curran, T., and Kahana, M. J. (2006). EEG oscillations and recognition memory: theta correlates of memory retrieval and decision making. *Neuroimage* 32, 978–987. doi: 10.1016/j.neuroimage.2006.02.018
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178. doi: 10.1111/1469-8986.3720163
- Karimi, H., Swaab, T. Y., and Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *J. Mem. Lang.* 102, 68–82. doi: 10.1016/j.jml.2018.05.003
- Klin, C. M., Guzmán, A. E., Weingartner, K. M., and Ralano, A. S. (2006). When anaphor resolution fails: partial encoding of anaphoric inferences. *J. Mem. Lang.* 54, 131–143. doi: 10.1016/j.jml.2005.09.001
- Klin, C. M., Weingartner, K. M., Guzmán, A. E., and Levine, W. H. (2004). Readers’ sensitivity to linguistic cues in narratives: how salience influences anaphor resolution. *Mem. Cogn.* 32, 511–522. doi: 10.3758/bf03195843
- Krahmer, E., and Deemter, K. V. (1998). On the interpretation of anaphoric noun phrases: towards a full understanding of partial matches. *J. Semant.* 15, 355–392. doi: 10.1093/jos/15.4.355
- Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657
- Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lau, E., Almeida, D., Hines, P. C., and Poeppel, D. (2009). A lexical basis for N400 context effects: evidence from MEG. *Brain Lang.* 111, 161–172. doi: 10.1016/j.bandl.2009.08.007
- Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.5.1*.
- Levine, W. H., Guzmán, A. E., and Klin, C. M. (2000). When anaphor resolution fails. *J. Mem. Lang.* 43, 594–617. doi: 10.1080/17470218.2010.520414
- Lewis, A. G., Schoffelen, J. M., Schriefers, H., and Bastiaansen, M. (2016). A predictive coding perspective on beta oscillations during sentence-level language comprehension. *Front. Hum. Neurosci.* 10:85. doi: 10.3389/fnhum.2016.00085
- Lewis, A. G., Wang, L., and Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? *Brain Lang.* 148, 51–63. doi: 10.1016/j.bandl.2015.01.003
- Lundqvist, M., Herman, P., and Lansner, A. (2011). Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. *J. Cogn. Neurosci.* 23, 3008–3020. doi: 10.1162/jocn\_a\_00029
- Mandera, P., Keuleers, E., and Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: a review and empirical validation. *J. Mem. Lang.* 92, 57–78. doi: 10.1016/j.jml.2016.04.001
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Martin, A. E., Nieuwland, M. S., and Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage* 59, 1859–1869. doi: 10.1016/j.neuroimage.2011.08.057
- Martinich, A. P. (1985). *The Philosophy of Language*. Oxford, UK: Oxford University Press.
- McKoon, G., and Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *J. Verbal Learn. Verbal Behav.* 19, 668–682. doi: 10.1016/s0022-5371(80)90355-2
- McKoon, G., and Ratcliff, R. (1998). Memory-based language processing: psycholinguistic research in the 1990s. *Annu. Rev. Psychol.* 49, 25–42. doi: 10.1146/annurev.psych.49.1.25
- Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., and Friederici, A. D. (2015). Frontal–posterior theta oscillations reflect memory retrieval during

- sentence comprehension. *Cortex* 71, 205–218. doi: 10.1016/j.cortex.2015.06.027
- Mitra, P. P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophys. J.* 76, 691–708. doi: 10.1016/s0006-3495(99)77236-x
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Reason* 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Murphy, G. L. (1984). Establishing and accessing referents in discourse. *Mem. Cogn.* 12, 489–497. doi: 10.3758/bf03198311
- Myers, J. L., and O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Process.* 26, 131–157. doi: 10.1080/01638539809545042
- Nieuwland, M. S. (2014). Who's he?" Event-related brain potentials and unbound pronouns. *J. Mem. Lang.* 76, 1–28. doi: 10.1016/j.jml.2014.06.002
- Nieuwland, M. S., Barr, D., Bartolozzi, F., Busch-Moreno, S., Donaldson, D., Ferguson, H. J., et al. (2019). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *bioRxiv* [Preprint]. doi: 10.1101/267815
- Nieuwland, M. S., and Martin, A. E. (2017). Neural oscillations and a nascent corticohippocampal theory of reference. *J. Cogn. Neurosci.* 29, 896–910. doi: 10.1162/jocn\_a\_01091
- Nieuwland, M. S., Otten, M., and Van Berkum, J. J. A. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *J. Cogn. Neurosci.* 19, 228–236. doi: 10.1162/jocn.2007.19.2.228
- Nieuwland, M. S., and Van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: evidence from ERPs. *Brain Res.* 1118, 155–167. doi: 10.1016/j.brainres.2006.08.022
- Nieuwland, M. S., and Van Berkum, J. J. A. (2008a). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: evidence from ERPs. *Brain Lang.* 106, 119–131. doi: 10.1016/j.bandl.2008.05.001
- Nieuwland, M. S., and Van Berkum, J. J. A. (2008b). The neurocognition of referential ambiguity in language comprehension. *Lang. Linguist. Compass* 2, 603–630. doi: 10.1111/j.1749-818X.2008.00070.x
- O'Brien, E. J., Duffy, S. A., and Myers, J. L. (1986). Anaphoric inference during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 346–352. doi: 10.1037//0278-7393.12.3.346
- O'Brien, E. J., Raney, G. E., Albrecht, J. E., and Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Process.* 23, 1–24. doi: 10.1080/01638539709544979
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Paczynski, M., and Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *J. Mem. Lang.* 67, 426–448. doi: 10.1016/j.jml.2012.07.003
- Peña, M., and Melloni, L. (2012). Brain oscillations during spoken sentence processing. *J. Cogn. Neurosci.* 24, 1149–1164. doi: 10.1162/jocn\_a\_00144
- Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., et al. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proc. Natl. Acad. Sci. U.S.A.* 113, 11366–11371. doi: 10.1073/pnas.1603312113
- Poesio, M., and Vieira, R. (1998). A corpus-based investigation of definite description use. *Comput. Linguist.* 24, 183–216.
- Pyke, A., West, R. L., and LeFevre, J. A. (2007a). "How readers retrieve referents for nouns in real time: a memory-based model of context effects on referent accessibility," in *Proceedings of the International Conference on Cognitive Modeling* (New York, NY: Taylor & Francis), 7–12.
- Pyke, A., West, R. L., and LeFevre, J. A. (2007b). "On-line reference assignment for anaphoric and non-anaphoric nouns: a unified, memory-based model in ACT-R," in *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (Nashville, TN: Cognitive Science Society), 1403–1408.
- Pyke, A. A. (2007). *Discriminating Anaphoric and Non-Anaphoric Definite Nouns: A Unified Memory-Based Model*. Doctoral dissertation, Carleton University, Ottawa.
- Rommers, J., Dijkstra, T., and Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *J. Cogn. Neurosci.* 25, 762–776. doi: 10.1162/jocn\_a\_00337
- Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257. doi: 10.1016/j.tics.2007.04.004
- Schumacher, P. B., and Hung, Y.-C. (2012). Positional influences on information packaging: insights from topological fields in German. *J. Mem. Lang.* 67, 295–310. doi: 10.1016/j.jml.2012.05.006
- Swaab, T. Y., Camblin, C. C., and Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *J. Cogn. Neurosci.* 16, 715–726. doi: 10.1162/089892904970744
- Tyler, L. K. (1983). The development of discourse mapping processes: the on-line interpretation of anaphoric expressions. *Cognition* 13, 309–341. doi: 10.1016/0010-0277(83)90013-6
- Van Berkum, J. J. A., Brown, C. M., and Hagoort, P. (1999a). Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182. doi: 10.1006/jmla.1999.2641
- Van Berkum, J. J. A., Hagoort, P., and Brown, C. M. (1999b). Semantic integration in sentences and discourse: evidence from the N400. *J. Cogn. Neurosci.* 11, 657–671. doi: 10.1162/089892999563724
- Van Berkum, J. J. A., Brown, C. M., Hagoort, P., and Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology* 40, 235–248. doi: 10.1111/1469-8986.00025
- Van Berkum, J. J. A., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. doi: 10.1016/j.brainres.2006.06.091
- Van Berkum, J. J. A., Zwitserlood, P., Bastiaansen, M. C. M., Brown, C. M., and Hagoort, P. (2004). So who's "he" anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension. *Suppl. J. Cogn. Neurosci.* 16:70.
- Van Gompel, R. P., Liversedge, S. P., and Pearson, J. (2004). *Antecedent Typicality Effects in the Processing of Noun Phrase Anaphors. The On-line Study of Sentence Comprehension: Eyetracking, ERP and Beyond* (New York, NY: Psychology Press), 119–137.
- Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., and McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *J. Cogn. Neurosci.* 3, 131–150. doi: 10.1162/jocn.1991.3.2.131
- Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015
- Van Petten, C., and Senkfor, A. J. (1996). Memory for words and novel visual patterns: repetition, recognition, and encoding effects in the event-related brain potential. *Psychophysiology* 33, 491–506. doi: 10.1111/j.1469-8986.1996.tb02425.x
- Vega-Mendoza, M., Pickering, M. J., and Nieuwland, M. S. (2018). Concurrent use of animacy and event-knowledge during comprehension: evidence from event-related potentials. *psyarxiv* [Preprint]. Available at <https://psyarxiv.com/2qbmp/> (accessed September 20, 2017).
- Voss, J. L., and Paller, K. A. (2009). Remembering and knowing: electrophysiological distinctions at encoding but not retrieval. *Neuroimage* 46, 280–289. doi: 10.1016/j.neuroimage.2009.01.048
- Walker, C. H., and Yekovich, F. R. (1987). Activation and use of script-based antecedents in anaphoric reference. *J. Mem. Lang.* 26, 673–691. doi: 10.1016/0749-596x(87)90109-4
- Wang, L., and Schumacher, P. B. (2013). New is not always costly: evidence from online processing of topic and contrast in Japanese. *Front. Psychol.* 4:363. doi: 10.3389/fpsyg.2013.00363
- Warnes, G. R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., et al. (2017). *gdata: Various R Programming Tools for Data Manipulation. R package version 2.18.0*.
- Weiss, S., and Mueller, H. M. (2012). "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. *Front. Psychol.* 3:201. doi: 10.3389/fpsyg.2012.00201
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1*.

- Wickham, H., François, R., Henry, L., and Müller, K. (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.0.1.
- Wickham, H., and Henry, L. (2019). *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. R package version 0.8.3.
- Wilke, C. O. (2019). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 0.9.4.
- Yang, C. L., Perfetti, C. A., and Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 55–89. doi: 10.1037/0278-7393.33.1.55

**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 © 2019 Nieuwland, Coopmans and Sommers. 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.





# Individual Differences in Verb Bias Sensitivity in Children and Adults With Developmental Language Disorder

Jessica E. Hall<sup>1\*</sup>, Amanda Owen Van Horne<sup>2</sup> and Thomas A. Farmer<sup>3</sup>

<sup>1</sup> Speech, Language, and Hearing Sciences, The University of Arizona, Tucson, AZ, United States, <sup>2</sup> Communication Sciences and Disorders, University of Delaware, Newark, DE, United States, <sup>3</sup> Department of Psychology, California State University, Fullerton, Fullerton, CA, United States

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Rachel Ryskin,  
Massachusetts Institute  
of Technology, United States  
Gabriela Simon-Cereijido,  
California State University, Los  
Angeles, United States

### \*Correspondence:

Jessica E. Hall  
jessicahall@email.arizona.edu

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 26 June 2019

**Accepted:** 28 October 2019

**Published:** 19 November 2019

### Citation:

Hall JE, Owen Van Horne A and  
Farmer TA (2019) Individual  
Differences in Verb Bias Sensitivity  
in Children and Adults With  
Developmental Language Disorder.  
Front. Hum. Neurosci. 13:402.  
doi: 10.3389/fnhum.2019.00402

A number of experiments support the hypothetical utility of statistical information for language learning and processing among both children and adults. However, tasks in these studies are often very general, and only a few include populations with developmental language disorder (DLD). We wanted to determine whether a stronger relationship might be shown when the measure of statistical learning is chosen for its relevance to the language task when including a substantial number of participants with DLD. The language ability we measured was sensitivity to verb bias – the likelihood of a verb to appear with a certain argument or interpretation. A previous study showed adults with DLD were less sensitive to verb bias than their typical peers. Verb bias sensitivity had not yet been tested in children with DLD. In Study 1, 49 children, ages 7–9 years, 17 of whom were classified as having DLD, completed a task designed to measure sensitivity to verb bias through implicit and explicit measures. We found children with and without DLD showed sensitivity to verb bias in implicit but not explicit measures, with no differences between groups. In Study 2, we used a multiverse approach to investigate whether individual differences in statistical learning predicted verb bias sensitivity in these participants as well as in a dataset of adult participants. Our analysis revealed no evidence of a relationship between statistical learning and verb bias sensitivity in children, which was not unexpected given we found no group differences in Study 1. Statistical learning predicted sensitivity to verb bias as measured through explicit measures in adults, though results were not robust. These findings suggest that verb bias may still be relatively unstable in school age children, and thus may not play the same role in sentence processing in children as in adults. It would also seem that individuals with DLD may not be using the same mechanisms during processing as their typically developing (TD) peers in adulthood. Thus, statistical information may differ in relevance for language processing in individuals with and without DLD.

**Keywords:** developmental language disorder, sentence processing, statistical learning, language development, mouse tracking, verb bias, artificial grammar learning, specific language impairment

## INTRODUCTION

Statistical learning is often studied in the context of language learning because researchers have considered statistical learning tasks as representative of the types of tasks that people face when learning language. Language is full of statistical regularities of different types, and statistical learning tasks can isolate some of these features to determine how they are learned at different points in development and in different populations. Indeed, individual differences studies have documented that variability in statistical learning ability can predict variability in performance on measures of language comprehension. Misyak and Christiansen, 2012 and Misyak et al. (2010) studies with adults, and Kidd (2012) and Kidd and Arciuli (2016) studies with children, for example, have shown positive correlations for performance on statistical learning tasks and comprehension tasks. Additionally, Lany et al. (2018) found evidence of a relationship between segmenting ability in a statistical learning task and how efficiently infants processed speech.

The relationship between statistical learning and language ability has been important for understanding developmental language disorder (DLD). DLD, formerly known as Specific Language Impairment, is a disorder that affects an individual's ability to effectively learn and use language. This deficit in language learning and use is not attributable to any other biomedical cause. It affects approximately 13% of children (when including children with low non-verbal intelligence scores as the new terminology of DLD mandates; Tomblin et al., 1997; Bishop et al., 2017) and has lifelong academic and social consequences (Conti-Ramsden et al., 2018). Several researchers have posited that learning the statistics of language may be a barrier for people with DLD and may have some causal explanation for the profiles we see in DLD. One of the first studies of statistical learning in DLD showed a relationship between statistical learning and receptive vocabulary (Evans et al., 2009). Arciuli and Conway (2018), however, recently argued that "... research on statistical learning and language acquisition in developmental disabilities should be broadened beyond group comparisons to consideration of individual differences and contextual factors that contribute to variability in language difficulties within and across disabilities" (p. 7). We agree that it is important to consider and test the relationship between statistical learning and language ability, especially given that recent research is casting some doubt onto what statistical learning tasks can reveal about language. For example, Spit and Rispen (2019) found that although gifted children showed better comprehension of object relative clauses than their same-age peers, performance on a serial reaction time task did not account for this variability. This finding is evidence that statistical learning *in general* may not have a strong relationship with language ability. The modality of the statistical learning task, especially if it differs from the language task, could potentially mask any relationship. Siegelman et al. (2018a) make a convincing argument that the vast experience people have with sounds in their native language impacts their performance on auditory learning tasks, unlike in visual statistical learning tasks in which people do not have similar amounts of entrenched experience. It is also possible that statistical learning involves

a number of underlying components that enable encoding and abstraction, and that statistical learning tasks vary in how they test these components (Arciuli, 2017).

Three challenges contribute to difficulty in establishing reliable relationships between statistical learning ability and variability in language learning and processing. One is the problem of how to appropriately measure language proficiency, given that it entails numerous skills. Standardized tests of grammatical proficiency may cover too broad a range of constructs to demonstrate a strong relationship, with a limited number of items on any one skill or knowledge type. Given that the multi-factorial nature of global proficiency tasks may reduce the likelihood of detecting relationships to statistical learning ability, a more appropriate strategy for empirically documenting such links involves employing language tasks designed to measure more narrow (sub)components of grammatical competency. For example, Misyak et al. (2010) documented such a relationship by utilizing a non-adjacent dependency learning task to predict the ease of processing relative clauses, a language task that involves tracking the non-adjacent dependency between the embedded and main verbs of the sentence.

Secondly, there is the appropriateness of the statistical learning task. For example, the serial reaction time task has been used many times, but we question its relevance to language learning for reasons of modality and the statistics involved. Kidd et al. (2018) make this point clearly in their paper: "Typically, studies quantify SL [statistical learning] as the ability to learn simple transitional probabilities, but SL-for-language likely requires more than this..." (Box 1, p. 163). Erickson and Thiessen (2015) also note that "language acquisition involves sensitivity to more kinds of statistical information than simple transitional probabilities" (p. 68) in their discussion of underlying processes of extraction and integration. Proficient language use does not end with learning that *the* always predicts *boy*, but requires learning that *the* predicts a set of words that share a syntactic distribution and semantic features. Statistical learning tasks should be chosen based on the relevance of the potential component mechanisms for the language skill being studied (as per the non-adjacent dependency example discussed in the preceding paragraph).

Relatedly, we take the view that language learning involves learning multiple types of probabilistic relationships among units existing at multiple representational levels. Thus accordingly, because language develops over time, we might expect statistical learning to correlate differently to language ability at different developmental time points (i.e., relationships may appear stronger after a skill is mastered than before). To this point, Arnon (2019) found reliability across three statistical learning tasks in adults but not in children. Thus, the final challenge is to choose the appropriate task for the age group being tested, taking into consideration theories of language acquisition and cognitive development in tandem.

We addressed these three challenges through the following study design. First, we designed tasks that seemed to rely upon one common skill: the ability to learn a word's syntactic distribution. For our language task, we used an adaptation of Snedeker and Trueswell's (2004) task that captures verb bias sensitivity. Verb bias is the product of a certain type of

grammatical category learning, the learning that some verbs are more likely to occur with specific words, phrases, or interpretations. For the statistical learning task, we chose an auditory statistical learning task that employed linguistic stimuli. We used an adaptation of the artificial grammar learning experiment from Reeder et al. (2013) that focuses on grammatical category learning. We used this design in experiments with children (Hall et al., 2018a) and child and adult populations with DLD (Hall et al., 2017, 2018b). Because distributional learning drives successful performance on the artificial grammar learning task and because of the distributional nature of the verb bias information for differentiated performance, we think these tasks tap into similar underlying components of grammatical processing and representational knowledge. Accordingly, we predict a significant relationship between scores on these two measures. We employed an implicit measure of grammatical comprehension ability, mouse tracking, to capture variability in sentence processing at a more fine-grained level than provided by off-line indices of comprehension. Preliminarily, we have examined use of verb bias by college students with DLD (Hall et al., 2019) and found that they were less sensitive than typically developing (TD) peers, suggesting indeed that the presence of a language disorder may be related in some way to deficits in the processing and representation of verb bias.

Finally, we chose an age group in which we expected these skills to be nearing adult proficiency but possibly impaired for children with DLD: ages 7–9. Snedeker and Trueswell (2004) showed that children as young as five were sensitive to verb bias, but verb bias had not yet been studied in children with DLD. Peter et al. (2015) showed evidence of stronger verb bias effects with age in a study of adults and children ages 3–6. At 7–9, children are just beginning to read, and reading may be a factor that further entrenches verb bias (see Perfetti et al., 2001; Shankweiler et al., 2008; Mani and Huettig, 2014; for evidence of a relationship between literacy and sentence processing). We hoped to minimize the impact of reading and so we did not choose older children. We did not choose younger ages because, as Peter et al. (2015) state, early on, verb bias is much weaker in young children because their cumulative experience is smaller, and thus more susceptible to “random fluctuations in the input.” Fluctuations could take the form of uncommon constructions in children’s storybooks and songs or in the speech of peers. These studies, in combination with findings from studies that show the manipulability of verb bias (Wells et al., 2009; Farmer et al., 2011; Fine and Jaeger, 2013; Fine et al., 2013; Ryskin et al., 2017), suggest that verb biases are still forming in children as old as 9 and continue to be shaped by experience throughout the lifespan. Further evidence of prolonged syntactic development throughout adolescence comes from brain imaging studies of sentence processing (Schneider et al., 2016, 2018; Schneider and Maguire, 2018).

Although we are interested in group differences, we took seriously the call of Kidd et al. (2018) to test this relationship with a more rigorous methodological approach. We elected to use a multiverse approach to more transparently characterize our findings. The multiverse is a relatively new method for analyzing

and reporting data to increase transparency and rigor (Steege et al., 2016). In the multiverse approach, the many choices that the researcher must make about data processing and selection are made plain. For example, options such as removing outliers or not, classifying SES by income alone or by income plus years of education, and binning participants into subgroups according to a numerical measure, among many others, are all presented, and the data are then analyzed for each option. By presenting all possible results (the multiverse), the researcher communicates a more accurate assessment of the robustness of the findings. Note that the multiverse approach is not a method for selecting or evaluating models; instead, it is a principled way to show the strength of findings given the number of arbitrary choices researchers must make when analyzing complex datasets. The multiverse approach is a method of avoiding both Type 1 and Type 2 errors because it allows a wider lens through which to view the data. We elected to use the multiverse given the large and confusing number of measures that could be used to determine learning within each task.

One challenge in using a multiverse approach is that the complexity involved nearly necessitates using data that are already available in the literature. With the exception of children’s performance on the verb bias task, all data in this paper have been previously reported. This includes sensitivity to verb bias by adults with and without DLD (Hall et al., 2019) and artificial grammar learning by both children with and without DLD (Hall et al., 2018a,b) and adults with and without DLD (Hall et al., 2017). Study 1 provides the findings for how children with and without DLD perform on the verb bias task, completing the data reporting required for the multiverse analysis. In Study 1, we draw on the previously reported adult data to make clear the range of performance and possible predictors of performance and to support interpretation of the child data. In Study 2 we then report our findings using the multiverse approach to determine evidence of a relationship between statistical learning and language.

## STUDY 1 INTRODUCTION

Snedeker and Trueswell (2004) used a visual world paradigm to determine differences in verb bias and visual referent sensitivity between TD adults and children. In their study, they used syntactically ambiguous sentences (e.g., *Feel the frog with the feather*) that required children to act out one of the two possible interpretations (either using the object as an *instrument* to complete the action or choosing an animal that was holding the object in which it is as seen as a *modifier*). Stimuli varied by the likelihood that the verb in the sentence was to appear with one of the two interpretations in corpus and sentence-completion norming data (*poke* occurs more often in instrument interpretations whereas *hug* occurs more often in modifier interpretations), or by the number of visual referents present (one frog vs. two frogs). TD children showed no differences in verb bias sensitivity compared with adults in both their choice of interpretation and in eye tracking measures of where they looked while completing the task. However, children did show different

patterns of choice and looking behavior than adults when two referents were present.

We adapted this task for use with mouse tracking in our study of verb bias sensitivity in college students with DLD (Hall et al., 2019). In our study, participants viewed a computer screen with illustrations of each interpretation of the ambiguous sentence (e.g., *The elephant pokes the camel with the feather*) in the two top corners. Participants were instructed to click on the interpretation that went with the sentence. We used the trajectory of their mouse movement to measure the amount that they were attracted toward the competing picture for a given trial. Previous mouse tracking studies have demonstrated that when very little competition is present, participants move the mouse in a straight trajectory toward their choice; whereas trials with a great amount of competition result in trajectories that curve toward the competitor (Spivey et al., 2005; Farmer et al., 2007; Spivey, 2007; see Freeman et al., 2011, for an overview). Thus, mouse trajectories provide a means to continuously measure dynamic competition during sentence processing through time-normalized  $x,y$  coordinates. In this way, we could measure the role of verb bias in the choice of interpretation and in the process of making that choice. We found that TD college students chose interpretations that were consistent with verb bias more often than their peers with DLD, and their mouse trajectories also reflected greater sensitivity to verb bias, with more curved mouse trajectories when they chose an interpretation that was inconsistent with verb bias than when it was consistent. The trajectories for the group with DLD did not show this pattern.

Two previous studies have shown poorer verb comprehension in children with language deficits relative to TD peers. Kelly and Rice (1994) showed that children with DLD had no preference for a change-of-state vs. motion interpretations for a novel verb form, whereas age-matched TD peers demonstrated a preference for change-of-state interpretations, suggesting that the children with DLD may not be as sensitive to subtle aspects of verb subcategory that lead TD children to have a change-of-state bias. Nation et al. (2003) showed that children with poor comprehension skills were as quick to anticipate the object of a verb as TD peers for verbs with specific semantic restrictions (e.g., “eat” predicts something edible). However, these participants spent less time overall looking at the target object than more skilled comprehenders. This is further evidence that although children with DLD may be sensitive to restrictive semantics associated with verbs, they may not have well-developed preferences for interpretations of verbs based on statistical likelihood.

In this study, we examine the degree to which grammar production skills, as measured by the Structured Photographic Expressive Language Test, Third Edition (SPELT-3, Dawson et al., 2003), a test that can diagnostically discriminate between children with and without DLD, and text exposure, as measured by the recognition of children’s book titles, predict verb bias sensitivity in children ages 7–9. We predict that, like the adults with DLD in our previous study, children with lower SPELT-3 scores will show less sensitivity to verb bias than more grammatically productive peers, both in the choices they make and at the level of cognitive competition, as revealed by mouse-trajectories. We might expect

more differences in childhood than in adulthood because adults could have developed compensatory strategies to aid them in sentence processing. We examine the role of text exposure because we expect that children who read or are read to more often will have more entrenched verb biases than children with less reading experience.

## STUDY 1 METHODS

### Participants

Participants were 55 children ages 7–9, 19 who were classified as having DLD and 36 who were classified as TD. All of the participants with DLD and 22 of the TD participants in this study also participated in the statistical learning study reported in Hall et al. (2018b); their demographics are reported in the first two rows of **Table 1** for the purpose of data transparency. Thirty-four of the 36 TD children participated in the study reported in Hall et al. (2018a). Six children with DLD participated in the current study after completing a treatment study on morpheme production (Owen Van Horne et al., 2017, 2018). All children had normal hearing verified by a screening (American Speech-Language-Hearing Association, 1997), had normal or corrected vision, passed a non-verbal intelligence test (Kaufman Brief Intelligence Test-2, matrices; Kaufman and Kaufman, 2004), and had no history of autism spectrum disorders or neurological disorders by parent report. Children in the DLD group scored a standard score of 95 or below on the SPELT-3 (Dawson et al., 2003). Although our children were older ages than those in Perona et al. (2005), which recommends the 95 cut off score for highest sensitivity and specificity, we think that this categorization is reasonable because all but three children in the DLD group received services for speech and language. Of those three, two received reading services. Children in the TD group scored above 95 on the SPELT-3 and had no history of speech or language difficulties. Children in both groups completed the Peabody Picture Vocabulary Test, 4th edition (PPVT-4, Dunn and Dunn, 2007), and the title recognition task (Montag and MacDonald, 2015). In the title recognition task, designed to measure the amount children ages 8–12 are exposed to text, the examiner reads aloud titles of real and fake children’s books, and the child responds yes if they have heard of that book before. Data were screened as described in section “Data Screening,” and data from three TD and three DLD participants were excluded. Participant demographics and test scores for those 49 participants included in the final analyses are reported in **Table 1**.

### Materials

We again used the mouse tracking adaptation of the visual world paradigm task from Snedeker and Trueswell (2004) that we used in Hall et al. (2019). Sentences in the experimental trials were syntactically ambiguous, e.g., “The giraffe brushes the zebra with the sponge.” Two possible interpretations were displayed, an instrument interpretation (the giraffe using a sponge to brush the zebra) and a modifier interpretation (the giraffe using its foot to brush a zebra that holds a sponge). We compared mouse trajectory curvature on trials when



**TABLE 1 |** Participant demographic and testing information by diagnostic category (DLD, developmental language disorder; and TD, typically developing), after excluding participants as described in the screening measures of section “Data Screening,” for child datasets in Studies 1 and 2.

	<i>n</i> <i>n</i> male		Age (months)		Maternal education (years)		PPVT-4 raw		PPVT-4 SS		SPELT-3 SS		KBIT-2 SS		TRT	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
DLD	16	12	92.9	8.6	15.0	2.2	133.1	13.0	103.8	9.2	84.3	10.1	106.8	14.3	−7.1	7.0
Age-matched TD	17	9	95.2	8.0	17.1	3.5	153.6	11.1	119.6	12.6	110.3	7.7	117.2	11.6	−2.2	6.9
Additional TD	16	4	105.1	14.9	18.2	4.4	162.9	19.0	120.9	11.1	111.4	5.7	109.4	13.1	4.0	12.5

PPVT, Peabody Pictures Vocabulary Test, 4th Edition. Raw, raw scores; SS, standard scores; SPELT-3, Structured Photographic Expressive Language Test, 3rd Edition; K-BIT 2, Kaufman Brief Intelligence Test, 2nd Edition, non-verbal subtest. Scores on the title recognition task (TRT) are out of a range of −30 to 30, with −30 being the lowest score.

participants chose an interpretation that matched the verb bias vs. trials in which the choice did not match bias. We defined sensitivity to verb bias as greater mouse curvature on mismatched trials than matched trials. We also had comprehension trials which displayed only one of the two correct interpretations. The alternative picture showed an impossible interpretation (the giraffe holding a sponge but not brushing the zebra). These trials provided a screener for children who did not understand the sentences. We expected more incorrect trials in the DLD group due to comprehension difficulties, though these should still be somewhat rare in both groups because of the simple nature of sentences. Pictures across all trial types were made to appear as similar as possible, with the object and animals roughly the same size so as to neutralize the salience of items in the visual display. Verbs in the sentences were either biased to appear with instrumental phrases (The butterfly hits the grasshopper with the flower) or biased to appear with modifier phrases (The gorilla hugs the cat with the blanket), based on the norming data and classification reported in Snedeker and Trueswell (2004). Filler trials had two pictures of different animals and sentences that asked participants to “Click on the (animal name) that (animal attribute).” These were used as a measure for overall mouse movement because DLD is associated with poorer motor control (Hill, 2001) and were entered in our models as a covariate. Practice trials in the beginning required matching color and shape (“The red circle is bigger than the blue”) and familiarized participants with the task. See the Appendix in Hall et al. (2019) for a description of experimental sentence and picture stimuli.

Participants completed eight practice trials first, and then 16 experimental trials, 16 comprehension trials, and 24 fillers for a total of 56 items, presented in a completely randomized order. Pictures were counterbalanced for position of the modifier and instrument interpretation, position of the impossible interpretation on comprehension trials, and position of the correct interpretation on filler trials. Within each of the experimental and comprehension trials there were eight sentences with instrument-biased verbs and eight with modifier-biased verbs. Direct objects in the sentences were from Snedeker and Trueswell (2004) and were chosen to have little impact on the interpretation of the sentence as instrument or modifier with the verb they appeared with.

We used MouseTracker software (Freeman and Ambady, 2010) to deliver the task and to measure mouse curvature on each trial. The mouse was reset to the same location at the beginning of the trial, and *x,y* coordinates were used to determine curvature. MouseTracker software calculates the maximum deviation of the mouse from an imaginary straight line drawn between actual start and end points of each mouse movement. Warning messages appeared at the end of a trial if the participant took longer than three seconds to initiate movement. The examiner read the message to the child and explained that she could look at the pictures as long as she liked *before* pressing the button, but she needed to choose as quickly as possible after she heard the sentence.

## Measures

Using the lme4 package (Bates et al., 2015) and the lmerTest package (Kuznetsova et al., 2017) in R version 3.5.1 (R Core Team, 2018), we ran linear mixed effects models to explore subject and linguistic factors influencing choice of interpretation and mouse trajectories for experimental trials.

## Choice of Interpretation

Because the interpretations shown during experimental trials were both possible, we examined overt choice to determine children’s sensitivity to the bias of the individual verb. We were also curious if children would be sensitive to the global bias in which instrument interpretations are overall more likely in English. We used a mixed effects logistic regression with likelihood of instrument choice as the dependent variable and the bias of the verb as a fixed factor. SPELT-3 served as a measure of participants’ grammar production skills and the title recognition task as a measure of text exposure. Categorical variables were effects coded, and the reference variable was instrument verb. Continuous variables were centered. The maximal random effects structure included a subject slope for verb bias and intercepts for subject and item. Akaike information criterion (AIC) was used to determine model fit, using the maximal random effects structure when the difference between models’ AIC was less than 2.

Consistent with the results of five-year-old typical children in Snedeker and Trueswell (2004), we predicted that children would show sensitivity to the bias of individual verbs and a slight tendency to choose instrument interpretations more often, and reflecting knowledge of the global instrument bias in English.

We predicted that this sensitivity would vary based on language proficiency and/or text exposure.

Mouse Movements

We used a linear mixed effects model in which the dependent variable was mouse curvature as measured by maximum deviation, and fixed effects included the bias of the verb, the consistency of choice of interpretation with verb bias, the expected strength of verb bias according to Snedeker and Trueswell (2004) norms (see **Table 2**), SPELT-3 scores, title recognition task scores, and interactions between consistency, bias, and proficiency measures. Strength of verb bias was a continuous variable used to account for variability in items attributed to the linguistic cue rather than the visual cues. We included participants' average maximum deviation for filler trials to control for individual differences in motor control. For all analyses, variables were dummy coded, and instrument bias and consistent choice served as the reference categories. The maximal random effects structure included subject slopes for verb bias, consistency of choice of interpretation, their interaction, as well as strength of bias, and random subject and item intercepts. AIC was used to determine model fit using the same criterion as above. We included random effects and measures of motor control and strength of bias because the latter two are attributable and meaningful in the consideration of differences between individuals with and without DLD, and therefore not random.

We predicted that measures of proficiency would predict sensitivity to verb bias, with straighter trajectories (small maximum deviation) across all trial types representing poor use of verb bias information. Greater verb bias sensitivity would be demonstrated by straighter trajectories on trials in which interpretation choice was consistent with verb bias, and more

curved trajectories (large maximum deviation) on trials when participants chose an interpretation inconsistent with bias.

Procedures

This is the same experiment as was reported in Hall et al. (2019) and the same procedures were followed for children as the adults in that study. We briefly discuss them here. Procedures were approved by the Institutional Review Board at the University of Iowa. Children participated in this study during a one-hour session that sometimes also included other tasks, including standardized testing and the artificial grammar learning task reported in Hall et al. (2018a,b).

Children sat at a laptop computer and listened to the examiner give instructions. The examiner told children to carefully view the two pictures located at the top corners of the screen on each trial to find the differences between the pictures. When they saw the differences, they could then click on the Start button to begin the trial and listen to the sentence. Children were instructed to move the mouse as quickly as possible to the picture that went with the sentence after the sentence played. If children had difficulty with this, they were told to wait until the star appeared on the screen to move the mouse. Children were reminded to choose as quickly as possible several times during the experiment, and they were given stickers and encouragement and occasionally short breaks if their attention waned. All children completed the experiment with their right hand. Many children reported never having used a computer mouse before, so although we did not collect handedness information, it likely did not matter because children were not especially dexterous and because we included a measure of movement on control trials as a covariate.

STUDY 1 RESULTS

Data Screening

We did not include practice trials in any analysis. We measured accuracy on comprehension trials to screen trials and participants. We excluded three participants with DLD and three TD participants for choosing incorrect interpretations for more than half of the 16 comprehension trials, leaving us with 37 participants. With these children excluded, a Mann-Whitney *U* test confirmed that participants with DLD had more incorrect responses than TD children, *U* = 113, *p* < 0.05. The average number of incorrect responses for a participant with DLD was 6.5 (*SD* = 1.3) and for TD children was 4.3 (*SD* = 2.3). **Table 1** provides demographic information for non-excluded participants only.

We also screened each experimental trial mouse trajectory for the remaining participants for aberrant mouse movements (i.e., non-interpretable looping cycling leftward and rightward; Freeman et al., 2008; excluding 37 trials by children with DLD and 33 trials by TD children; 70 trials total). We also excluded trials with a reaction time exceeding 5000 ms (10 trials by children with DLD; 14 trials total); and trials in which initiation time exceeded 2000 ms (1 trial by children with DLD; 2 trials total). Overall, 18.8% of experimental trials for children with DLD and 7.4% of the TD children's experimental trials, or 11.2% of

**TABLE 2 |** List of verbs by bias type and the nouns that appear with the verbs in the sentences.

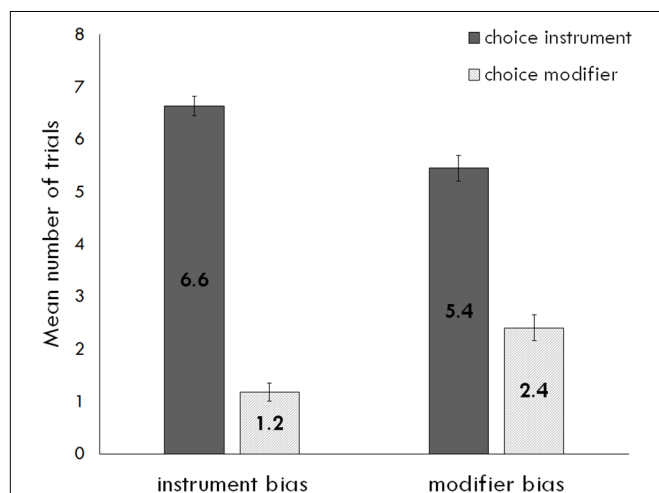
Bias	Verb	Instrument/modifier	Strength
Instrument	Hit	Flower	24
	Tickle	Fan	24
	Poke	Feather	29
	Clean*	T-Shirt	38
	Bop*	Ball	38
	Brush*	Sponge	43
	Cover*	Book	43
	Feed*	Glass	43
Modifier	Look at	Glass	19
	Hug	Blanket	19
	Find	Stick	24
	Talk to	Tube	24
	Sing to	Funnel	29
	Yell at*	Funnel	30
	Listen to*	Tube	38
	Choose*	Fork	81

*Strength is determined by the percentage of time that the verb appeared with the biased interpretation in the norming sentence completion study reported in Appendix B of Snedeker and Trueswell (2004). Asterisks denote "strongly biased" verbs.*

**TABLE 3 |** Summary of logistic regression analysis for variables predicting the probability that child participants choose instrument interpretation in the verb bias task.

Factor	Variance	SD	$\beta$	SE	<i>p</i>
<b>Random factors</b>					
Subject	0.65	0.81			
Item	0.89	0.94			
<b>Fixed factors</b>					
(Intercept)			1.50	0.32	<0.0001
Bias (reference category = instrument)			0.45	0.27	0.10
SPELT-3 standard score, centered			−0.01	0.01	0.60
Title recognition task, centered			−0.01	0.02	0.43
Bias × SPELT-3			0.01	0.01	0.33
Bias × Title recognition task			−0.01	0.01	0.54
Model $\chi^2 = 5.92$					

SPELT-3, *Structured Photographic Expressive Language Test, 3rd Edition*.

**FIGURE 1 |** Mean number of trials (out of 8) in which participants chose each interpretation (instrument and modifier) for instrument- and modifier-biased trials. Error bars represent standard error.

total data, were discarded. We did not exclude any participants for missing trials because mixed effects models can adequately handle missing data. The children with the largest number of missing trials in each diagnostic group were a child with DLD with 7 missing trials and a TD child with 6 missing trials. The mean total number of missing trials for children with DLD was 3.0 ( $SD = 1.8$ ) and 1.2 ( $SD = 1.3$ ) for TD children, a difference that was significant according to a two-tailed independent sample *t*-test,  $t(47) = 4.00$ ,  $p < 0.01$ .

Finally, we ran a linear mixed effects model with a random subject intercept with only filler trials included to test for baseline motor control differences between groups. There was no difference between children with and without DLD for maximum deviation of mouse trajectories on control trials,  $p = 0.48$ .

## Choice of Interpretation

We first examined choice of interpretation. The best fit model included random intercepts for subject and item. The dependent

**TABLE 4 |** Results of the mixed effects linear model for maximum deviation of mouse trajectories as influenced by consistency of choice with verb bias, choice of interpretation, measures of text exposure (title recognition task) and language proficiency (SPELT-3 standard score), and their interactions, as well as expected strength of verb bias and average maximum deviation on control trials.

Factor	Variance	SD	$\beta$	SE	<i>p</i>
<b>Random factors</b>					
Subject intercept	0.004	0.07			
Verb bias	0.07	0.26			
Consistency of choice	0.04	0.20			
<b>Fixed factors</b>					
(Intercept)			0.14	0.06	<b>0.01</b>
Verb bias (reference category = instrument)			0.11	0.09	0.20
Consistency of choice of interpretation (reference category = consistent)			0.14	0.10	0.15
SPELT-3 standard score, centered			−0.002	0.002	0.25
Title recognition task score, centered			0.004	0.003	0.17
Strength of verb bias			−0.002	0.001	0.18
Average maximum deviation on control trials			0.50	0.18	<b>0.01</b>
Verb bias × Consistency of choice			−0.29	0.15	<b>0.01</b>
Verb bias × SPELT			−0.003	0.005	0.51
Verb bias × Title recognition task			0.003	0.007	0.66
Consistency of choice × SPELT			0.0006	0.006	0.91
Consistency of choice × Title recognition task			0.008	0.008	0.33
Verb bias × Consistency of choice × SPELT			0.005	0.006	0.47
Verb bias × Consistency of choice × Title recognition task			−0.02	0.01	<b>0.09</b>

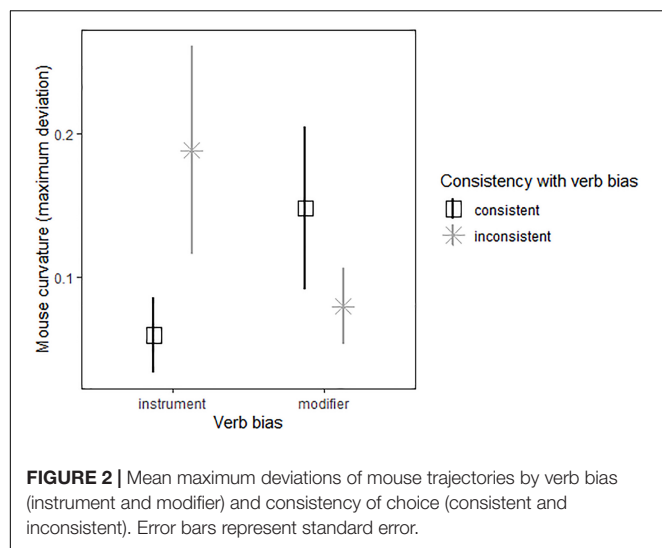
Bold indicated significance ( $p = 0.05$ ) or near significance ( $p < 0.1$ ).

variable was the probability that a participant would select an instrument interpretation on a given trial. See **Table 3** for log odds (reported as  $\beta$ ) and **Figure 1** for an illustration of means by trial type. Recall that the instrument interpretation is the more likely overall interpretation of “with the X” phrases in English. Children were sensitive to this global bias: participants were 82% likely to choose instrument on a given trial,  $z = 4.71$ ,  $p < 0.0001$ . The bias of the verb did not significantly influence choice of interpretation,  $z = 1.65$ ,  $p = 0.10$ , and measures of text exposure and language proficiency did not significantly interact with bias or have any effect on participants’ likelihood of choosing instrument,  $ps > 0.3$ .

## Mouse Movements

Although we did not find that children’s grammatical proficiency or exposure to text influenced their choice of interpretation, it is possible they will predict sensitivity as represented in mouse trajectories. Sensitivity to bias in mouse trajectories would be indicated by a significant main effect or interaction with consistency of interpretation. This would be interpreted as more curved trajectories, and thus more competition, when participants chose responses that were inconsistent with bias.

Model results are reported in **Table 4** and mean maximum deviation of mouse trajectory by trial type illustrated in **Figure 2**. The best fit model included random subject slopes for bias and consistency. There were two significant factors: an interaction

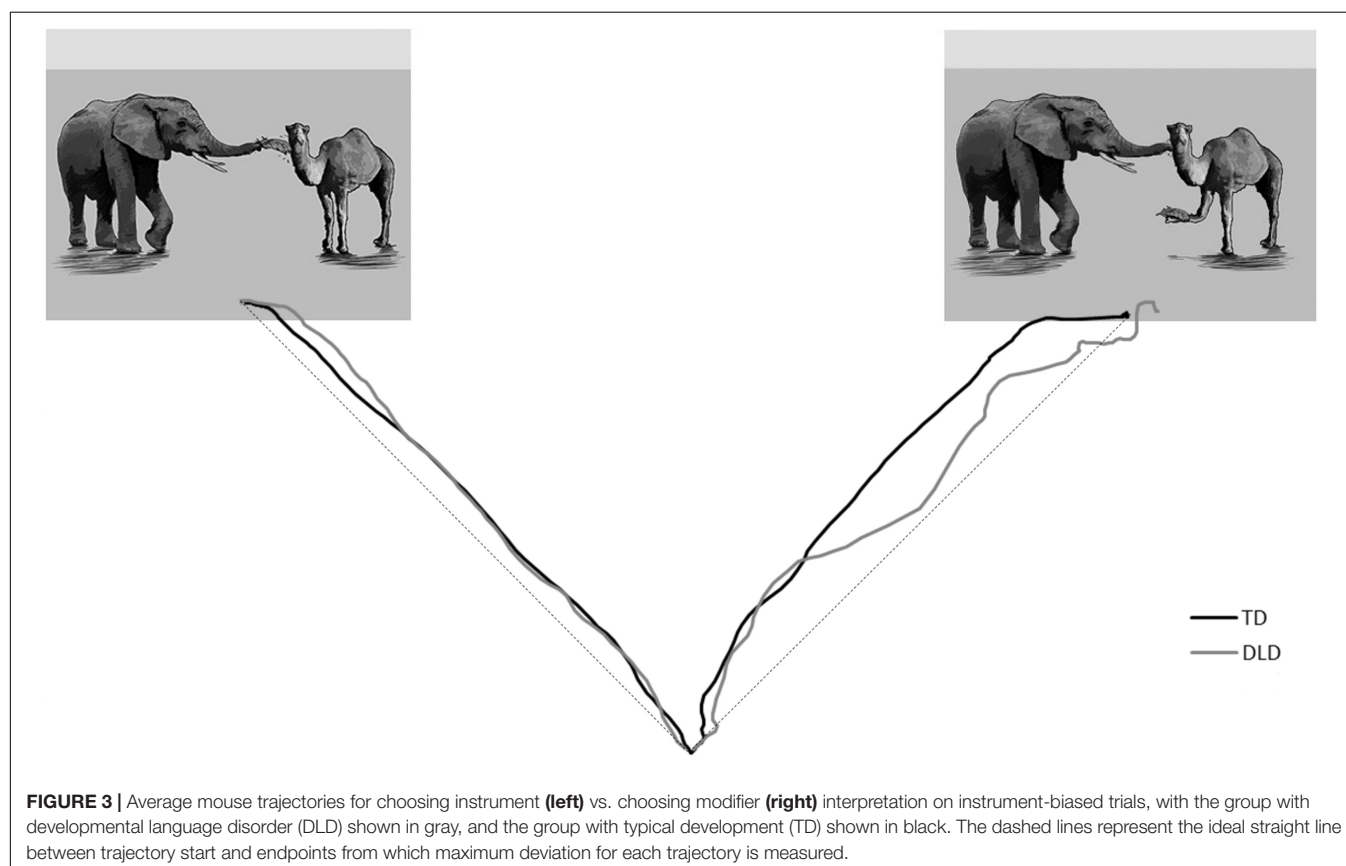


effects, we found that participants showed a larger difference between choosing inconsistently vs. consistently on instrument-biased trials than on modifier-biased trials. Participants showed curved trajectories when choosing against bias on instrument-biased trials, but somewhat straight trajectories when choosing with bias. **Figure 3** provides an illustration of averaged mouse trajectories by diagnostic group on trials with instrument-biased verbs, with choosing consistently with bias shown on the left and choosing against bias shown on the right. Participants showed the opposite pattern for modifier-biased trials, though the gap was smaller. Although **Figure 3** appears to show differences in mouse trajectory for the two diagnostic groups, participants' mouse movements were not significantly related to participants' language proficiency or text exposure, but movements on control trials positively predicted their movements on experimental trials.

## STUDY 1 DISCUSSION

between bias and consistency,  $t(620.4) = 2.52$ ,  $p = 0.01$ , and maximum deviation on control trials,  $t(87.8) = 2.75$ ,  $p < 0.01$ . There was a marginal three-way interaction between bias, consistency, and the title recognition task which measured text exposure,  $t(663.1) = 1.70$ ,  $p = 0.09$ , an effect which moved to  $p = 0.47$  when the participant with the highest title recognition task score was removed from the dataset. Interpreting these

We found that children with and without DLD ages 7–9 were primarily influenced by global bias in their choice of pictures. However, children showed sensitivity to local verb bias information in their mouse movements. There was a greater deviation toward the unchosen picture when choosing against bias on instrument-biased trials than on modifier-biased trials.





It was not surprising that children chose instrument interpretations most often, given the global instrument bias. The global bias also likely contributed the different patterns of mouse trajectories between instrument- and modifier-biased verbs. In our previous study with college students using the same stimuli (Hall et al., 2019), participants also showed a preference for the instrument interpretation and similar mouse trajectory patterns, with stronger evidence of verb bias sensitivity for instrument-biased verbs. Audio and visual stimuli in our task may have influenced children's choice differently from Snedeker and Trueswell's (2004) study. That study showed that children were likely still learning to integrate visual and linguistic cues, with behavioral differences between adults and children when an additional visual cue was added. In fact, the eye tracking data from that study revealed that children's eye movements were beginning to pattern more like adults, even though children's choice of interpretation did not yet reflect this. The findings from Peter et al. (2015) also provide evidence for a longer developmental trajectory for verb bias/cue integration, with differences in verb bias effects among 3- and 6-year-olds and adults in a syntactic priming paradigm. It was surprising that for both choice of interpretation and mouse movements, we found no correlation with measures of language or text exposure. This suggests that perhaps other aspects of cognition, such as working memory or cognitive control (see for example, Just and Carpenter, 1992; Novick et al., 2005; Lewis et al., 2006; Martin, 2016), are driving the maturation of cue integration in sentence processing.

Being able to efficiently predict upcoming information in a linguistic signal may have a profound impact on overall comprehension. If children with DLD are slower and less efficient in making their predictions than typical peers, they risk missing crucial information for making timely connections during conversations. We have no evidence of group differences among children from this data set, but Hall et al. (2019) indicates that differences do exist in adulthood using the same task. This suggests that studies of adolescent language development may be important for fully understanding the functional differences observed in adult outcomes (Conti-Ramsden and Durkin, 2007; Carroll and Dockrell, 2010; Hesketh and Conti-Ramsden, 2013). The current study suggests that syntactic prediction during processing may not yet be adult-like at these ages given that typical children as well as children with DLD showed little evidence of verb bias in their choice of interpretation.

The main contribution of this study is to take a first step examining individual differences in verb bias sensitivity in children with and without DLD. In general, results suggest that children ages 7–9 with and without DLD do not consistently use verb bias information to resolve ambiguity, though they are sensitive to verb bias. Importantly for this paper, there is also sufficient variability in performance for consideration of how individual differences in statistical learning might contribute, despite (or perhaps because of) our finding that measures of language proficiency, and text exposure did not meaningfully predict verb bias sensitivity.

## STUDY 2 INTRODUCTION

In this study we adopt a “multiverse” approach (Steege et al., 2016) to examine whether a particular statistical learning task predicts performance on the verb bias task described in Study 1. Data is drawn from the verb bias study reported in Hall et al. (2019) and the artificial grammar learning studies reported in Hall et al. (2018a,b), and Hall et al. (2017). We walk through some of the rationale for the different measurement choices first, and then present the findings.

## STUDY 2 METHODS

### Participants

#### Adults

To ensure enough participants to adequately power a test of relationship, we added 31 additional TD adult participants to the dataset of the 33 adult participants from TD and DLD groups in the verb bias study reported in Hall et al. (2019) and the artificial grammar learning study reported in Hall et al. (2017). The additional 31 adult participants were recruited from the University of Iowa Elementary Psychology Research Exposure participant pool and were screened by self-report for being monolingual and having no history of language or cognitive impairment. All of the participants in this TD group met qualifying criteria on the tasks (at least 60% accuracy on the one-back task in the artificial grammar learning task, and at least 50% accuracy on comprehension trials in the verb bias task), and as such, data from all participants were included. Demographic information for all 64 adult participants are presented in **Table 5**, again with the participants from the original studies presented in the first two rows. For more information on how adult participants with DLD were identified, please see Hall et al. (2019).

#### Children

Child participants are the same as those from Study 1, with the six children excluded in Study 1 also excluded here. Demographic information are presented in **Table 1**.

### Analysis

We used performance on the artificial grammar learning task as a continuous variable in a mixed effects linear model predicting performance on the verb bias task. There are many arbitrary ways to measure variables by which to look for a relationship between the tasks, which led us to adopt this “multiverse” approach (Steege et al., 2016). **Table 6** lists measures for each participant group, task, verb set, and dataset, with abbreviations used in the section “Results.” We included a random subject intercept in all models because the Akaike Information Criterion (AIC) indicated this was the best fit for the previous models we ran analyzing the verb bias data in both children and adults. We do not consider alternative random effects structures for the models for the sake of space, but we recognize that these also could impact findings. Code to run analyses in R version 3.5.1

**TABLE 5 |** Participant demographic and testing means and standard deviations by diagnostic category (DLD, developmental language disorder; and TD, typically developing), after excluding participants as described in the screening measures in Hall et al. (2019) and Hall et al. (2017) for adult datasets in Study 2.

			Age (years)		Education (years)		KBIT-2 standard score		Spelling (out of 15)		Token (out of 44)		PPVT-4 raw score		Author recognition task (–65 to 65)	
	<i>n</i>	<i>n</i> male	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
DLD	17	8	20.7	1.1	13.9	1.1	99.1	10.5	4.2	2.6	35.2	5.5	199.4	11.2	16.4	12.7
Matched TD	16	8	21.0	1.9	14.1	1.9	108.9	12.3	11.7	2.5	40.1	3.1	206.8	6.8	18.9	7.0
Additional TD	31	14	20.8	1.5	Not collected		Not collected		Not collected		Not collected		202.5	10.3	24.8	15.1

Scores on Kaufman Brief Intelligence Test, 2nd Edition, non-verbal subtest (KBIT-2) are standard scores with a normative mean of 100 and a standard deviation of 15. Scores on the spelling and token tasks are raw counts of items correct out of 15 and 44, respectively. Scores on the Peabody Pictures Vocabulary Test 4th Edition (PPVT-4) are raw scores. Scores on author recognition task (ART) are out of a range of –65 to 65, with 65 being the highest possible score.

(R Core Team, 2018) and sample data are available on github at <https://github.com/jessica-hall/multiverse/>.

## Participants to Include

For all models, we tested both adults and children (Adult 1 and Child 1 in **Table 6**) and we also tested a subgroup dataset of TD participants only (Adult 2 and Child 2). We did this because we found group differences in verb bias sensitivity in adults (Hall et al., 2019), and thus the participants with DLD may have relied on other information to perform the verb bias task and therefore would not show a relationship between performance on both tasks.

## Dependent Variable: Verb Bias Measure

We had two types of measures for this task, and thus we consider two types of dependent variables in our models: choice of interpretation and mouse trajectories. The measure for choice of interpretation is consistency with verb bias, dummy coded

as “1” for consistent and “0” for inconsistent, in a logistic regression analysis (VB1 in **Table 6**). Thus, the hypothesis tested in these models is whether good learning in the artificial grammar learning task predicts *responses consistent with verb bias* in the verb bias task. The rationale here is that because the statistical learning task requires explicit evaluation of items, perhaps it will show more relation to the more explicit decision of which interpretation participants choose. For the mouse trajectories, we consider consistency of choice of interpretation as an independent variable that interacts with the statistical learning measure, with the maximum deviation of the mouse trajectory value as the dependent variable (VB2). An interaction between these variables would indicate a relationship between distributional learning in an artificial setting and distributional learning in the real world. The hypothesis tested in these models is whether good learning in the artificial grammar learning task predicts *more attraction to the unselected response* when choosing an interpretation inconsistent with verb bias. The rationale for using this dependent variable is that the mouse trajectory measure can capture a wider spectrum of differences in sentence processing than simply which picture participants chose and therefore will be a more sensitive measure of individual differences.

Next, there are several possibilities to consider for which trials to include. One is the full dataset. A second is a dataset restricted to instrument-biased verbs only because in both child and adult studies, participants showed more sensitivity to the bias of instrument-biased verbs (VB3). A third alternative is to restrict the dataset further to only instrument-biased trials in which participants chose a modifier interpretation, because choosing modifier on instrument-biased trials is the instance in which we expect to see the greatest evidence of verb bias sensitivity (greater maximum deviation values; VB4). In the most restricted models, then, there is no covariate measure of consistency of choice of interpretation because we are only considering one choice.

Finally, because we found a relationship with the expected strength of verb bias according to the norming data provided by Snedeker and Trueswell (2004), we test each of these alternatives using the full set of verbs (S1 in **Table 6**), using the full set of verbs with a strength interaction term (S2), and using a restricted set of only the strongly biased verbs for the adults (S3) or only the weakly biased verbs for the children (S4). **Table 2** provides

**TABLE 6 |** Choices of measures, definitions, and abbreviations for multiverse analysis.

### 1. Participant datasets

Adult 1: Adult participants with DLD and TD

Adult 2: Adult TD participants only

Child 1: Child participants with DLD and TD

Child 2: Child TD participants only

### 2. Verb bias measures

VB1: Mean consistency of choice of interpretation on 0–1 scale;

0 = inconsistent, 1 = consistent; logistic regression model

VB2: Maximum deviation (MD), interaction with consistency as a categorical variable; all trials; linear regression

VB3: MD, interaction with consistency; instrument-biased trials only; linear regression

VB4: MD instrument-biased trials with choice of interpretation = modifier; linear regression

### 3. Verb sets

S1: Full set of verbs

S2: Full set of verbs and strength variable

S3: Strongly biased verbs only

S4: Weakly biased verbs only

### 4. Artificial grammar learning measures, difference in mean standardized rating of each item type

AGL1: Novel minus ungrammatical, entire test

AGL2: Novel minus ungrammatical, first half of test

strength of verb bias ratings for each verb. We switch from strongly biased verbs (S3) for adults to weakly biased verbs (S4) for children because examination of data from Study 1 indicated that children showed stronger verb bias effects for weakly biased verbs, in contrast to the pattern that adults showed in Hall et al. (2019). The rationale for including this measure is that it allows us to capture some of the variability that may be associated with the linguistic element of the stimuli rather than the visual elements and therefore represent a clearer picture of the role of verb bias in sentence processing.

### Independent Variable: Statistical Learning Measure

Because performance on the statistical learning task has been described with these same participant datasets, we provide only a brief description of what is to be learned in the task and review prior results briefly.

In this task, participants listened to an artificial language that contained “gaps” of information. In the language, words of the same “category” had similar, but not perfectly overlapping, distributions. This task provides a good approximation of the type of learning required for verb bias. In the case of verb bias, one deduces subcategorizations from hearing sets of verbs appear in similar but not perfectly overlapping distributions. Similarly, in this artificial grammar learning task, one must attend to the item and the way in which it distributes into syntactical contexts and how those are interpreted in order to deduce categories and succeed at learning the grammar. At test, novel grammatical items contained combinations that were not heard during exposure but that were grammatically possible according to the shared distributional features, as well as ungrammatical items that contained unheard grammatically impossible distributional features. The key test of learning in the task is the difference in participants’ ratings of novel grammatical and ungrammatical test items. Participants rated items on a visual analog scale that we translated to values from 0 to 100, with 0 being ungrammatical and 100 being grammatical. We used  $z$  values to create a standardized measure for each participant because there was individual variation in how the scale was used. The simplest measure of distributional learning in the artificial grammar learning task is to take the difference in average ratings for novel grammatical and ungrammatical test items (AGL1 in Table 6). A large positive difference in ratings between novel and ungrammatical items would indicate learning of the grammar. As reported in Hall et al. (2018b), we found that on average, child and adult participants with and without DLD rated novel items higher than ungrammatical items, with no differences between diagnostic or age groups.

However, because we found an effect of testing order (items tested earlier received higher ratings than items at the end of the test), we also considered a separate measure of the difference in ratings for novel, and ungrammatical items from the first half of the test only (AGL2 in Table 6). Adding order to our model in our previous study distinguished both diagnostic groups and age groups. As we discuss in Hall et al. (2018b), the order effect may indicate sensitivity to a changing distribution as ungrammatical items are heard during the test phase. It is possible that in averaging ratings for novel items *throughout*

the test, we would have means for each participant near zero because of positive ratings early on and negative ratings later. We would not be able to distinguish learners who showed strong order effects and learners who always rated grammatical novel items near the midpoint of the scale (zero). This is an important distinction to make because we would consider the former to be good at distributional learning and the latter not as good. Therefore, we include two measures of statistical learning, one from the entire test and the other with items from first half only. General performance was such that 15 of 16 children with DLD, 20 of 24 TD children, 17 of 17 adults with DLD, and 14 of 17 TD adults had a positive difference in mean ratings in items over the entire test, indicating learning. The amount of difference ranged from  $-0.19$  to  $1.5$  for the children and  $-0.37$  to  $1.61$  for the adults. These numbers changed minimally within each group when considering the first half of the test only.

### Summary of Multiverse Methodology

To summarize, we consider a total of 48 models for each age group, adults and children: for the choice of interpretation as dependent variable model, we have 2 participant subsets  $\times$  3 strength of verb bias measures  $\times$  2 artificial grammar learning measures. For the mouse trajectory as dependent variable model, we have 2 participant subsets  $\times$  3 strength of verb bias measures  $\times$  3 measures of mouse trajectory  $\times$  2 artificial grammar learning measures.

## STUDY 2 RESULTS

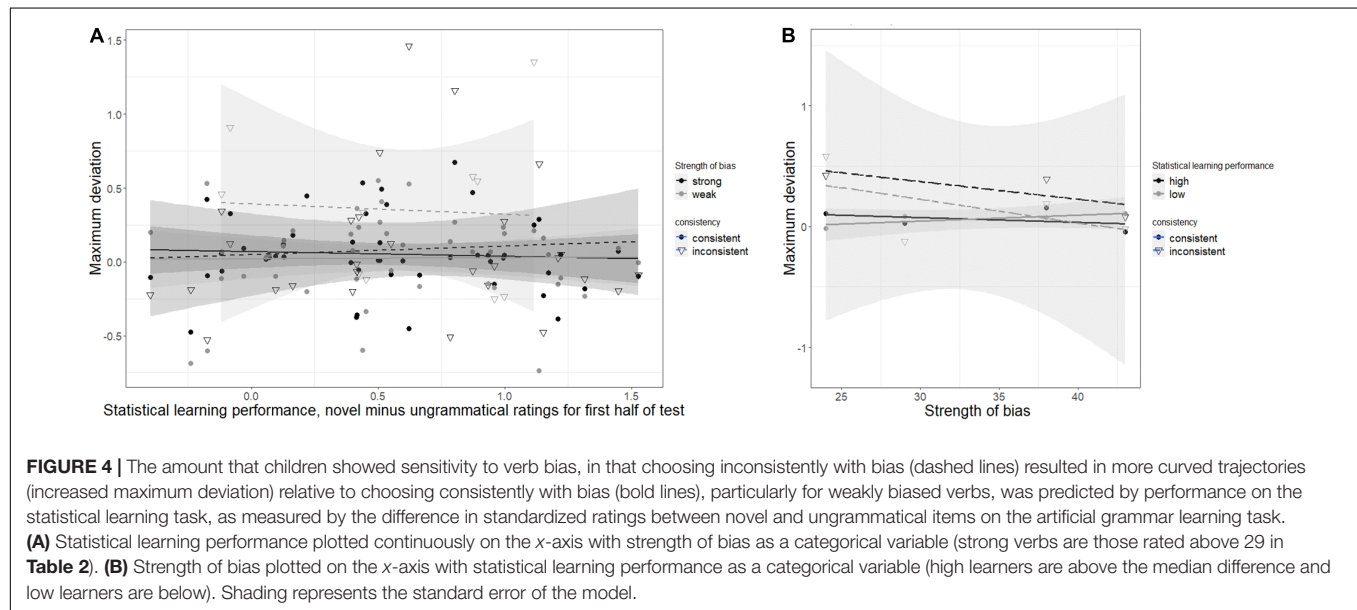
### Children

We list beta estimates, standard errors, and  $p$  values for all of the critical effects (as explained above) as well as the number of participants and number of observations for each model in Table 7. None of the 48 models with child participants returned any significant result for the critical effects. Two models were marginally significant,  $p < 0.10$ . One possibility is that there is, in fact, no relationship between this AGL task and the verb bias measures. Another is that selection of the proper comparisons is critical for observing the anticipated result. The two marginally significant models were models that had mouse trajectories (as measured by maximum deviation) as the dependent variable and had instrument-biased verbs only (VB3), and included a variable for strength of verb bias (S2). The statistical learning measure was for the first half of the test only (AGL2). The results were similar for each of the two datasets run, one with all participants (Child 1,  $p = 0.09$ ), and the other with only TD participants (Child 2,  $p = 0.07$ ) and are reported in **Supplementary Materials**. For these models, children with higher statistical learning performance showed more curved trajectories when their choice was inconsistent than when their choice was consistent. This effect was true for weakly biased verbs. Strongly biased verbs actually showed the opposite pattern, with much greater curvature overall for consistent trials compared with inconsistent trials. **Figure 4** presents

**TABLE 7 |** Beta estimates ( $\beta$ ), standard errors (SE), and  $p$  values for the critical variable of each model, and number of participants ( $n$ ) and observations ( $n$  obsv) for each model run for child datasets.

Verb bias task measures and verb sets	Participant datasets and artificial grammar learning task measures															
	Child 1								Child 2							
	AGL 1			AGL 2			<i>n</i>	<i>n obsv</i>	AGL 1			AGL 2			<i>n</i>	<i>n obsv</i>
	$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>			$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>		
VB1																
S1	−0.17	0.24	0.48	−0.04	0.16	0.82	49	678	−0.13	0.28	0.63	0.09	0.18	0.63	33	470
S2	0.01	0.02	0.53	−0.01	0.01	0.36	49	678	0.02	0.02	0.42	−0.009	0.01	0.51	33	470
S4	−0.28	0.34	0.40	0.11	0.22	0.62	49	342	−0.09	0.40	0.83	0.35	0.25	0.17	33	237
VB2																
S1	0.06	0.11	0.56	0.07	0.07	0.33	49	678	−0.03	0.12	0.83	−0.02	0.08	0.79	33	470
S2	0.002	0.007	0.78	0.006	0.01	0.25	49	678	0.005	0.01	0.56	0.007	0.005	0.16	33	470
S4	0.13	0.15	0.40	0.05	0.1	0.64	49	342	−0.11	0.16	0.49	−0.14	0.1	0.18	33	237
VB3																
S1	0.08	0.20	0.68	−0.01	0.14	0.97	49	337	0.08	0.23	0.72	−0.13	0.15	0.39	33	234
S2	0.009	0.03	0.72	0.03	0.02	<b>0.09</b>	49	337	0.02	0.03	0.57	0.04	0.02	<b>0.07</b>	33	234
S4	0.01	0.36	0.97	−0.29	0.29	0.33	48	126	−0.04	0.32	0.89	−0.44	0.27	0.10	33	89
VB4																
S1	0.19	0.23	0.42	0.02	0.16	0.92	32	50	0.25	0.28	0.37	−0.15	0.17	0.39	21	34
S2	0.003	0.03	0.92	0.02	0.02	0.21	32	50	0.009	0.03	0.79	0.03	0.02	0.21	21	34
S4	0.20	0.44	0.66	−0.25	0.36	0.50	10	14	0.19	0.55	0.74	−0.41	0.44	0.37	8	11

Definitions of abbreviations can be found in Table 6. Bold indicated significance ( $p = 0.05$ ) or near significance ( $p < 0.1$ ).



these interactions in two plots to fully illustrate these effects. **Figure 4A** shows the continuous effect of statistical learning performance and **Figure 4B** demonstrates the continuous effect of verb bias strength. As can be seen most clearly in Panel B children with high statistical learning performance (in black) showed a stronger verb bias effect for weakly biased verbs (a larger gap between the dashed and bold lines) than strongly biased verbs, and children with low statistical

learning performance (in gray) did not demonstrate a verb bias effect for strongly biased verbs. Note that in **Figure 4A**, for both very low and very high statistical learning performance, children never chose inconsistently with bias for weakly biased verbs, most easily seen by the narrower (but taller) shaded area around the dashed gray line. This restriction of range for certain item types may explain the marginal significance for these models.



**TABLE 8 |** Beta estimates ( $\beta$ ), standard errors (SE), and  $p$  values for the critical variable of each model, and number of participants ( $n$ ) and observations ( $n$  obsv) for each model run for adult datasets.

Verb bias task measures and verb sets	Participant datasets and artificial grammar learning task measures															
	Adult 1								Adult 2							
	AGL1			AGL2			<i>n</i>	<i>n obsv</i>	AGL1			AGL2			<i>n</i>	<i>n obsv</i>
	$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>			$\beta$	SE	<i>p</i>	$\beta$	SE	<i>p</i>		
VB1																
S1	0.04	0.07	0.59	−0.001	0.07	0.98	64	930	0.03	0.08	0.71	0.01	0.08	0.89	47	688
S2	0.007	0.006	0.19	0.001	0.006	0.82	64	930	0.03	0.007	<b>0.058</b>	0.008	0.006	0.18	47	688
S3	0.19	0.10	<b>0.054</b>	0.09	0.10	0.35	64	469	0.20	0.11	<b>0.07</b>	0.13	0.11	0.23	47	348
VB2																
S1	−0.002	0.03	0.94	0.008	0.03	0.76	64	930	0.03	0.03	0.35	0.03	0.03	0.28	47	688
S2	0.001	0.002	0.50	0.001	0.002	0.74	64	930	0.001	0.002	0.69	0.001	0.002	0.59	47	688
S3	0.02	0.04	0.60	0.02	0.04	0.61	64	469	0.06	0.04	0.20	0.05	0.04	0.28	47	348
VB3																
S1	0.004	0.05	0.92	−0.01	0.05	0.81	64	472	0.01	0.05	0.85	−0.04	0.05	0.47	47	349
S2	−0.003	0.006	0.59	−0.002	0.006	0.7	64	472	0.0003	0.007	0.97	0.002	0.007	0.82	47	349
S3	−0.007	0.06	0.90	−0.02	0.06	0.72	64	291	0.004	0.07	0.96	−0.05	0.07	0.44	47	216
VB4																
S1	−0.02	0.06	0.73	−0.001	0.06	0.98	41	75	−0.02	0.07	0.75	0.03	0.07	0.70	29	49
S2	−0.001	0.008	0.86	−0.001	0.007	0.92	41	75	−0.003	0.009	0.75	−0.001	0.008	0.93	29	49
S3	−0.04	0.06	0.53	−0.01	0.07	0.85	36	53	−0.04	0.07	0.61	0.02	0.08	0.79	25	36

Definitions of abbreviations can be found in **Table 6**. Bold indicated significance ( $p = 0.05$ ) or near significance ( $p < 0.1$ ).

**TABLE 9 |** Results for mixed effects logistic regression of factors predicting the probability of a choice consistent with bias in the verb bias task by adult participants, with a dataset that included only strongly biased verbs.

Factor	Variance	SD	B	SE	$p$
<b>Random factors</b>					
Subject	0.00	0.00			
<b>Fixed factors</b>					
(Intercept)			0.65	0.10	<b>&lt;0.0001</b>
Artificial grammar learning performance (novel – ungrammatical, entire test)			0.19	0.10	<b>0.054</b>

Bold indicated significance ( $p = 0.05$ ) or near significance ( $p < 0.1$ ).

## Adults

**Table 8** displays beta estimates, standard errors, and  $p$  values for all of the critical effects as well as the number of participants and number of observations for each model. Of the 48 models with adult participants, only one returned significant results for the critical effects,  $p = 0.054$ . This was the model with the difference in ratings for novel and ungrammatical items for the whole testing period (AGL1) on the artificial grammar learning task predicting the likelihood of a response consistent with verb bias (VB1), with only strongly biased verbs included (S3) and the dataset with all participants (Adult1 dataset). We report results for this model in **Table 9** and **Figure 5** provides an illustration.

A similar model for the dataset with only TD participants (Adult 2) was marginally significant,  $p = 0.07$ . Finally, one additional model with only TD participants which also included these variables and an interaction with strength rather than a subset of strongly biased verbs (S2) was

borderline at  $p = 0.058$ . Results for these models are reported in **Supplementary Materials**. All models show a trend for participants with higher statistical learning scores more likely to choose interpretations consistent with verb bias on the verb bias task than participants with lower statistical learning scores.

## STUDY 2 DISCUSSION

Our previous studies (Hall et al., 2017, 2018a,b) demonstrated that children and adults with DLD are capable of learning from distributional dependencies in an artificial grammar learning task similarly to their TD peers. However, there was considerable spread in performance by all groups. We predicted that how well individuals learned distributional dependencies in the artificial language task would have bearings on how well they use distributional information to resolve ambiguous sentences in real language. We found some evidence that individual differences in statistical learning predicted performance on the verb bias task in adult participants but not in children, but the findings are not robust. This was not especially surprising given that adults with DLD showed differences from TD peers on the verb bias task (see Hall et al., 2019), but Study 1 did not demonstrate that individual differences in language proficiency or text exposure predicted performance by child participants. The relationship was found only for predicting consistency of choice of interpretation and not for mouse trajectories in the adult participants. Indeed, children did not appear to be using verb bias when choosing an interpretation, and thus this may be why we did not see this relationship for them. That the

relationship in adults seemed to be driven by the TD participants provides further evidence that adults with DLD are not using distributional information in the same way as TD peers when disambiguating sentences in the verb bias task. It is possible that some of the adults with DLD, like the child participants, were not using verb bias to disambiguate the sentences in the sentence processing task.

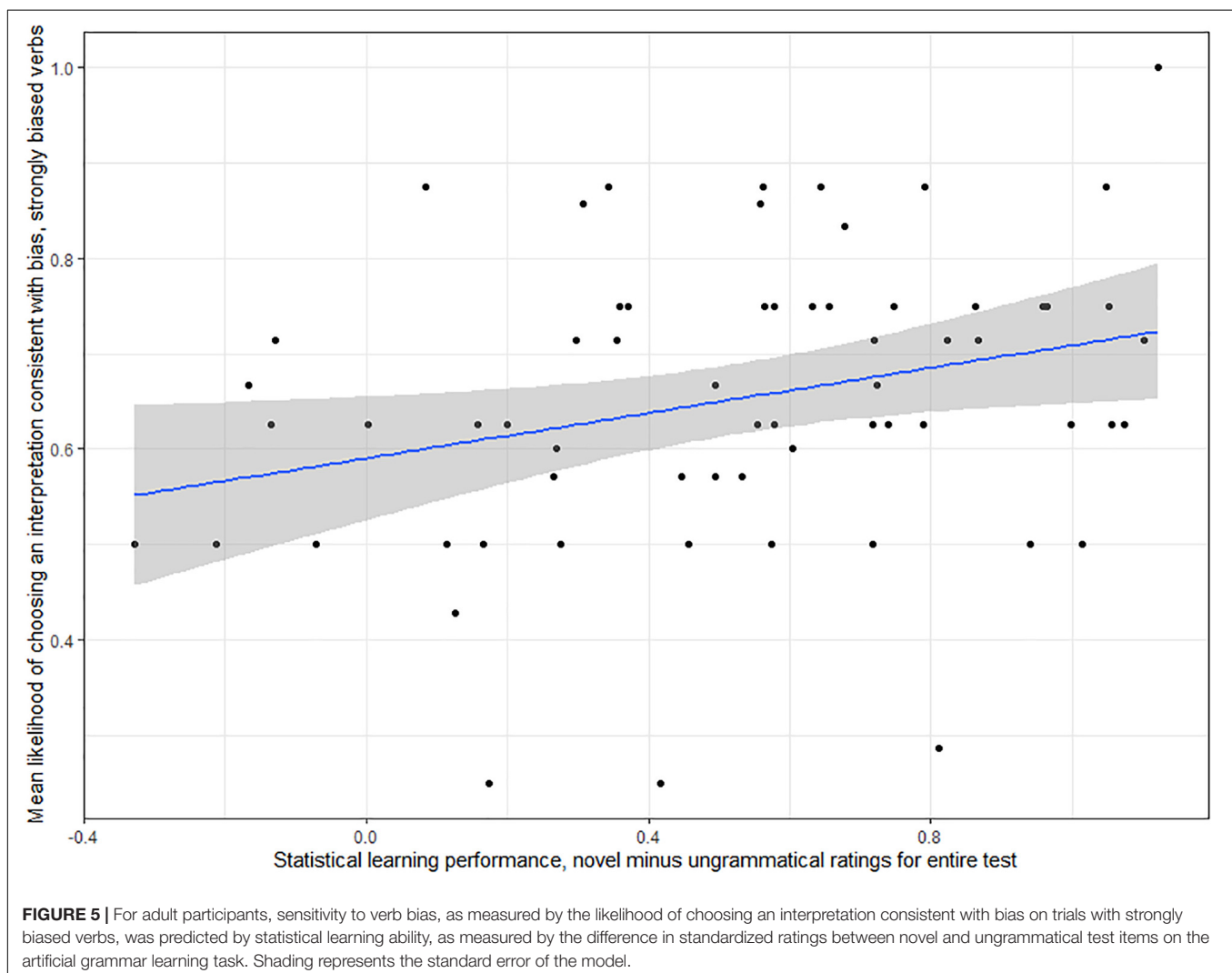
At the suggestion of a reviewer, we examined the internal consistency coefficient (ICC) for our primary independent and dependent measures in the significant model. As might be expected from previous results that showed large standard deviations for ratings in the statistical learning task with typical children (Hall et al., 2018a), we obtained very low measures of internal consistency for the artificial grammar learning task measures. Because the measure for of statistical learning depended on ratings for novel items, it was likely impacted by the great amount of noise, even in the adult data. We also obtained low measures of reliability for the verb bias task measures. The poor values on these measures of reliability conducted *post hoc* suggest that the tasks are not well suited for measuring individual

differences, at least for the number of items on the tasks in the present study. Low reliability may explain why most of the models in our multiverse analysis were not significant.

## Individual Differences

Given the number of comparisons run, it is reasonable to question whether the results obtained were simply spurious effects. Indeed, we would feel more confident if there was a more consistent result regardless of how the measures were selected. The value of the multiverse analysis is to demonstrate the robustness of the findings; because the majority of models were not significant, the significant findings from this study are not robust. Nonetheless, we believe it to be valuable to reflect on whether there is a rational explanation for why three comparisons reached or were near significance for the adults and the remainder were not.

First, the two verb bias task measures differed on the degree of explicitness. The choice of interpretation was likely a better match for the explicit grammaticality test in the artificial grammar learning task than the implicit mouse trajectory measure. A more



implicit measure of learning in the artificial grammar learning task (such as that in Lammertink et al., 2019, and López-Barroso et al., 2016) may have better predicted mouse trajectories on the verb bias task. Other studies with implicit measures have found positive relationships with children; for example, Kidd (2012) showed that 4- and 5-year-olds' performance on serial reaction time task corresponded with their ability to remember a primed sentence structure. It is beyond the scope of this paper to delineate whether implicit and explicit processes are discrete processes, but we suggest this as an area worthy of further investigation.

Second, the choice of interpretation may have been more impervious to factors like motivation, alertness, or attentiveness that may have added to the variability in mouse trajectories. The choice of interpretation may not have changed much because it may not require much effort to understand the sentences, but the speed and dexterity of mouse movements could change dramatically over time as participants become more fatigued or bored with the experiment, or, in the case of children, more adjusted to using the mouse (some had never seen one before). We might also attribute mouse trajectories as reflecting the ability to use distributional information to do speeded language processing. In the artificial grammar learning task, participants are not given a time limit to make their grammaticality decisions. In addition, the cognitive load is fairly low: participants listen to a three-word "sentence" and hold it in memory as they compare it to stored mental representations for items heard five to fifteen minutes earlier. They are not told to remember any specific items or even asked if the sentence is the same as ones heard previously, easing the task to some extent. The verb bias task, on the other hand, requires that participants listen to a somewhat complex sentence while looking at visually complex stimuli and then move as quickly as possible to one of two pictured interpretations. Differences in mouse movements therefore may have been more affected by differences in executive function or speed of processing. This could have had the effect of stabilizing the within-participant variability, and in fact the adults had greater within-participant variability on the choice of interpretation measures than the mouse trajectory measures.

Also of note is that the internal consistency of the measures is low likely in part because of the low number of items. We obtained a large number of negative ICC values across all measures, indicating great within-subject variability for both statistical learning and sentence processing measures. For measures with positive values, the number of items was often quite low (restricting to weakly biased verbs halves the dataset, and further restricting it to instrument-biased verbs halves it again; see **Tables 7, 8** for number of observations within each model). With more items, tasks might have been more reliable, and therefore more suited to showing a strong relationship. Although the study of the psychometric properties of statistical learning tasks is in its infancy, reliability-related issues may contribute to the difficulty in identifying experimental links between statistical tasks and measures of online language processing, especially with child participants (Arnon, 2019). It is possible better reliability

may be obtained through larger numbers of test items, but adding more items will impact the age of potential child participants capable of completing tasks. There is also the problem of increasing participants' exposure to ungrammatical combinations with more test items in an artificial grammar learning task, which could influence results and how results are interpreted.

Attention to the psychometric properties of statistical learning and online comprehension tasks would strengthen the inferences possible in this field. As noted by Arnon (2019), increasing the type of items tested (Siegelman et al., 2017) and using online methodologies (Siegelman et al., 2018b) are two ways to improve the reliability of statistical learning tasks without fatiguing child participants. We hope to take steps in future research that will allow us to improve reliability in these types of measures when used with young children.

## Developmental Differences

We found clear diagnostic group differences by adults in the verb task (Hall et al., 2019) but not in the artificial grammar learning task (Hall et al., 2017). This may have been due to greater task demands in the verb bias task related to the complexity of real language and the variation in individuals' experience with language. Although the language that people with DLD hear may not differ substantially from what people with TD hear (Leonard, 2014; though Karmiloff-Smith, 2009, has questioned whether impairments or deficits are compounded upon by how others interact with individuals with developmental disorders), the accumulated experience of DLD may result in a linguistic experience that is not as rich or deep as peers. For example, difficulty learning language could result in weaker semantic and syntactic representations (see Sheng and McGregor, 2010; Alt et al., 2013; Haebig et al., 2017), which then limit the expressiveness of the individual's own speech, as well as the efficiency and the precision with which the individual understands others' speech. This may have a cascading effect, such that weak representations in childhood limit how later information is stored – even if the child is exposed to the same information – or may lead to the child seeking out different types of interactions as they age, leading to actually different experiences and interactions in adolescence. In the artificial grammar learning task, on the other hand, the language is tightly controlled. There are no competing experiences, the learning is not under time pressure, the exposure period is small, and all participants have identical exposures. There is little variation; in short, it is a toy language, and the controlled nature of the language likely made it an easier task than the verb bias task for all participants.

Snedeker and Trueswell (2004) demonstrated that verb bias cues outweighed referential cues for 5-year-old children, which differentiated them from the adult participants in their study of verb bias. Although children made choices based squarely on verb bias, their eye movements indicated emerging consideration for referential cues (the number of animals present influenced how often they looked at objects indicating a modifier interpretation). It is possible that at ages 7–9, children are now learning to integrate and weight different cues during sentence processing

and not relying so squarely on verb bias. However, the process is not complete, and so we still see differences between them and adults. That we see the opposite pattern of Snedeker and Trueswell (2004) should not be surprising because we did not have referential cues here. And so, although we know that children are capable of learning from distributional information similarly to adults (Hall et al., 2018b) and that they are sensitive to the different distributional properties of verbs in the task we used, we can infer that children show different patterns of interpretation than adults in both their mouse movements and overt choices because they are in the process of learning to use other cues to interpret sentences.

Regarding development and verb bias, it is possible that our tasks were not as similar as we hoped, in that although they both involved distributional information, our statistical learning task involved tracking adjacent dependencies (which words occurred next to each other) whereas verb bias, in this case, was a non-adjacent dependency (a noun appeared between the verb and the ambiguous *with the x* phrase). It is possible that working memory or other cognitive limitations could have impacted children's performance on the verb bias task differently from the statistical learning task (e.g., Thiessen et al., 2013), and that these differences may have been better captured in a task that tracked learning of non-adjacent distributional information. This is an example of how a specific type of statistical learning may be more relevant for an emerging skill at this point in development than another.

It is important to continue to study distributional learning using both real and artificial language stimuli to better understand the mechanisms involved and how they might facilitate grammar acquisition and use in typical populations and in populations with DLD. Furthermore, both cross-sectional and longitudinal studies should assess performance in the intervening periods between early childhood and adulthood to better elucidate the developmental trajectories of both typical and atypical populations (McMurray et al., 2010, 2018; Rigler et al., 2015).

## GENERAL DISCUSSION

In the present work, we examined the roles of grammatical proficiency, text exposure, and statistical learning for explaining individual differences in sentence processing by children and adults with and without DLD. Our general purpose was to better understand the relationship between statistical learning and language learning and processing. In Study 1, we found that children with DLD and their TD peers showed some sensitivity to verb bias in an implicit mouse tracking measure even when their explicit behavior did not reflect this sensitivity. Instability in the formation of verb biases as part of typical development may have contributed to the pattern of findings for children. In our second study, we found that our measure of statistical learning predicted how adults interpreted ambiguous sentences using verb bias in only one of 48 possible models. It is possible that we saw stronger evidence for a relationship for TD adults than adults with DLD because adults with DLD

use verb bias information differently or do not have the same access to the information during sentence processing as TD peers. However, findings for a relationship were not robust and reliability for all measures was quite low. Together, these results suggest that language processing is influenced by the statistics in the language environment and one's ability to attend to them and use them, but we need more reliable tasks to better detect and understand this relationship. The performance by children in comparison to adults on the verb bias task, taken in combination with the findings on statistical learning, suggests that there may be differences between initial learning of statistical information in linguistic environments and using that information efficiently during complex language processing combined with other cues.

While we acknowledge some uncertainty about which aspects of the verb bias stimuli may have affected how children interpreted sentences, results from Study 1 provide novel insights. We now know that, although children with and without DLD may be sensitive to verb bias information, the global instrument bias and integration of visual and linguistic cues also affects how verb bias information is deployed during processing, and likely has a long developmental timeline. Unique contributions of Study 2 are that TD adults may rely on the same mechanisms to learn from distributional information and to predict distributional information while processing language. These results provide further evidence that statistical learning may contribute to variation in how individuals process and interpret language, at least in adulthood. These studies allow more nuanced discussion of the mechanisms responsible for efficient sentence processing and the developmental timescales of these mechanisms.

Results from these studies provide further evidence that verb bias continues to develop beyond school age, and that differences observed in adults with and without DLD suggest that that verb bias is an area of weakness in DLD, albeit one that may appear somewhat hidden until later in development when verb biases and cue integration are more fully formed in the TD population. Given the impact that verb bias can have on comprehension and communication, it is an area worthy of further study.

Results from Study 2 illustrate a need for more transparent methods for reporting results from studies of complex mechanisms, such as those purported to support multifaceted skills like language. Because *post hoc* explanations of data are tempting and often seem rational given trends in the literature, we recommend best practice be either using highly transparent methodology such as multiverse analyses or preregistering both the tasks chosen and the final measures along with potential explanations and predicted outcomes to reduce the temptation to report only the most exciting findings. We also encourage researchers to examine tasks' psychometric properties and report measures of reliability in studies of statistical learning and language processing. For progress in research into the cognitive science of language, a commitment to open science is necessary to ensure that results can be verified and replicated. Without extensive reporting of how and why variables were chosen and measured, our work will always be exploratory.



## DATA AVAILABILITY STATEMENT

Code to run analyses in R version 3.5.1 (R Core Team, 2018) and sample data are available on github at <https://github.com/jessica-hall/multiverse/>. Full datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Iowa Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

JH was the lead author of this manuscript, contributed the majority of the writing during her postdoc at the University of Arizona, conducted the experiments, and analyzed data as part of her dissertation requirement while a student at the University of Iowa. AO and TF provided substantial insight and ideas during the development of stimuli and data analysis. AO contributed to the writing and editing of drafts of these datasets, and provided the insight into developmental language disorder at different ages as well as access to populations seen in her labs.

## REFERENCES

- Alt, M., Meyers, C., and Alt, P. M. (2013). Using ratings to gain insight into conceptual development. *J. Speech Lang. Hear. Res.* 56, 1650–1661. doi: 10.1044/1092-4388(2013/11-0317)
- American Speech-Language-Hearing Association (1997). *Guidelines for Audiologic Screening*. Available at: <http://www.asha.org/policy/GL1997-00199/>
- Arciuli, J. (2017). The multi-component nature of statistical learning. *Philos. Trans. R Soc. B* 372:20160058. doi: 10.1098/rstb.2016.0058
- Arciuli, J., and Conway, C. M. (2018). The promise—and challenge—of statistical learning for elucidating atypical language development. *Curr. Dir. Psychol. Sci.* 27, 492–500. doi: 10.1177/0963721418779977
- Arnott, I. (2019). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behav. Res. Methods*. doi: 10.3758/s13428-019-01205-5 [Epub ahead of print].
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Catalise-2 Consortium, Adams, C., et al. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *J. Child Psychol. Psychiatry* 58, 1068–1080. doi: 10.1111/jcpp.12721
- Carroll, C., and Dockrell, J. (2010). Leaving special school: post-16 outcomes for young adults with specific language impairment. *Eur. J. Special Needs Educ.* 25, 131–147. doi: 10.1111/j.1460-6984.2012.00166.x
- Conti-Ramsden, G., and Durkin, K. (2007). Phonological short-term memory, language and literacy: developmental relationships in early adolescence in young people with SLI. *J. Child Psychol. Psychiatry* 48, 147–156. doi: 10.1111/j.1469-7610.2006.01703.x
- Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., and Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *Int. J. Lang. Commun. Disord.* 53, 237–255. doi: 10.1111/1460-6984.12338

## FUNDING

Data collection and analysis were supported by NIH-NIDCD F31DC015370 awarded to JH while at the University of Iowa, and NIH-NIDCD 5R01DC011742 awarded to Dr. Karla McGregor, who was a primary sponsor on JH's grant. Writing was supported by NIH-NIDCD F32DC017373 awarded to JH at the University of Arizona.

## ACKNOWLEDGMENTS

We thank Karla McGregor for her guidance and feedback in the creation of this manuscript, Tim Arbisi-Kelm and Nichole Eden for their assistance carrying out these studies, Susan Cook for introducing us to the multiverse method, and Bob McMurray, Gerry Altmann, and Sarah Brown-Schmidt for their input on data analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00402/full#supplementary-material>

- Dawson, J. I., Stout, C. E., and Eyer, J. A. (2003). *Structured Photographic Expressive Language Test—Third Edition (SPELT-3)*. DeKalb, IL: Janelle Publications.
- Dunn, L. M., and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4)*. Minneapolis, MN: Pearson Assessments.
- Erickson, L. C., and Thiessen, E. D. (2015). Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Dev. Rev.* 37, 66–108. doi: 10.1016/j.dr.2015.05.002
- Evans, J. L., Saffran, J. R., and Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* 52, 321–335. doi: 10.1016/B978-0-444-52891-9.00022-1
- Farmer, T. A., Cargill, S. A., and Spivey, M. J. (2007). Gradiency and visual context in syntactic garden-paths. *J. Mem. Lang.* 57, 570–595. doi: 10.1016/j.jml.2007.04.003
- Farmer, T. A., Monaghan, P., Misyak, J. B., and Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: a reply to Staub et al. (2009). *J. Exp. Psychol.* 37, 1318–1325. doi: 10.1037/a0023063
- Fine, A. B., and Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cogn. Sci.* 37, 578–591. doi: 10.1111/cogs.12022
- Fine, A. B., Jaeger, T. F., Farmer, T. A., and Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One* 8:e77661. doi: 10.1371/journal.pone.0077661
- Freeman, J. B., and Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav. Res. Methods* 42, 226–241. doi: 10.3758/BRM.42.1.226
- Freeman, J. B., Ambady, N., Rule, N. O., and Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *J. Exp. Psychol.* 137, 673–690. doi: 10.1037/a0013875
- Freeman, J. B., Dale, R., and Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Front. Psychol.* 2:59. doi: 10.3389/fpsyg.2011.00059
- Haebig, E., Weber, C., Leonard, L. B., Deevy, P., and Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *J. Neurodev. Disord.* 9:22. doi: 10.1186/s11689-017-9201-1

- Hall, J. E., Owen Van Horne, A. J., and Farmer, T. A. (2018a). Distributional learning aids linguistic category formation in school-age children. *J. Child Lang.* 45, 717–735. doi: 10.1017/S0305000917000435
- Hall, J. E., Owen Van Horne, A. J., McGregor, K. K., and Farmer, T. A. (2018b). Individual and developmental differences in distributional learning. *Lang. Speech Hear. Serv. Sch.* 49, 694–709. doi: 10.1044/2018\_LSHSS-STLT1-17-0134
- Hall, J. E., Owen Van Horne, A. J., McGregor, K. K., and Farmer, T. (2017). Distributional learning in college students with developmental language disorder. *J. Speech Lang. Hear. Res.* 60, 3270–3283. doi: 10.1044/2017\_jslhr-17-0013
- Hall, J. E., Owen Van Horne, A. J., McGregor, K. K., and Farmer, T. A. (2019). Deficits in the use of verb bias information in real-time processing by college students with developmental language disorder. *J. Speech Lang., Hear. Res.* 62, 337–355. doi: 10.1044/2018\_JSLHR-L-17-0267
- Hesketh, A., and Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of specific language impairment. *PLoS One* 8:e56314. doi: 10.1371/journal.pone.0056314
- Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *Int. J. Lang. Commun. Disord.* 36, 149–171. doi: 10.1080/13682820118418
- Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149. doi: 10.1037//0033-295x.99.1.122
- Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Dev. Psychol.* 45, 56–63. doi: 10.1037/a0014506
- Kaufman, A. S., and Kaufman, N. L. (2004). *K-BIT-2: Kaufman brief intelligence test*, 2nd Edn. Circle Pines, MN: American Guidance Service, Inc.
- Kelly, D. J., and Rice, M. L. (1994). Preferences for verb interpretation in children with specific language impairment. *J. Speech Lang. Hear. Res.* 37, 182–192. doi: 10.1044/jshr.3701.182
- Kidd, E. (2012). Individual differences in syntactic priming in language acquisition. *Appl. Psycholinguist.* 33, 393–418. doi: 10.1037/a0025405
- Kidd, E., and Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Dev.* 87, 184–193. doi: 10.1111/cdev.12461
- Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends Cogn. Sci.* 22, 154–169. doi: 10.1016/j.tics.2017.11.006
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lammertink, I., Van Witteloostuijn, M., Boersma, P., Wijnen, F., and Rispens, J. (2019). Auditory statistical learning in children: novel insights from an online measure. *Appl. Psycholinguist.* 40, 279–302. doi: 10.1017/s01421716418000577
- Lany, J., Shoaib, A., Thompson, A., and Estes, K. G. (2018). Infant statistical-learning ability is related to real-time language processing. *J. Child Lang.* 45, 368–391. doi: 10.1017/S0305000917000253
- Leonard, L. B. (2014). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.
- Lewis, R. L., Vasisht, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007
- López-Barroso, D., Cucurell, D., Rodríguez-Fornells, A., and de Diego-Balaguer, R. (2016). Attentional effects on rule extraction and consolidation from speech. *Cognition* 152, 61–69. doi: 10.1016/j.cognition.2016.03.016
- Mani, N., and Huetting, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *J. Exp. Child Psychol.* 126, 264–279. doi: 10.1016/j.jecp.2014.05.004
- Martin, A. E. (2016). Language processing as cue integration: grounding the psychology of language in perception and neurophysiology. *Front. Psychol.* 7:120. doi: 10.3389/fpsyg.2016.00120
- McMurray, B., Danelz, A., Rigler, H., and Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Dev. Psychol.* 54, 1472–1491. doi: 10.1037/dev0000542
- McMurray, B., Samelson, V. M., Lee, S. H., and Tomblin, J. B. (2010). Individual differences in online spoken word recognition: implications for SLI. *Cognit. Psychol.* 60, 1–39. doi: 10.1016/j.cogpsych.2009.06.003
- Misyak, J. B., and Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331. doi: 10.1111/j.1467-9922.2010.00626.x
- Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Front. Psychol.* 1:31. doi: 10.3389/fpsyg.2010.00031
- Montag, J. L., and MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *J. Exp. Psychol.* 144, 447–468. doi: 10.1037/xge0000054
- Nation, K., Marshall, C. M., and Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *J. Exp. Child Psychol.* 86, 314–329. doi: 10.1016/j.jecp.2003.09.001
- Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn., Affect., Behav. Neurosci.* 5, 263–281. doi: 10.3758/cabn.5.3.263
- Owen Van Horne, A. J., Curran, M., Larson, C., and Fey, M. E. (2018). Effects of a complexity-based approach on generalization of past tense-ed and related morphemes. *Lang., Speech Hear. Serv. Sch.* 49, 681–693. doi: 10.1044/2018\_LSHSS-STLT1-17-0142
- Owen Van Horne, A. J., Fey, M., and Curran, M. (2017). Do the hard things first: a randomized controlled trial testing the effects of exemplar selection on generalization following therapy for grammatical morphology. *J. Speech Lang. Hear. Res.* 60, 2569–2588. doi: 10.1044/2017\_JSLHR-L-17-0001
- Perfetti, C. A., Van Dyke, J., and Hart, L. (2001). The psycholinguistics of basic literacy. *Annu. Rev. Appl. Linguist.* 21, 127–149. doi: 10.1017/s0267190501000083
- Perona, K., Plante, E., and Vance, R. (2005). Diagnostic accuracy of the structured photographic expressive language test. *Lang. Speech Hear. Serv. Sch.* 36, 103–115. doi: 10.1044/0161-1461(2005/010)
- Peter, M. S., Chang, F., Pine, J. M., Blything, R., and Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *J. Mem. Lang.* 81, 1–15. doi: 10.1016/j.jml.2014.12.002
- R Core Team, (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna: R Core Team.
- Reeder, P. A., Newport, E. L., and Aslin, R. N. (2013). From shared contexts to syntactic categories: the role of distributional information in learning linguistic form-classes. *Cogn. Psychol.* 66, 30–54. doi: 10.1016/j.cogpsych.2012.09.001
- Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., and McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Dev. Psychol.* 51, 1690–1703. doi: 10.1037/dev0000044
- Ryskin, R. A., Qi, Z., Duff, M. C., and Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *J. Exp. Psychol.* 43, 781–794. doi: 10.1037/xlm0000341
- Schneider, J. M., Abel, A. D., Ogiela, D. A., McCord, C., and Maguire, M. J. (2018). Developmental differences in the neural oscillations underlying auditory sentence processing in children and adults. *Brain Lang.* 186, 17–25. doi: 10.1016/j.bandl.2018.09.002
- Schneider, J. M., Abel, A. D., Ogiela, D. A., Middleton, A. E., and Maguire, M. J. (2016). Developmental differences in beta and theta power during sentence processing. *Dev. Cogn. Neurosci.* 19, 19–30. doi: 10.1016/j.dcn.2016.01.001
- Schneider, J. M., and Maguire, M. J. (2018). Developmental differences in the neural correlates supporting semantics and syntax during sentence processing. *Dev. Sci.* 2019:e12782. doi: 10.1111/desc.12782
- Shankweiler, D., Mencl, W. E., Braze, D., Tabor, W., Pugh, K. R., and Fulbright, R. K. (2008). Reading differences and brain: cortical integration of speech and print in sentence processing varies with reader skill. *Dev. Neuropsychol.* 33, 745–775. doi: 10.1080/87565640802418688
- Sheng, L., and McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *J. Speech Lang. Hear. Res.* 53, 146–159. doi: 10.1044/1092-4388(2009/08-0160)

- Siegelman, N., Bogaerts, L., Christiansen, M. H., and Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160059.
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., and Frost, R. (2018a). Linguistic entrenchment: prior knowledge impacts statistical learning performance. *Cognition* 177, 198–213. doi: 10.1016/j.cognition.2018.04.011
- Siegelman, N., Bogaerts, L., Kronenfeld, O., and Frost, R. (2018b). Redefining “learning” in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cogn. Sci.* 42, 692–727. doi: 10.1111/cogs.12556
- Snedeker, J., and Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognit. Psychol.* 49, 238–299. doi: 10.1016/j.cogpsych.2004.03.001
- Spit, S., and Rispens, J. (2019). On the relation between procedural learning and syntactic proficiency in gifted children. *J. Psycholinguist. Res.* 48, 417–429. doi: 10.1007/s10936-018-9611-6
- Spivey, M. J. (2007). *The Continuity of Mind*. New York, NY: Oxford University Press.
- Spivey, M. J., Grosjean, M., and Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10393–10398. doi: 10.1073/pnas.0503903102
- Steegen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.* 11, 702–712. doi: 10.1177/1745691616658637
- Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (2013). The extraction and integration framework: a two-process account of statistical learning. *Psychol. Bull.* 139, 792–814. doi: 10.1037/a0030801
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., and O’Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40, 1245–1260. doi: 10.1044/jslhr.4006.1245
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., and MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognit. Psychol.* 58, 250–271. doi: 10.1016/j.cogpsych.2008.08.002

**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.

The reviewer GS-C declared a shared affiliation, with no collaboration, with one of the authors, TF, to the handling Editor at the time of review.

Copyright © 2019 Hall, Owen Van Horne and Farmer. 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.



# Maintenance Versus Transmission Deficits: The Effect of Delay on Naming Performance in Aphasia

Nadine Martin<sup>1\*</sup> and Gary S. Dell<sup>2</sup>

<sup>1</sup> Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA, United States, <sup>2</sup> Beckman Institute and Department of Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

## OPEN ACCESS

### Edited by:

Vitória Plai,  
Radboud University Nijmegen,  
Netherlands

### Reviewed by:

Maryellen C. MacDonald,  
University of Wisconsin–Madison,  
United States  
Royce Anders,  
Université de Lyon, France

### \*Correspondence:

Nadine Martin  
nmartin@temple.edu

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 19 July 2019

**Accepted:** 01 November 2019

**Published:** 27 November 2019

### Citation:

Martin N and Dell GS (2019)  
Maintenance Versus Transmission  
Deficits: The Effect of Delay on  
Naming Performance in Aphasia.  
Front. Hum. Neurosci. 13:406.  
doi: 10.3389/fnhum.2019.00406

We propose that deficits in lexical retrieval can involve difficulty in transmission of activation between processing levels, or difficulty in maintaining activation. In support, we present an investigation of picture naming by persons with aphasia in which the naming response is generated after a 1 s (sec) cue to respond in one condition or a 5 s cue to respond in another. Some individuals did better after 5 s, some did worse after 5 s, and some were not impacted by the delay. It is suggested that better performance after 5 s indicates a transmission deficit and that worse performance after 5 s indicates a maintenance deficit. To support this hypothesis, we adapted the two-step semantic-phonological model of lexical retrieval (Schwartz et al., 2006) so that it can simulate the passage of time and can simulate lesions in transmission (its semantic and phonological connection strength parameters) and/or maintenance (its decay parameter). The naming error patterns after 1 and 5 s for each participant were successfully fit to the model. Persons who did better after 5 s were found to have low connection strength parameters, persons who did worse after 5 s were simulated with an increased decay rate, and persons whose performance did not differ with delay were found to have lesions of both types. Some potential theoretical and clinical implications are discussed.

**Keywords:** short-term memory, naming, temporal processing, word retrieval, aphasia

## INTRODUCTION

Aphasia, a language impairment that follows brain damage is accompanied by reduced verbal short-term memory (STM) capacity that is commensurate with the severity of the language impairment (Martin and Saffran, 1997; Martin and Ayala, 2004; Martin and Gupta, 2004). We attribute the association between aphasia and reduced verbal STM to a very old idea: a person's ability to maintain the semantic or phonological representations of words depends on mechanisms that carry out the retrieval of these representations when speaking and listening. If one has difficulty producing and understanding a word, one will have trouble maintaining it. Although this claim is often made (Berndt and Mitchum, 1990; Saffran, 1990; Martin et al., 1996) and debated (Shelton et al., 1992; Martin and Freedman, 2001; Martin R.C., 2005), it has been difficult to specify with sufficient precision that it can be used to understand and remediate aphasia. In this paper, we describe a model of aphasia that may explain the production and short-term maintenance of single words and present data that test the model.



Computational models of language production and aphasia that are based on spreading activation (e.g., Dell et al., 1997, 2004; Foygel and Dell, 2000; Rapp and Goldrick, 2000; Ueno et al., 2011; Walker and Hickok, 2016) represent words and their sounds as a network of units connected through weighted links. Aphasic production deficits are viewed as a failure of spreading activation to activate the correct units, relative to incorrect ones, explaining the nature and frequencies of paraphasias that occur in picture naming or word repetition tasks. These models attribute aphasia to either a transmission failure (e.g., weak or noisy connection weights) or a failure to maintain activation of a unit (e.g., overly fast decay of activation). In our work, we have used both accounts to simulate aphasia, and particularly to simulate individual persons with aphasia as opposed to aphasic syndromes (Martin et al., 1994, 1996; Schwartz et al., 2006; Dell et al., 2007, 2013; Nozari et al., 2010). It turns out, though, that transmission and maintenance failures are difficult to distinguish. One can fail to activate the *k*/of the target “cat” because connections to it are weak, or because its activation decays away.

Although current models have been able to distinguish the “*where*” of a deficit (e.g., lexical-semantic vs. lexical-phonological), they are less able to distinguish deficit “*mechanisms*,” e.g., transmission vs. maintenance (Foygel and Dell, 2000). There are two reasons for this. First, most data on lexical deficits in aphasia come from production tasks that do not manipulate or measure the temporal dynamics of production. Second, the models themselves make no claims about the passage of time and its effects on accuracy of word retrieval. An exception to these generalizations involved early studies that found that the rare semantic errors made in word repetition tasks by persons with aphasia can be promoted when there is more time before the response (e.g., Martin et al., 1994, 1996). In recent work, we have investigated the effects of time passage on word retrieval and we proposed that such errors are caused specifically by an overly strong decay of activation.

We have investigated the effects of time passage on word retrieval by adding a temporal component (response delay) to word retrieval tasks (Martin et al., 1996, 2018; Martin and Dell, 2017). These studies have revealed some intriguing findings that we investigate further in this study: some aphasic individuals perform more poorly after a time delay while others benefit from additional time to respond. Here, we provide some data from the Temple Assessment of Language and Short-term memory in Aphasia (TALSA; Martin et al., 2018) that demonstrate the change in accuracy of naming following a response delay and (2) test the hypothesis that better or worse performance on delayed naming tasks maps onto deficits of transmission or maintenance, respectively. To test this claim, we created a new version of the model of word production, the Semantic-Phonological Model (SP), which has been used in many of our studies of word production, but most recently in a study of Dell et al.’s (2013) that identified the neural correlates of semantic and phonological components of word processing.

## The Present Study

For this study, we adapted the SP model to better represent the passage of time and treated both connection strength values and

decay rate as lesionable parameters. Both of these alterations were necessary to apply the model to data showing changes in accuracy of word production after a 5 s response delay. We demonstrate that reduced connection strength can account for performance that improves after 5 s and increased decay rate can explain worse performance after a response delay. We also show that the new SP model, which we call the “slow” SP-decay model, can directly fit the error proportions in naming that occur after different response delays, including worse performance, better performance and no change in accuracy levels. Also, as in our previous modeling work, the goal is not just to model overall correctness, but also the proportions of the error types, such as non-word errors and various kinds of lexical errors.

The first part of the study is empirical. We sought behavioral evidence for temporal dimensions of impairment in lexical processing by evaluating the picture naming performance of individuals with chronic aphasia. We administered the picture naming test under two response delay conditions (1 and 5 s), allowing us to observe the effects of a time delay on accuracy. Based on a prior study (Martin and Dell, 2017), we expected to find a few individuals that were worse after a 5 s response delay, while for others, accuracy would increase after a 5 s response delay. We also expected many to show little difference, or at least differences that are not easily detectable.

The second part of the study introduces the slow SP model. Unlike most previous versions of the model, it represents the passage of time so that response delays can be modeled and includes decay as a lesionable parameter. The expectation is that the naming error pattern made by individuals whose performance is worse after a 5 s delay could be characterized by a weak (larger) decay parameter, and individuals whose naming benefits from the extra time in the 5 s delay could instead be fit by assuming weak (lower) connection strength parameters. The data and model fitting potentially have both theoretical and clinical implications. They can test the temporal assumptions of the model and can identify deficits and potential treatments related to those aspects.

**Part 1. The effects of response delay on accuracy of picture naming in people with aphasia.**

## MATERIALS AND METHODS

### Participants

#### Participants With Aphasia

The 90-item picture naming subtest of the Temple Assessment of Language and Short-term Memory in Aphasia (TALSA) was administered in two different time periods, with slight differences in the administration format, but no difference in the item content (see details in description below). In the most recent administration of the test (2015–2018), 24 people with chronic aphasia completed the TALSA naming test. In an earlier administration period (2008–2012), 21 people with chronic aphasia completed the test but six of these individuals were among those who were tested in the 2015–2018 period. For these six, we used the data sets from the most recent testing period. Thus, there were 39 participants (15 from the early

testing period and 24 from the recent one). The classical aphasia types represented in this group included, Broca, Wernicke, Conduction, Anomia, and Transcortical Motor. Participants with aphasia were at least 6 months post-onset and had single or multiple left-hemisphere lesions resulting from a cerebrovascular accident (CVA).

**Table 1** shows the etiologies of aphasia, months post-onset at the time of testing, the aphasia quotient from the Western Aphasia Battery-Revised (Kertesz, 2006) and the period in which they were tested (2008–2012 or 2015–2018). There were 15 female and 24 males in the sample. Average age was 57 years [standard deviation (SD): 9.48 range: 32–78]. The average number of months post-onset at time of testing was 82 months (SD: 77.79) and ranged from 6 to 333 months. All but one of the participants were high school educated, and the years of education ranged from 7 to 19 years, with an average of 14 years (SD: 2.57).

All participants were administered the Western Aphasia Battery-Revised (Kertesz, 2006). This standardized screening test for language abilities in aphasia assesses language abilities such as naming, repetition and comprehension. It yields an Aphasia Quotient summarizing the overall language ability with a score between 0 and 100. The average Aphasia Quotient for this sample of people with aphasia ( $n = 39$ ) was 75 (SD: 17.10) and scores ranged from 33.8 to 100.

### Control Participants

Eleven individuals without aphasia or brain damage completed the same TALSA picture naming test that the persons with aphasia did. There were three males and eight females with an average age of 66.43 years (SD: 10.1 range: 47–80). Years of education ranged from 12 to 20 with an average of 15.57 years (SD: 2.80).

All participants voluntarily enrolled in this research program and signed a consent form approved by the Internal Review Board at Temple University.

## Materials

### Naming Test

All participants completed the TALSA's 90-item picture naming (Martin et al., 2018). Picture names were 1–3 syllables in length. We used frequency ratings from Pastizzo and Carbone (2007) and divided the stimuli into high frequency (>25 occurrences per million, range 27 to 673) and low frequency (<25 occurrences per million).

### Administration of the Naming Test

As noted above, we administered two versions of the test that varied in the format of administration, but not in content. The pictures and target names were identical in both versions of the test. In the first version, administered between 2008 and 2012, the 90 picture items were divided into three sets of 30 items. Each set was assigned to one of three response delay conditions, 1 s unfilled, 5 s unfilled and 5 s filled delay. Syllable length and word frequency were balanced across all three sets. After each of the three sets was administered in one of the three response delay conditions, they were then administered a second and third time (in separate testing sessions) in the other two response delay

conditions. Thus, all 90 stimuli were administered in all three response delay conditions. For this study, we report only the data from the first two conditions, 1 s unfilled and 5 s unfilled, as we were interested in the effects of time, but not interference.

In 2015, we revised the administration of the test, presenting the 90 picture items blocked in a single response delay condition (e.g., 1 s unfilled response delay) and then in separate sessions, administering the same 90 items (randomized order) in the other two response delay conditions. The order of administering the test in the three response delay conditions was randomized across participants. Again, only the data from the unfilled conditions are reported in this study.

### Testing Procedure

For both versions, pictures were presented on a computer via e-prime software (Psychology Software Tools Inc., 2012) for 4 s, with a beep cue to name the picture 1 s (1-sec) or 5 s (5-sec) after it went off the screen. The next picture was presented 4 s after this cue and hence the participant had to respond within this 4 s period.

### Scoring Procedures

Scoring and response categorization followed the guidelines of the Philadelphia Naming Test (Roach et al., 1996; Dell et al., 1997). The first complete response was counted as the response of interest. This is the first naming attempt with minimally a consonant-vowel or vowel-consonant (with schwa not being counted as a vowel). Attempts should not be self-interrupted, have a clear downward or upward intonation and is followed by a distinct pause.

### Reliability of Scoring

Naming responses were transcribed by two research speech-language pathologists in the Aphasia Rehabilitation Research Laboratory where the testing took place.

Inter-rater reliability of scoring was evaluated using Cohen's Kappa statistic on a random selection of participants (8 test administrations) which accounted for 15% of the data from the 2015–2018 sample. There was substantial agreement between the two scorers,  $k = 0.774$  ( $p < 0.000$ ) (Landis and Koch, 1977).

## RESULTS

### Control Participants

For the 11 control participants the average score on the 1 s unfilled response delay condition was 0.98 (SD: 0.02, range: 0.94–1.00). On the 5 s unfilled response delay condition, the average score was 0.98 (SD: 0.02, range: 0.93–1.00).

### Participants With Aphasia

From the 39 sets of data collected during the two testing periods, we removed 12 whose correct naming proportion on the both the 1 and 5 s response delay conditions was greater than 0.90 correct. The remaining 27 sets of data were further analyzed to identify significant increases or decreases in accuracy following a 5 s response delay, in comparison to the 1 s response delay condition.

**TABLE 1 |** Information on (1) etiology of stroke, time post-onset and aphasia severity and (2) period when tested and inclusion of the participant's data in the naming analysis.

Participant ID	Time post-onset (months)	Etiology	WAB-R AQ <sup>1</sup>	Period tested	Qualified for naming analysis
FS1	11	LCVA <sup>2</sup> , interval resorption of hemorrhage within the left temporal lobe.	76.4	2008–2011	Y
SX3	291	LCVA, hemorrhagic infarct in the left basal ganglia.	92.8	2015–2018	Y
KC3	155	LCVA involving the middle and superior temporal gyrus, middle and inferior frontal gyrus, inferior parietal lobe (supramarginal and angular gyrus), and extending down to the lateral ventricle with damage to the basal ganglia. Temporal pole is preserved.	64.1	2015–2018	Y
EH4	123	LCVA, left hemisphere stroke with damage to the insular cortex, middle and inferior frontal gyrus extending to the parietal lobe. Temporal lobe is intact.	81.4	2015–2018	Y
CM5	82	LCVA, left parietal infarct with stable petechial hemorrhages of the bilateral centrum semiovale, cerebellum and brainstem.	70.0	2015–2018	N
DD6	78	LCVA, affecting left frontal parietal regions	55.6	2008–2011	Y
CT7	133	LCVA, left middle cerebral artery (MCA) infarct with extension into the posterior limb of the left internal capsule.	33.6	2008–2011	Y
MI10	86	LCVA	71.5	2008–2011	Y
EC15	103	LCVA	84.9	2008–2011	Y
TB16	69	LCVA, LMCA affecting watershed areas of LMCA/posterior cerebral artery (PCA) with hemorrhagic transformation	55.6	2008–2011	N
IU19	12	LCVA	82.0	2008–2011	N
SL21	106	Left CVA (parietal aneurysm).	89.0	2008–2011	Y
QH22	22	LCVA, left transverse and sigmoid cerebral sinus thrombosis with secondary bleeding into the ischemic zone; subsequent left hemispherectomy with evacuation of intracranial hemorrhage on the left side.	84.9	2008–2011	Y
EC25	333	LCVA, left frontoparietal craniotomy and clipping of left PCA aneurysm.	62.5	2015–2018	Y
KL27	37	LCVA, thalamic CVA	83.5	2008–2011	N
HI 28	21	Left occipital lobe infarct with several smaller satellite infarcts surrounding the posterior horn of the left lateral ventricle with several small acute infarcts within the left centrum semiovale and corona radiata; old right corona radiata and left sub-insular lacunar infarcts	65.3	2015–2018	Y
UN29	12	LCVA, left MCA, edema posterior aspect of left frontal lobe/left temporal parietal region with subacute petechial hemorrhage left basal ganglia and increased edema left caudate nucleus and left internal capsule consistent with evolving infarct.	33.8	2008–2011	Y
QC30	79	Left parietal infarct (non-hemorrhagic); chronic infarcts affecting left periventricular region, right corona radiata, bilateral basal ganglia and bilateral thalami.	93.2	2008–2011	N
SC32	14	LCVA	N/A	2008–2011	N
KU33	6	LCVA, left posterior temporal occipital lobe infarct.	90.5	2008–2011	Y
LT34	12	LCVA, left MCA occlusion involving the posterior left temporal lobe and left parietal lobe.	86.6	2008–2011	N
UP35	77	LCVA affecting the posterior 2/3s of the inferior frontal gyrus, subcortical white matter beneath the middle and superior frontal gyri, and the anterior superior insula cortex. The temporal lobe was intact.	88.6	2015–2018	N
DC37	23	LCVA, extensive left ACA, MCA and PCA infarctions. Extensive left craniotomy with stable subdural hemorrhage.	92.8	2015–2018	Y
KM38	224	LCVA, infarct affecting LMCA territory and portion of the LACA territory; extensive damage to frontal portions of the temporal and parietal lobes, down to lateral ventricles, sparing superior middle and frontal lobes. Insula and basal ganglia are severely damaged.	80.3	2015–2018	Y
CN39	27	LCVA, LMCA aneurysm and left internal carotid artery (ICA) occlusion, affecting left insular cortex, posterior 2/3s of the inferior frontal gyrus, inferior middle frontal gyrus, anterior margin of the angular gyrus and inferior insula. Temporal lobe is intact.	76.3	2015–2018	Y
HE41	71	LCVA, aneurysm in the region of the left MCA bifurcation/trifurcation, large hypoaattenuating lesions within the left temporal, frontal, and parietal lobes into the frontoparietal lobes.	74.8	2015–2018	Y
DC44	56: 5	LCVA, chronic infarct in left basal ganglia extending into left periventricular frontal white matter with mild to moderate chronic ischemic white matter changes in right thalamus with left Wallerian degeneration.	93.5	2015–2018	N
XH46	23	LCVA, LMCA infarct with damage to the middle and inferior frontal gyrus and anterior insula. Some damage extending to the basal ganglia (head of the caudate).	73.1	2015–2018	Y

(Continued)

TABLE 1 | Continued

Participant ID	Time post-onset (months)	Etiology	WAB-R AQ <sup>1</sup>	Period tested	Qualified for naming analysis
KG47	159	LCVA, infarct affecting territory of LMCA and LPCA including left parietal lobe and extending posteriorly to the left occipital lobe. There are tiny areas of acute ischemia in the cerebellar hemispheres, along the right parietal convexity and in the body of the corpus callosum. The insula and superior temporal gyrus anterior to temporoparietal junction are spared, but posterior white matter in the superior temporal gyrus is affected.	99.7	2015–2018	Y
UM48	44	LCVA, LMCA infarct affecting superior insula posteriorly, anterior parietal and posterior frontal lobes, as well as posterior middle and posterior inferior frontal gyrus. Indications of white matter ischemia.	89.2	2015–2018	N
KT53	25	LCVA, large left hemisphere lesion affecting posterior superior and posterior middle temporal gyri, with bulk of damage to inferior parietal lobule. Indications of posterior branch of MCA infarct. Medial parietal lobe is spared.	48.8	2015–2018	Y
NF54	31	LCVA, LMCA including the opercular region and extending posteriorly in the temporal lobe along the optic radiations	89.1	2015–2018	N
KK55	129	LCVA, moderate left frontal and temporal parietal infarct – left MCA distribution; hemorrhage medial to the left temporal region extending partially into left lenticular nucleus.	78.7	2015–2018	Y
MN56	114	LCVA, left temporoparietal region of hypodensity and sulcal effacement consistent with late acute/early subacute left MCA infarct.	81.1	2015–2018	N
BQ58	65	LCVA, large hypoattenuating lesions within the left temporal, frontal, and parietal lobes into the cortex of the frontoparietal lobes.	33.6	2015–2018	Y
BC60	25	Left anterior MCA distribution infarct involving the frontal and insular lobes, mild surface hypodensity may represent thrombosed MCA branches or petechial hemorrhages. Old tiny lacunar infarct in L caudate head.	71.4	2015–2018	Y
KG62	111	Large infarction left MCA distribution with small focal areas of hemorrhage; also complete occlusion of the left MCA secondary to thrombus.	66.3	2015–2018	Y
CI63	187	LCVA, left frontal MCA territory infarct and older left insular/frontoparietal infarct. Regions affected include posterior temporal lobe, inferior frontal gyrus, and insula. Most of the lesion is subcortical, with white matter projections from anterior temporal lobe interrupted. Subcortical structures indicate Wallerian degeneration and substantial deep white matter loss. Temporal pole is relatively preserved.	61.4	2015–2018	Y
DS68	13	LCVA, left insular Infarct and middle temporal lobe. Primary cortical damage is to the posterior insula, affecting white matter including the arcuate fasciculus. White matter damage extends from the posterior third ventricle to the anterior portion of the lateral ventricle.	82.7	2015–2018	Y

<sup>1</sup>WAB-R = Western Aphasia Battery-R (Kertesz, 2006). <sup>2</sup>LCVA, left cerebral vascular accident.

The final column of **Table 1** indicates whether a participant's data was included in this further analysis (Y) or not (N).

**Table 2** shows the proportions correct, the difference between proportions correct as a function of delay, and the significance and effect size (Cohen's *D*) of those differences. Six participants (22%) demonstrated a significant change in accuracy after a 5 s response delay. Three showed better performance after 5 s (KG47, CI63, and KC3) and three showed worse performance (DS68, SL21, and UN29).

Before we turn to modeling these data, it is useful to consider whether there are true differences due to delay in the sample, and whether these cases are possible examples of such differences. After all, in a sample of 27 individuals, one would expect one or two of them to be associated with a significant effect of delay by chance even if the manipulation had no true effect. The fact that there were six significant cases is somewhat reassuring. Perhaps more important is the

sizes of the effects obtained, as measured by Cohen's *D*, for example, for DS68,  $D = 0.58$ , for KG47,  $D = -0.92$ , for SL21,  $D = 0.52$  and UN29,  $D = 0.87$ . (A positive value indicates worse performance on the 1 s delay.) With 0.80 considered to be a large effect and 0.50, a medium effect, the effect sizes support the legitimacy of these differences. Although we cannot be certain that we have identified just those individuals whose naming is affected by the delay, we are reasonably confident that such people exist and that the set of six that we have selected includes some.

In the next study, we will use the new version of the model to test the hypothesis that better performance after a delay arises from a transmission deficit (low connection strengths) and worse performance arises from a maintenance deficit (increased decay). We also will test whether the model can in general fit the response patterns. Finally we will model data from three participants whose data



**TABLE 2 |** Performance on the TALSA naming test ( $n = 90$ ) with two testing conditions: 1- and 5-s response delay and test of the difference between these conditions.

Participant ID	Response delay		5-s minus 1-s	z Statistic	p Value	ln odds ratio	Cohen's D
	1-s	5-s					
DS68	0.77	0.53	−0.23	3.23	0.00	1.06	0.58
KG47	0.84	0.97	0.12	2.56	0.01	−1.68	−0.92
CI63	0.57	0.73	0.16	2.33	0.02	−0.74	−0.41
SL21	0.90	0.78	−0.12	2.18	0.03	0.94	0.52
KC3	0.67	0.80	0.13	2.01	0.04	−0.69	−0.38
UN29	0.10	0.02	0.08	1.99	0.05	1.59	0.87
KU33	0.67	0.79	−0.12	1.83	0.07	−0.63	−0.34
FS1	0.88	0.96	−0.08	1.81	0.07	−1.10	−0.60
XH46	0.52	0.64	−0.12	1.66	0.10	−0.51	−0.28
MI10	0.66	0.77	−0.11	1.64	0.10	−0.55	−0.30
EH4	0.83	0.73	0.10	1.62	0.11	0.60	0.33
KG62	0.52	0.63	−0.11	1.51	0.13	−0.46	−0.25
CN39	0.76	0.84	−0.09	1.48	0.14	−0.56	−0.31
QH22	0.81	0.88	−0.07	1.23	0.22	−0.51	−0.28
KT53	0.36	0.42	−0.07	0.92	0.36	−0.28	−0.15
EC15	0.87	0.83	0.04	0.63	0.53	0.26	0.14
MT50	0.68	0.71	−0.03	0.49	0.63	−0.16	−0.09
SX3	0.83	0.81	0.02	0.39	0.70	0.15	0.08
DC37	0.63	0.65	−0.02	0.31	0.76	−0.10	−0.05
BQ58	0.89	0.90	−0.01	0.24	0.81	−0.12	−0.06
HE41	0.83	0.82	0.01	0.20	0.84	0.08	0.04
HI28	0.81	0.80	0.01	0.19	0.85	0.07	0.04
CT7	0.74	0.73	0.01	0.17	0.87	0.06	0.03
KM38	0.71	0.70	0.01	0.16	0.87	0.05	0.03
DD6	0.66	0.64	0.02	0.16	0.88	0.05	0.03
EC25	0.68	0.68	0.00	0.00	1.00	0.00	0.00
KK55	0.88	0.88	0.00	0.00	1.00	0.00	0.00

showed little or no difference in accuracy in the 1 and 5 s response delay conditions, to show that such cases are also consistent with the model.

## Preparation of the Data for Modeling

As we explain below, the output of the interactive two-step model includes correct responses and five categories of errors: Semantic, Formal, Mixed, Unrelated, and Non-word errors. For the most part, responses on this test by people with and without aphasia fall into these categories. However, some responses fall into categories not produced in this model and some fall into the category of “Other” (e.g., naming just a part of the picture, man → shirt). When an “Other” error is made, that test item is removed from the total number of items tested. Two response types that are not produced by the model are ‘No Responses’ (saying nothing or otherwise reporting failure, e.g., “can’t”) and ‘Descriptions’ (providing a description of the portrayed object, e.g., “Some kind of animal, I think”). These responses are not removed from the analysis, but rather are distributed across the five model error types, in proportion to how frequently each of those error types occurs with that individual. Thus, this treatment does not change the proportion correct, nor does it change the relative proportions of the error types. The six sets of

response distributions that will be modeled are presented in Table 3.

## Part 2. Computational study. Modeling the transmission and maintenance deficits in naming.

The data from the naming study indicated two patterns of change in naming accuracy (better or worse) following a response delay. Here, we use the interactive two-step Semantic-Phonological (SP) model of word processing, to account for these patterns. The SP model of word retrieval consists of an interconnected network of semantic, lexical, and output phonological units, and a further set of connections between auditorily presented verbal input and the output phonological units (Figure 1). All connections are bidirectional, thus making the model’s flow of activation interactive. In naming, lexical access starts with a jolt of activation to the target word’s semantic features and then flows through the network. The activation function is linear with a decay component. Specifically, activation of a unit at a time step is equal to a fraction of its activation at the previous time step (the lost activation determined by the decay rate) plus any new activation delivered by its activated neighbors through weighted connections. Also, during each time step, a unit’s activation is perturbed by a normally distributed value with mean zero, and a standard deviation that is a linear function of the unit’s current activation (with a non-zero intercept). More

**TABLE 3 |** Participants with significant change in accuracy on picture naming test after a 5 s response delay: distributions of responses (proportions) after 1 and 5 s response delays.

Participant	Response delay	N	Correct	Semantic	Formal	Mixed	Unrelated	Non-word
KC3	1 s	n = 86	0.67	0.13	0.01	0.00	0.00	0.14
	5 s	n = 89	0.80	0.14	0.00	0.04	0.00	0.00
CI63	1 s	n = 90	0.56	0.07	0.03	0.00	0.01	0.34
	5 s	n = 89	0.73	0.01	0.03	0.00	0.00	0.21
KG47	1 s	n = 88	0.84	0.13	0.00	0.00	0.00	0.00
	5 s	n = 87	0.97	0.00	0.00	0.00	0.00	0.00
SL21	1 s	n = 89	0.91	0.02	0.00	0.01	0.00	0.06
	5 s	n = 85	0.82	0.01	0.00	0.03	0.01	0.16
UN29	1 s	n = 90	0.10	0.03	0.03	0.00	0.28	0.59
	5 s	n = 88	0.02	0.02	0.10	0.00	0.40	0.33
DS68	1 s	n = 90	0.77	0.08	0.02	0.00	0.00	0.13
	5 s	n = 90	0.53	0.18	0.00	0.00	0.00	0.29

activated units are noisier, but even units with no activation experience some noise. After a fixed number of time steps for activation to spread, the most active word unit of the appropriate grammatical category is selected, completing the first “step” of lexical access. Errors at this step are lexical (e.g., semantic, CAT→DOG; unrelated, CAT→LOG; formal, CAT→MAT, or mixed semantic-formal, CAT→RAT). A jolt of activation to the selected word unit initiates the second step. Activation then spreads throughout the network again for a fixed number of time steps, culminating in the selection of the most activated phonological units. Errors at this step are typically non-words (e.g., CAT→“cag”) but can also be formally related to the target word (e.g., CAT→“mat”). Errors occur because of noise and spreading activation, which activates units other than the target units. Please see Schwartz et al. (2006) for details.

The model has successfully simulated patterns of error by speakers with aphasia in: (1) *Naming*, by assuming there are weak connections between semantic and word units (parameter s) or word and phonological units (parameter p) (Schwartz et al.,

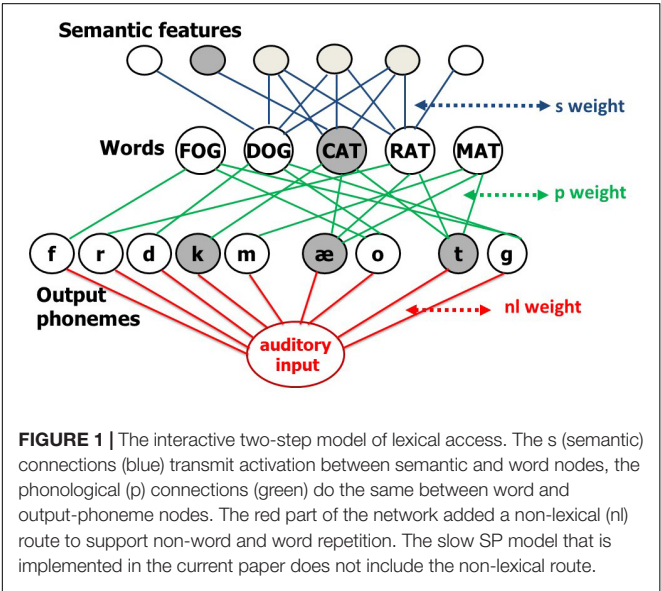
2006) and (2) *Word and non-word repetition* (Dell et al., 2007; Nozari et al., 2010), by including a mechanism that allows for production of phonological sequences that are not already stored in the lexicon (Hanley et al., 2004). This non-lexical route (**Figure 1**) lies in the connections between auditory input and output phonological units, and this connection strength is the parameter *nl*. Word repetition may involve both the non-lexical route and the lexical route corresponding to the second step of lexical access from meaning. To repeat a word, the model starts with a jolt of activation to the word unit and, for some individuals (see Nozari and Dell, 2013), a secondary jolt to the non-lexical route input unit.

The need to separate the s and p parameters is apparent from the error patterns of many of the persons with aphasia (e.g., Schwartz et al., 2006). A pattern with many non-word and formal errors, but few semantic errors suggests a low value of p, whereas a pattern with no non-word errors, but many lexical errors points to a low value of s. Also, it turns out that word repetition ability depends heavily on the value of p, and not on the value of s (Dell et al., 2013).

The current form of the SP model cannot be applied to the naming data obtained under different delays. In the model, activation spreads for a short and fixed period, essentially spreading all at once. Hence, there is no mechanism to explain how time affects processing. Consequently, we created the slow version of the naming model (without the non-lexical route) in which activation levels change more slowly and can be tracked over many time steps. We did this simply by reducing the amount of activation that spreads in each time step and the amount of decay that each unit undergoes in each step.

In the original model, normal performance was achieved with the s and p connection weights at 0.04 and with decay at 0.6. This yields a naming pattern of 97% correct, 2% semantic errors, and 1% mixed errors. Changing s and p to 0.0003 and decay to 0.001 and leaving other model properties unchanged creates a very similar model, except that activation patterns take more time to develop.<sup>1</sup>

<sup>1</sup>We emphasize that the slow model is not a response-time model in the way that other production models are (e.g., Levelt et al., 1999; Oppenheim et al., 2010;



**TABLE 4 |** Slow version of interactive activation model: proportion of naming responses correct at each time step in the SP model under two connection weight conditions.

Parameter settings	Time step							
	No delay					Delay		
	1	2	3	4	8	12	25	75
<b>Normal conditions</b>								
s = 0.0003	Decay = 0.001	0.62	0.85	0.91	0.94	0.97	0.97	0.72
p = 0.0003							0.91	
<b>Weight lesion</b>								
s = 0.0001	Decay = 0.001	0.09	0.14	0.21	0.28	0.47	0.59	0.51
p = 0.0001							0.65	

## Modeling Normal and Impaired Performance

Table 4 shows the simulation of normal performance and compares this to a lesion in the connection weights (parameters *s* and *p*), which reduces the transmission of activation in the network. Using the slow model, normal performance (97% correct) was simulated after between 8 and 20 time steps. Importantly, if we create a lesion in the weight parameters (reducing *s* and *p* to 0.0001), the model's accuracy is low, but improves with time. After 8 time steps, performance is poor (47% correct) but improves when more time passes (e.g., 65% at time step 25).

## Modeling the Pattern of Naming That Improves After a 5 s Response Delay

We then used the slow model to simulate the naming performance of the three people from the behavioral study

(Roelofs, 2014). It has no mechanism for varying selection time as a function of activation.

reported in Part 1 who showed significant improvement in naming following a 5 s response delay, KC3, KG47, CI63. These data are shown in Table 5 and include the proportion of correct and erroneous naming responses produced by each participant when naming was delayed by 1 and 5 s. Below those data are the proportions of correct and erroneous naming responses produced by the model after 8 time steps and after 25 time steps and the parameters used to fit the model to the naming pattern. We used 8 and 25 time steps to simulate 1- and 5-s response delays, respectively. We assumed that 1-s corresponds to 5 time steps and thus the 5-s delay corresponds to 25 steps. But at the 1 s delay, the actual naming response typically occurred a bit later than 1 s on average. Hence, we assumed eight steps for this delay.

The fitting process was informal, as our goal was only to establish whether the model's lesions can in principle create the kinds of differences that we see. We simply tried values of the *s* and *p* parameters that yielded performance in the range of each participant. For each case, the model captures the increase in accuracy after 5 s and also the changes in rates of different error types, especially a reduction in the non-word errors. To quantify the degree of fit, Table 5 shows the uncorrected root mean squared deviations (RMSDs) between the model and participant response-category proportions. The RMSD is calculated using the six proportions of each delay and thus there is a separate RMSD determined for the 1 s data and for the 5 s data.

## Modeling the Pattern of Naming That Becomes Worse After a 5 s Response Delay

Our next aim was to determine whether the slow SP model can account for the pattern of naming responses in which performance is worse after a 5 s response delay. It turns out

**TABLE 5 |** Modeling the pattern of better naming after a response delay.

Participant/ model	Model parameters	Response delay/ time steps	N	Response types						RMSD
				Correct	Semantic	Formal	Mixed	Unrelated	Non-word	
KC3		1-s delay	<i>n</i> = 86	0.67	0.13	0.01	0.00	0.00	0.14	
Model	$s^1 = 0.00009$	8 time steps		0.67	0.10	0.08	0.02	0.06	0.08	0.047
KC3	$p^2 = 0.0002$	5-s delay	<i>n</i> = 89	0.80	0.14	0.00	0.04	0.00	0.00	
Model	$DR^3 = 0.001$	25 time steps		0.79	0.08	0.04	0.02	0.02	0.05	0.038
KG47		1-s delay	<i>n</i> = 88	0.84	0.13	0.00	0.00	0.00	0.00	
Model	$s = 0.00011$	8 time steps		0.82	0.09	0.04	0.02	0.04	0.00	0.031
KG47	$p = 0.00045$	5-s delay	<i>n</i> = 87	0.97	0.00	0.00	0.00	0.00	0.00	
Model	$DR = 0.001$	25 time steps		0.88	0.07	0.01	0.02	0.01	0.00	0.047
CI63		1-s delay	<i>n</i> = 90	0.56	0.08	0.03	0	0	0.33	
Model	$s = 0.000019$	8 time steps		0.57	0.02	0.08	0.01	0.01	0.31	0.035
CI63	$p = 0.000109$	5-s delay	<i>n</i> = 89	0.73	0.01	0.03	0.00	0.00	0.22	
Model	$DR = 0.001$	25 time steps		0.71	0.04	0.05	0.01	0.01	0.18	0.024

<sup>1</sup>*s* = semantic weight parameter. <sup>2</sup>*p* = phonological weight parameter. <sup>3</sup>*DR* = decay rate parameter.

**TABLE 6 |** Slow version of interactive activation model: proportion of naming responses correct at each time step in the SP model, comparing connection weight, and decay lesions.

Parameter settings	Time step							
	No delay				Delay			
	1	2	3	4	8	12	25	75
<b>Normal conditions</b>								
s = 0.0003	Decay = 0.001	0.62	0.85	0.91	0.94	0.97	0.97	0.91
p = 0.0003								0.72
<b>Weight lesion</b>								
s = 0.0001	Decay = 0.001	0.09	0.14	0.21	0.28	0.47	0.59	0.65
p = 0.0001								0.51
<b>Decay lesion</b>								
s = 0.0003	Decay = 0.1	0.64	0.86	0.91	0.94	0.94	0.90	0.44
p = 0.0003								0.04

that the slow version of the model cannot simulate poorer performance after a delay if the possible lesions are restricted to the *s* and *p* parameters. What is needed is a postulation of a decay rate deficit as opposed to a deficit of connection strength. When a decay rate lesion is applied to the slow model (Table 6) and the *s* and *p* weight parameters are held constant, performance is better with no delay (94%) than at the delay (44% at step 25). In this way, the slow SP-decay model may explain the patient differences; in one case there is weakness in information transmission, in the other, there is a weakness in maintenance.

We used this model to simulate the naming performance of the three people from the behavioral study whose naming was significantly worse when a response was delayed by 5 s (SL21, UN29, and DS68, Table 7). The model captures the decrease in accuracy after 5 s as well as aspects of

the changes in error patterns, particularly the increase in non-word errors.

## Modeling the Pattern of Naming That Shows No Change in Accuracy After a 5 s Response Delay

Thus, far, the slow SP model accounts for those error patterns that become worse or better after a response delay. Can the model account for naming patterns that show no change after a 5 s response delay? We suspect that naming performance is not affected substantially by a 5 s response delay for many if not most people with aphasia. This was true for the sample. There are two types of individuals for whom delay matters little (according to the model). First, there are individuals whose performance is generally very good (e.g., with normal parameters, delay has little effect, see Table 4). Second, those individuals whose lesions include both reduced weights and increased decay are not particularly worse or better after a 5 s response delay, even if their overall level of accuracy is reduced. Table 8 shows three examples of such cases, EC25, HI28, and KM38. The slow SP-decay model fit the data pattern with a lesion in connection weights as well as decay rate. Importantly, the predicted error patterns were unaffected by whether 1 or 5 s of model time had passed. Figure 2 summarizes the modeling of all nine cases. Fits with reduced decay rates simulated a loss in accuracy after 5 s, while fits with reduced connection weights simulated a gain in accuracy. Fits with both lesion types (mixed lesions) simulated three example cases with little change in accuracy as a function of time.

## DISCUSSION

In this study, we aimed to provide evidence for word retrieval impairments that arise from impaired activation transmission

**TABLE 7 |** Modeling the pattern of worse naming after a response delay.

Participant/ model	Model parameters	Response delay/ time steps	N	Response types						RMSD
				Correct	Semantic	Formal	Mixed	Unrelated	Non-word	
SL21		1-s delay	n = 89	0.90	0.02	0.00	0.01	0.00	0.06	
Model	s <sup>1</sup> = 0.000295	8 time steps		0.91	0.02	0.02	0.01	0.00	0.05	0.010
SL21	p <sup>2</sup> = 0.00029	5-s delay	n = 85	0.78	0.01	0.00	0.03	0.00	0.11	
Model	DR <sup>3</sup> = 0.055	25 time steps		0.77	0.06	0.04	0.02	0.01	0.12	0.026
UN29		1-s delay	n = 90	0.10	0.00	0.07	0.00	0.06	0.77	
Model	s = 0.0003	8 time steps		0.10	0.06	0.11	0.01	0.10	0.63	0.067
UN29	p = 0.0003	5-s delay	n = 88	0.02	0.03	0.12	0.00	0.22	0.60	
Model	DR = 0.35	25 time steps		0.04	0.04	0.08	0.01	0.08	0.75	0.084
DS68		1-s delay	n = 90	0.77	0.08	0.02	0.00	0.00	0.13	
Model	s = 0.00024	8 time steps		0.82	0.04	0.03	0.01	0.01	0.09	0.032
DS68	p = 0.00022	5-s delay	n = 90	0.53	0.18	0.00	0.00	0.00	0.29	
Model	DR = 0.085	25 time steps		0.45	0.08	0.09	0.02	0.05	0.31	0.070

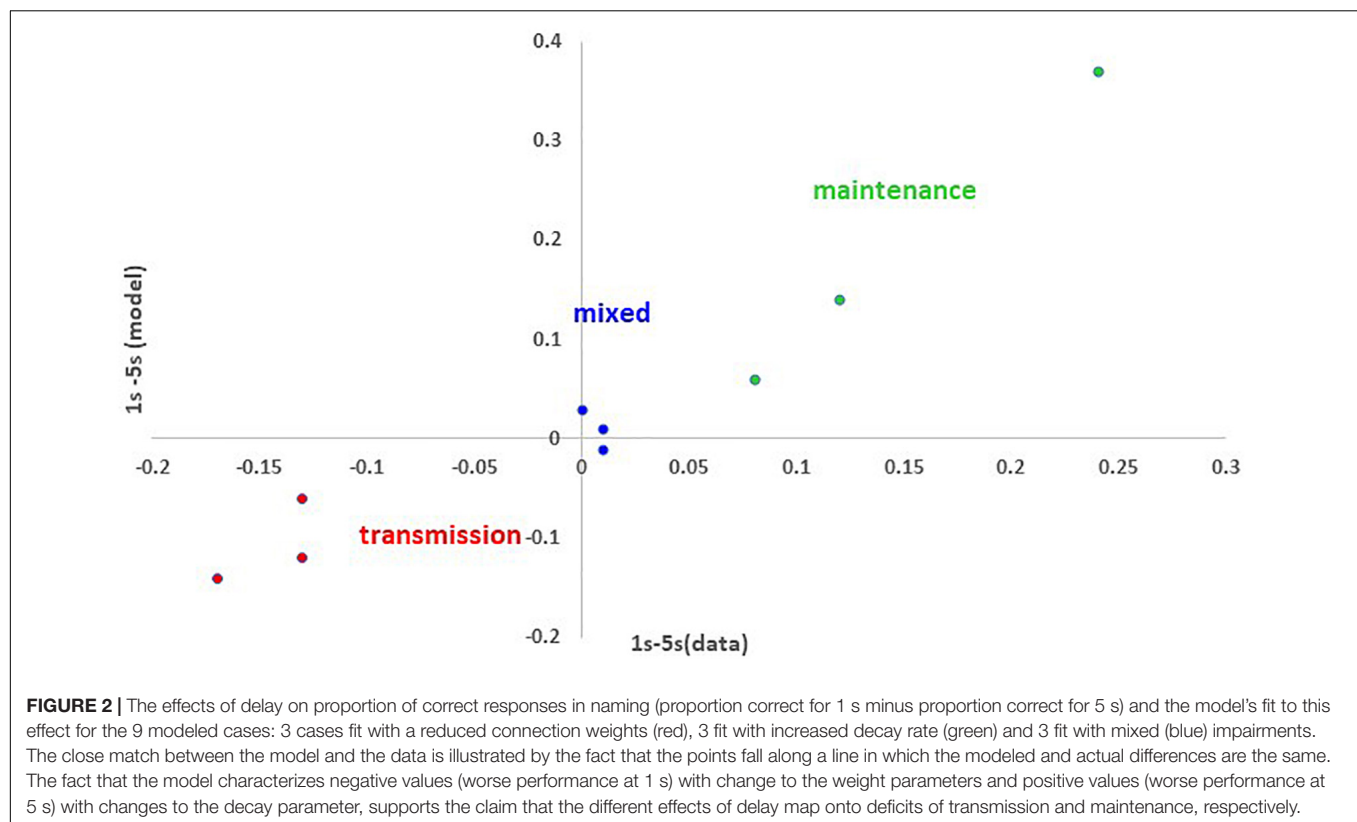
<sup>1</sup>s = semantic weight parameter. <sup>2</sup>p = phonological weight parameter. <sup>3</sup>DR = decay rate parameter.



**TABLE 8 |** Modeling the pattern of no change in accuracy after a response delay.

Participant/ model	Model parameters	Response delay/ time steps	N	Response types						RMSD
				Correct	Semantic	Formal	Mixed	Unrelated	Non-word	
EC25		1-s delay	<i>n</i> = 90	0.68	0.11	0.03	0.03	0.00	0.14	
Model	$s^1 = 0.000145$	8 time steps		0.70	0.05	0.06	0.02	0.02	0.15	0.030
EC25	$p^2 = 0.00016$	5-s delay	<i>n</i> = 90	0.68	0.19	0.00	0.00	0.00	0.13	
Model	$DR^3 = 0.04$	25 time steps		0.67	0.08	0.06	0.02	0.03	0.15	0.055
HI28		1-s delay	<i>n</i> = 90	0.81	0.01	0.01	0.00	0.00	0.17	
Model	$s = 0.00039$	8 time steps		0.77	0.01	0.05	0.01	0	0.17	0.023
HI28	$p = 0.00015$	5-s delay	<i>n</i> = 90	0.80	0.01	0.01	0.00	0.00	0.18	
Model	$DR = 0.03$	25 time steps		0.78	0.04	0.04	0.01	0	0.13	0.028
KM38		1-s delay	<i>n</i> = 82	0.71	0.12	0.00	0.00	0.00	0.16	
Model	$s = 0.00013$	8 time steps		0.72	0.07	0.06	0.02	0.03	0.11	0.040
KM38	$p = 0.00018$	5-s delay	<i>n</i> = 84	0.70	0.18	0.02	0.02	0.02	0.06	
Model	$DR = 0.035$	25 time steps		0.71	0.08	0.05	0.02	0.03	0.11	0.048

<sup>1</sup>*s* = semantic weight parameter. <sup>2</sup>*p* = phonological weight parameter. <sup>3</sup>*DR* = decay rate parameter.



and/or activation maintenance. This aim is motivated by a model of word processing (Dell et al., 1997) that postulates activation parameters of connection strength and decay rate, that regulate the retrieval and short-term maintenance of lexical-semantic and phonological representations of words. Each parameter affects the success of word retrieval in a different way. Impaired connection weights slow down activation transmission between

semantic, lexical, and phonological levels. The *s* and *p* connection weights differentially impact semantic-lexical transmission and lexical-phonological transmission, respectively. Impaired decay rate leads to excessive loss of activation by all units at all levels. Another way to think about it is that activation transmission, regulated by the *s* and *p* connection weight parameters, reflects how activated units change the activations of other units, while

the decay rate parameter determines how a unit's activation changes regardless of its inputs from other units.

We used a picture naming task with two different response delays to identify differences in transmission and maintenance abilities by persons with aphasia, and we sought to characterize such differences with the model. The model was able to account for the three patterns of change in naming accuracy as a function of delay that we observed: increased accuracy over a delay, decreased accuracy over a delay and little change in accuracy. Improvement in naming accuracy after a delay was modeled with a reduction of semantic and phonological weights, while keeping the decay rate parameter close to a level that simulates accurate word retrieval. To account for the naming pattern of decreased accuracy after a 5 s response delay, it was necessary to make decay rate a lesionable parameter separate from the connection weight parameter. In early simulations of naming and repetition in aphasia (Martin et al., 1996; Dell et al., 1997), decay rate and connection weight were lesioned globally (i.e., throughout the semantic-lexical-phonological network). More recent computational accounts of aphasia using this model lesioned only connection weights, but separately for lexical-semantic connections and lexical-phonological connections. The identification of individuals whose naming accuracy declined following a response delay, necessitated modifying the SP model to allow lesioning of decay rate. In this way, the slow SP model is a more complex model (more lesionable parameters; 3 instead of 2) than the earlier models. But it is also accounting for twice as much data (changes in error patterns as a function of delay) as the original models, thus more than making up for its additional complexity.

Finally, it is important to consider that many individuals in this study were not affected very much by the delay. These can be fit by the model with mixed lesions, that is, with lesions affecting both decay and connection strengths. If it is assumed that within the population of persons with aphasia, the parameters are largely independent random variables, one would expect that most individuals would be in this mixed category. We know from the large modeling study of Dell et al. (2013) that the *s* and *p* parameters are completely independent in a group of 103 persons with aphasia. If the same is true for decay with respect to the other parameters, then the relative uncommonness of the "pure" transmission and maintenance deficits that we found is expected.

The original SP model can also simulate word and non-word repetition (e.g., Dell et al., 2007; Nozari et al., 2010). The model assumes that words are repeated by activating a representation of the input and transmitting this activation directly to output phonology (non-lexical route) and indirectly to output phonology via lexical nodes (lexical route). Some patients use both routes while others appear to only use the lexical route (see Nozari and Dell, 2013). We could approach the simulation of word repetition after 1 or 5 s delays in the same way that we have done for naming, that is, by allowing time to pass in the model. For example, participant DS68's naming was characterized by slow SP-decay model in terms of a decay lesion (Table 7). Using DS68's parameters derived from naming, we can predict the participant's word repetition (assuming repetition by just the lexical route) by transmitting activation to the lexical nodes and

having that activation spread to the phonology, subject to the altered decay rate. And, using the slow model's ability to simulate time, we can predict how repetition will be affected by delay. Specifically, DS68's word repetition is predicted to be 94% correct at a 1 s delay and 65% correct at a 5 s delay. We mention this case because we actually have some data from DS68's on a word repetition subtest from the TALSA battery ( $n = 45$  items). DS68's performance on this test, which assesses repetition after a 1 and 5 s response delay was quite similar to the model's predicted performance: 87% correct after a 1 s delay and 58% after a 5 s delay. Thus, the assumed decay impairment derived from the naming data was mirrored in repetition and accurately modeled. Although this is just one case, it exemplifies predictions about repetition that can be made and tested. We stress, though, that success in applying the model to repetition, and more generally to the many phenomena that the original model was applied to over the years, is uncertain. The slow version of the model is not the "same" model as the original version. At this point, we are only confident that the model can explain the naming performance changes with delay, and we are reasonably confident that variation in the slow model's *s* and *p* parameters affects lexical and non-lexical errors similarly as in the original model. For example, lower values of *p* promote non-word errors.

## CONCLUSION

When aphasia was first characterized in the 19<sup>th</sup> century, the focus was on tasks, for example, naming being impaired while repetition is not. Later in the 20<sup>th</sup> century, theorists described aphasia in terms of impairments to components of linguistic knowledge (e.g., semantics, syntax, phonology) that are necessary to perform those tasks. More recent accounts have emphasized that aphasia is primarily a processing impairment affecting access to linguistic representations rather than a loss of language knowledge (e.g., McNeil, 1982; McNeil and Pratt, 2001). The most recent perspective has sought to characterize the nature of those processing impairments. The goal is not just to say what representations are impaired, but the nature of the impairment. This motivates an emphasis on cognitive abilities such as short-term memory (Saffran, 1990; Martin et al., 1994; Martin and Saffran, 1997), working memory (Wright and Shisler, 2005; Wright and Fergadiotis, 2012; Majerus, 2018), attention (Tseng et al., 1993; Murray et al., 1998; Hula and McNeil, 2008; Martin and Allen, 2012) and executive functions (Miyake et al., 2000; Martin and Allen, 2008; Allen et al., 2012). Thus, a theory of aphasia is evolving to encompass both representational and processing components.

As the theoretical models of aphasia include both linguistic and cognitive aspects of language function, it is anticipated that our approaches to rehabilitation of aphasia will follow suit. For example, current assessments of aphasia are able to identify the linguistic stages of word retrieval (semantic and/or phonological) that are impaired, guiding the focus of treatments to one linguistic stage or another (e.g., semantic feature analysis, Boyle and Coelho, 1995; phonological components treatment, Leonard et al., 2008). Our research builds on evidence for two

cognitive processes that support word retrieval, transmission and maintenance of activation, demonstrating that impairment to each has differential effects on the time course of word retrieval. As these contrasting impairments become better understood, treatments for anomia potentially will incorporate methods for their remediation. In fact, some treatments are beginning to consider the temporal aspects of interventions (e.g., Kalinyak-Fliszar et al., 2011; Conroy et al., 2018). Although we do not have a theory of how the deficits that we have investigated should be treated, we suggest that varying the temporal demands of responding when pictures and words are trained may be a useful tool, one that may work differently for individuals with maintenance and transmission impairments.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board, Temple University, Philadelphia, PA, United States. The patients/participants provided their written informed consent to participate in this study.

## REFERENCES

- Allen, C., Martin, R. C., and Martin, N. (2012). Relations between short-term memory deficits. *Semant. Proc. Exec. Function Aphasiol.* 26, 428–461. doi: 10.1080/02687038.2011.617436
- Berndt, R. S., and Mitchum, C. C. (1990). "Auditory and lexical information sources in immediate recall: evidence from a patient with deficit to the phonological short-term store," in *Neuropsychological Impairments of Short-Term Memory*, eds G. Vallar, and T. Shallice, (Cambridge: Cambridge University Press), 115–144. doi: 10.1017/cbo9780511665547.008
- Boyle, M., and Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *Am. J. Speech Lang. Pathol.* 4, 94–98. doi: 10.1044/1058-0360.0404.94
- Conroy, P., Drosopoulou, C. S., Humphreys, G. F., Halai, A. D., and Lambon Ralph, M. A. (2018). Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia. *Brain* 141, 1815–1827. doi: 10.1093/brain/awy087
- Dell, G. S., Lawler, E. N., Harris, H. D., and Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cogn. Neuropsychol.* 21, 125–145. doi: 10.1080/02643290342000320
- Dell, G. S., Martin, N., and Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: predicting word repetition from picture naming. *J. Mem. Lang.* 56, 490–520. doi: 10.1016/j.jml.2006.05.007
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., and Gagnon, D. A. (1997). Lexical access in aphasic and non-aphasic speakers. *Psychol. Rev.* 104, 801–838. doi: 10.1037/0033-295x.104.4.801
- Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., and Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: identifying the neural correlates of a computational model of word production. *Cognition* 128, 380–396. doi: 10.1016/j.cognition.2013.05.007
- Foygel, D., and Dell, G. S. (2000). Models of impaired lexical access in speech production. *J. Mem. Lang.* 43, 182–216. doi: 10.1006/jmla.2000.2716

## AUTHOR CONTRIBUTIONS

Both authors contributed equally to the content of this research. NM provided data and expertise in the assessment, evaluation and interpretation of performances by people with and without aphasia who participated in the behavioral studies. GD directed the computational modeling study and the development of the SP model used in these studies. GD and NM worked together on the model fits to the naming data. Both authors contributed to the interpretation of the results.

## FUNDING

Research reported in this publication was supported by the *National Institute on Deafness and other Communication Disorders Center* of the National Institutes of Health under award numbers R01DC013196 and R01DC016094. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## ACKNOWLEDGMENTS

We are very grateful to the participants who participated in this study. Special thanks go to Kevin McCaffery, Mary Glazer, Julie Schlesinger, Jessica Obermeyer, and Siena Sun, who helped with collection and organization of the data for this study.

- Hanley, J. R., Dell, G. S., Kay, J., and Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: a comparison of single and dual route models of auditory repetition. *Cogn. Neuropsychol.* 21, 147–158. doi: 10.1080/02643290342000339
- Hula, W., and McNeil, M. (2008). Models of attention and dual-task performance as explanatory constructs in aphasia. *Semin. Speech Lang.* 29, 169–187. doi: 10.1055/s-0028-1082882
- Kalinyak-Fliszar, M., Kohen, F., and Martin, N. (2011). Remediation of language processing in aphasia: improving activation and maintenance of linguistic representations in (verbal) short-term memory. *Aphasiology* 25, 1095–1113.
- Kertesz, A. (2006). *Western Aphasia Battery-Revised (WAB-R)*. Austin, TX: Pro-ed.
- Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310
- Leonard, C., Rochon, E., and Laird, L. (2008). Treating naming impairments in aphasia: findings from a phonological components analysis treatment. *Aphasiology* 22, 923–947. doi: 10.1080/02687030701831474
- Levelt, W. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75.
- Majerus, S. (2018). Working memory treatment in aphasia: a theoretical and quantitative review. *J. Neurolinguistics* 48, 157–175. doi: 10.1016/j.jneuroling.2017.12.001
- Martin, N., and Ayala, J. (2004). Measurements of auditory-verbal STM in aphasia: effects of task, item and word processing impairment. *Brain Lang.* 89, 464–483. doi: 10.1016/j.bandl.2003.12.004
- Martin, N., and Dell, G. (2017). Adding temporal dynamics to models of impaired language production. *Conference Abstract: Academy of Aphasia 55th Annual Meeting*. doi: 10.3389/conf.fnhum.2017.223.00110
- Martin, N., Dell, G. S., Saffran, E. M., and Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: testing the consequences of a decay impairment in an interactive spreading activation model of language. *Brain Lang.* 47, 609–660. doi: 10.1006/brln.1994.1061

- Martin, N., and Gupta, P. (2004). Exploring the relationship between word processing and verbal STM: evidence from associations and dissociations. *Cogn. Neuropsychol.* 21, 213–228. doi: 10.1080/02643290342000447
- Martin, N., Minkina, I., Kalinyak-Fliszar, M., and Kohen, F. P. (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *J. Neurolinguistic Res.* 48, 199–225. doi: 10.1016/j.jneuroling.2018.02.006
- Martin, N., and Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: evidence for common underlying processes. *Cogn. Neuropsychol.* 14, 641–682. doi: 10.1080/026432997381402
- Martin, N., Saffran, E. M., and Dell, G. S. (1996). Recovery in deep dysphasia: evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain Lang.* 52, 83–113. doi: 10.1006/brln.1996.0005
- Martin, R. C. (2005). Components of short-term memory and their relation to language processing: evidence from neuropsychology and neuroimaging. *Curr. Direct. Psychol. Sci.* 14, 204–208. doi: 10.1111/j.0963-7214.2005.00365.x
- Martin, R. C., and Allen, C. M. (2008). A disorder of executive function and its role in language processing. *Semin. Speech Lang.* 29, 201–210. doi: 10.1055/s-0028-1082884
- Martin, R. C., and Freedman, M. (2001). Short-term retention of lexical-semantic representations: implications for speech production. *Memory* 9, 261–280. doi: 10.1080/09658210143000173
- McNeil, M. R. (1982). “The nature of aphasia in adults,” in *Speech, Language, and Hearing: Vol. III. Pathologies of Speech and Language*, eds N. J. Lass, L. V. McReynolds, Northern, and D. E. Yoder, (Philadelphia, PA: W.B. Saunders), 692–740.
- McNeil, M. R., and Pratt, S. R. (2001). Defining aphasia: some theoretical and clinical implications of operating from a formal definition. *Aphasiology* 15, 901–911. doi: 10.1080/02687040143000276
- Miyake, A., Emerson, M. J., and Friedman, N. P. (2000). Assessment of executive functions in clinical settings: problems and recommendations. *Semin. Speech Lang.* 21, 169–185.
- Murray, L., Holland, A. L., and Beeson, P. M. (1998). Spoken language of individuals with mild fluent aphasia under focused and divided-attention conditions. *J. Speech Lang. Hear. Res.* 41, 213–227. doi: 10.1044/jslhr.4101.213
- Nozari, N., and Dell, G. S. (2013). How damaged brains repeat words: a computational approach. *Brain Lang.* 126, 327–337. doi: 10.1016/j.bandl.2013.07.005
- Nozari, N., Kittredge, A. K., Dell, G. S., and Schwartz, M. F. (2010). Naming and repetition in aphasia: steps, routes, and frequency effects. *J. Mem. Lang.* 63, 541–559. doi: 10.1016/j.jml.2010.08.001
- Oppenheim, G. M., Dell, G. S., and Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition* 114, 227–252. doi: 10.1016/j.cognition.2009.09.007
- Pastizzo, M. J., and Carbone, R. F. Jr. (2007). Spoken word frequency counts based on 1.6 million words in American English. *Behav. Res. Methods* 39, 1025–1028. doi: 10.3758/bf03193000
- Psychology Software Tools Inc. (2012). [E-Prime 2.0]. Available at: pstnet.com
- Rapp, B., and Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychol. Rev.* 107, 460–499. doi: 10.1037//0033-295x.107.3.460
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., and Brecher, A. (1996). The philadelphia naming test: scoring and rationale. *Clin. Aphasiol.* 24, 121–133.
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: the WEAVER ++/ARC model. *Cortex* 59, 33–48. doi: 10.1016/j.cortex.2014.07.001
- Saffran, E. M. (1990). “Short-term memory impairment and language processing,” in *Advances in Cognitive Neuropsychology and Neurolinguistics*, ed A. Caramazza (Hillsdale: Erlbaum).
- Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., and Sobel, P. (2006). A case series test of the two-step interactive model of lexical access: evidence from picture naming. *J. Mem. Lang.* 54, 228–264. doi: 10.1016/j.jml.2005.10.001
- Shelton, J., Martin, R. C., and Yaffee, L. (1992). “Investigating a verbal short-term memory deficit and its consequences for language processing,” in *Cognitive Neuropsychology in Clinical Practice*, ed D. Margolin, (New York, NY: Cambridge University Press).
- Tseng, C.-H., McNeil, M., and Milenkovic, P. (1993). An investigation of attention allocation deficits in aphasia. *Brain Lang.* 45, 276–296. doi: 10.1006/brln.1993.1046
- Ueno, T., Saito, S., Rogers, T. T., and Lambon Ralph, M. A. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* 72, 385–396. doi: 10.1016/j.neuron.2011.09.013
- Walker, G. M., and Hickok, G. (2016). Bridging computational approaches to speech production: the semantic-lexical-auditory-motor model (SLAM). *Psychon. Bull. Rev.* 23, 229–352.
- Wright, H. H., and Fergadiotis, G. (2012). Conceptualizing and measuring working memory and its relationship to aphasia. *Aphasiology* 26, 258–278. doi: 10.1080/02687038.2011.604304
- Wright, H. H., and Shisler, R. J. (2005). Working memory in aphasia: theory, measures and clinical implications. *Am. J. Speech Lang. Pathol.* 14, 107–118. doi: 10.1044/1058-0360(2005/012)

**Conflict of Interest:** NM and GD confirm that this submitted work was carried out without any personal, professional, or financial relationships present that could potentially be construed as a conflict of interest.

Copyright © 2019 Martin and Dell. 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.





# Distinct Neural Processes for Memorizing Form and Meaning Within Sentences

Matteo Mascelloni<sup>1,2,3\*</sup>, Roberto Zamparelli<sup>4</sup>, Francesco Vespignani<sup>5</sup>, Thomas Gruber<sup>6</sup> and Jutta L. Mueller<sup>1\*</sup>

<sup>1</sup> Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany, <sup>2</sup> School of Psychology and Counselling, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia, <sup>3</sup> Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia, <sup>4</sup> Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy, <sup>5</sup> DIPSCO, University of Trento, Trento, Italy, <sup>6</sup> Institute of Psychology, University of Osnabrück, Osnabrück, Germany

## OPEN ACCESS

### Edited by:

Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### Reviewed by:

Aine Ito,  
Humboldt University of Berlin,  
Germany  
Mante Sjouke Nieuwland,  
Max Planck Institute  
for Psycholinguistics, Netherlands

### \*Correspondence:

Matteo Mascelloni  
matteomascelloni@gmail.com  
Jutta L. Mueller  
jutta.mueller@uos.de

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 10 August 2019

**Accepted:** 07 November 2019

**Published:** 05 December 2019

### Citation:

Mascelloni M, Zamparelli R,  
Vespignani F, Gruber T and Mueller JL  
(2019) Distinct Neural Processes  
for Memorizing Form and Meaning  
Within Sentences.  
Front. Hum. Neurosci. 13:412.  
doi: 10.3389/fnhum.2019.00412

In order to memorize sentences we use both processes of language comprehension during encoding and processes of language production during maintenance. While the former processes are easily testable via controlled presentation of the input, the latter are more difficult to assess directly as language production is typically initiated and controlled internally. In the present event-related potential (ERP) study we track subvocal rehearsal of sentences, with the goal of studying the concomitant planning processes with the help of a silent cued-production task. Native German participants read different types of sentences word-by-word, then were prompted by a visual cue to silently repeat each individual word, in a *rehearsal phase*. In order to assess both local and global effects of sentence planning, we presented correct sentences, syntactically or semantically violated sentences, or random word order sequences. Semantic violations during reading elicited an N400 effect at the noun violating the selectional restrictions of the preceding verb. Syntactic violations, induced by a gender incongruity between determiner and noun, led to a P600 effect at the same position. Different ERP patterns occurred during the silent production phase. Here, semantically violated sentences elicited an early fronto-central negativity at the verb, while syntactically violated sentences elicited a late right-frontal positivity at the determiner. Random word order was accompanied by long-lasting slow waves during the production phase. The findings are consistent with models of hierarchical sentence planning and further indicate that the ongoing working memory processes are qualitatively distinct from comprehension mechanisms and neurophysiologically specific for syntactic and lexical-semantic level planning. In conclusion, active working memory maintenance of sentences is likely to comprise specific stages of sentence production that are indicated by ERP correlates of syntactic and semantic planning at the phrasal and clausal level respectively.

**Keywords:** sentence repetition, language production, working memory, syntax, semantics, ERP, slow wave, mental rehearsal

## INTRODUCTION

Having language at our disposal serves multiple purposes. While it undisputedly works as a means of communication between individuals, it also serves as a code for cognition within the individual (Pinker and Bloom, 1990; Bolhuis et al., 2014; Asoulin, 2016). A cognitive function which uses language in complex ways is working memory, which is conceived as the cognitive function supporting “the few temporarily active thoughts” (Cowan, 2010, p. 51). In many cases, our thoughts are verbal in nature, which is why models of working memory include some kind of language-based processes. An important mechanism included in many current models of verbal working memory is subvocal rehearsal as a mechanism of maintaining arbitrary verbal material (Baddeley and Hitch, 1974; Cowan, 1999; Camos et al., 2009). In addition to phonological information, as instantiated in subvocal rehearsal, higher order linguistic representations such as semantic and syntactic information have also been shown to contribute to working memory processing. A prime example involving both phonological/articulatory and higher order linguistic processes is working memory for sentences. Typically, the process of memorizing sentences is easy, even when verbatim recall is required (e.g., dictation). The present study aims to contribute to the understanding of how well-formed sentences are retained more efficiently and more accurately than unstructured lists of words.

It has long been known that our memory for words embedded in sentences far exceeds the typical short-term memory span of  $\pm 7$  items (Brener, 1940; Miller and Selfridge, 1950; Marks and Miller, 1964). Readers experience difficulties when trying to correctly repeat a random word sequence such as “out prince swamp the of white the of are carriage horses pulling dangerous the” after reading it. The same words can easily be repeated, however, when they are organized within a sentence: “white horses are pulling the carriage of the prince out of the dangerous swamp.” This so-called “sentence superiority effect” is a very robust observation and measurable even in two-word lists (Perham et al., 2009) and meaningless “jabberwocky” sentences (Marks and Miller, 1964; Bonhage et al., 2014). It can be measured with different tasks, including recall (Baddeley et al., 2009; Allen et al., 2018) and recognition (Bonhage et al., 2014; Allen et al., 2018). Many studies have demonstrated that the sentence-superiority effect is due to rapid access to stored linguistic knowledge and conceptual/semantic processes which improve the way our memory encodes, maintains and retrieves meaningful sentences (Potter and Lombardi, 1990; Lombardi and Potter, 1992; Jefferies et al., 2004; Baddeley et al., 2009; Perham et al., 2009; Schweppe et al., 2011; Bonhage et al., 2014). Yet, how exactly subvocal rehearsal benefits from, or interacts with higher order linguistic information has not been investigated at a fine-grained level. On the one hand, there is evidence that subvocal rehearsal is dispensable if higher order linguistic information is available (e.g., syntactic and/or semantic relations), as suggested by the robustness of the sentence superiority effect in the face of articulatory suppression (Baddeley et al., 2009; Bonhage et al., 2014). On the other hand, studies comparing memory for different types of sentences

provide evidence that subvocal rehearsal plays a role even for remembering well-formed sentences (Meltzer et al., 2016, 2017).

The present study tests how higher order linguistic information and subvocal rehearsal interact. This is done by investigating specifically how syntactic and semantic information contribute to sentence memory and, in particular, to subvocal rehearsal as a working memory maintenance mechanism. As this research question involves processes discussed in the working memory as well as in the language production literature, we will review studies from both fields with a specific focus on neurophysiological processes at the sentential level, i.e., working memory for sentences and sentence production.

Subvocal rehearsal is a well-investigated, yet not uncontroversial mechanism for the short-term memorization of verbal and verbalizable material, and is part of multi-component as well as process models of working memory (Cowan, 1999; Baddeley, 2003; but see Lewandowsky and Oberauer, 2015). There is ample evidence for the psychological reality of subvocal articulation processes during memory processing, even though the efficiency of such processes has been questioned (Souza and Oberauer, 2018). Early conceptualizations of short-term memory already included subvocal rehearsal as a mechanism of maintaining verbal information (Waugh and Norman, 1965; Sperling, 1967; Atkinson and Shiffrin, 1968). Evidence for such a mechanism stems, for example, from the observation that concurrent articulation of irrelevant speech (e.g., the articulation of “ne na da na ne na...”) interferes with the maintenance of verbal material (Murray, 1967). Sequences consisting of longer words are more difficult to remember than sequences consisting of shorter words (Baddeley et al., 1975, 1984). This can be explained by the additional time needed to pronounce longer words, an explanation which is supported by the observation that articulatory suppression eliminates this effect (Baddeley et al., 1984). Further, participants in memory tasks frequently report using subvocal articulation strategically (Dunlosky and Kane, 2007; Morrison et al., 2016). Instructions to rehearse word lists aloud seem to improve performance especially in participants with low working memory spans (Turley-Ames, 2003). Thus, while the role and importance of subvocal rehearsal remain debated, it clearly plays a role in short-term maintenance of arbitrary verbal information. Lastly, brain areas that are involved during overt language production, such as premotor cortex (BA6) and parts of the inferior frontal cortex (BA44) (Indefrey and Levelt, 2004; Saur et al., 2008) also play a role during the maintenance phase in verbal working memory (Chein et al., 2003; Buchsbaum and D’Esposito, 2008; Bonhage et al., 2014). For this reason, current psychological and neurocognitive approaches to working memory posit the involvement of the same cognitive and sensorimotor processes related to language production in verbal working memory tasks (Buchsbaum and D’Esposito, 2008, 2019; Acheson and MacDonald, 2009; Majerus, 2013).

While it can be reasonably assumed that processes of language production are involved in the memory retention of arbitrary verbal information, they become less important when higher-order linguistic information, such as syntactic and semantic structures within the sentence, come into play. Early studies have already proposed that memory advantages for sentences

may be due to the processing of syntactic and semantic dependencies between items (Miller and Selfridge, 1950; Marks and Miller, 1964). Later studies showed that successful sentence maintenance involves neither extensive rehearsal nor attention-demanding processes, but rather relies on long-term memory representations and automatic language processing mechanisms (Jefferies et al., 2004; Baddeley et al., 2009). Accordingly, an explicit model of immediate sentence recall, the conceptual regeneration hypothesis, assumes that conceptual-semantic, but not phonological/articulatory processes are mainly involved in immediate sentence memory (Potter and Lombardi, 1990; Lombardi and Potter, 1992). The original hypothesis was based on the observation of a specific type of error during sentence recall: Synonyms of words in sentences that were presented next to the sentences were often reproduced in replacement of the correct word in the sentence (Potter and Lombardi, 1990). Later studies, however, provided a more multifaceted picture by showing in similar experimental designs that phonological and syntactic information can also interfere with sentence memory (Rummer and Schweppe, 2005; Schweppe and Rummer, 2007; Schweppe et al., 2011). Thus, it seems highly likely that linguistic codes at all levels, from articulatory to conceptual, play a certain role in immediate sentence memory.

In agreement with the findings from behavioral studies, a recent fMRI study demonstrated that working memory for sentences, compared to unstructured word sequences, involves a widely distributed network of brain areas related to semantic processing during encoding, and decreased activation of subvocal rehearsal-related areas during maintenance (Bonhage et al., 2014). This and other studies suggest that the working memory benefits during maintenance (consisting of a smaller amount of rehearsal-related activity and performance increase) may be contingent on enhanced processing costs during the encoding phase (Bor et al., 2003, 2004; Bonhage et al., 2014). In an EEG study on the memorization of sentences vs. unstructured word sequences, sentence maintenance was accompanied by reduced oscillatory power in the theta, alpha, and beta bands (Bonhage et al., 2017), frequencies which have all been related to working memory load (Jensen and Tesche, 2002), and in the case of theta oscillations, to the application of rehearsal strategies (Meltzer et al., 2017). Other electrophysiological studies have used event-related potentials (ERP) to investigate the retention of verbal material either in working memory tasks or in sentence processing tasks. Studies using working memory tasks have reported long-lasting frontal negativities for the costs of retention of verbal compared to non-verbal material (Lang et al., 1992; Ruchkin et al., 1992). While some authors have related the frontal slow waves directly to phonological rehearsal processes (Lang et al., 1992; Ruchkin et al., 1992), others, who reported similar slow waves for non-verbalizable conditions, have interpreted it as being related to attentional control of working memory contents (Bosch et al., 2001; Murphy et al., 2006). Studies assessing working memory costs during sentence processing have reported similar frontal negative shifts for sentences or sentence parts which were hypothesized to impose increased working memory processing loads (King and Kutas, 1995; Fiebach et al., 2002).

Together, neurophysiological studies on verbal working memory show that the availability of higher order linguistic information can reduce general brain activation related to subvocal rehearsal during the maintenance phase. In these studies, rehearsal is treated as a uniform function that can occur to a higher or lower degree, depending on the type of material and memory strategy. In fMRI studies, the presence of rehearsal is typically identified based on the involvement of brain regions that are usually correlated with articulation, specifically posterior inferior frontal and premotor areas (Bor et al., 2004; Bonhage et al., 2014). In EEG studies, rehearsal has been inferred from the presence of specific concurrent increased slow-wave amplitudes (Lang et al., 1992; Ruchkin et al., 1992) or certain oscillatory patterns (Griesmayr et al., 2010; Bonhage et al., 2017) in response to the processing of rehearsed material. Yet, the nature and sequence of the preparatory and execution processes during rehearsal has not been brought to light by these neurophysiological studies. A more fine-grained analysis of the processes involved in language production, if they occur, is still largely amiss. We suggest that models of sentence production are highly informative about how such processes are likely to be applied.

On-line language production has proven more difficult to investigate than language comprehension, specifically at the sentential level. This is due to the internal nature of the different stages of planning and execution in language production, which are only indirectly accessible. In general, psycholinguistic models of sentence production postulate that (i) there is a certain degree of planning ahead in sentence production and that (ii) there are separable planning stages, e.g., at the conceptual level, at the level of abstract lexical forms and at the level of concrete phonological forms (Garrett, 1975, 1982; Smedt and Kempen, 1987; Levelt, 1994). Tasks used in studies on sentence planning have to include some type of concrete instructions, often picture-based, specifying which sentence is to be produced. Many studies use different kinds of distractor items (Meyer, 1996; Wagner et al., 2010; Bürki et al., 2016; Klaus et al., 2017) or complexity manipulations (Ferreira, 1991; Smith and Wheeldon, 1999) in order to interfere with specific stages of sentence production. Longer or shorter onset latencies for production are then interpreted as reflecting either increased processing costs or facilitation during the planning stage, stemming from the corresponding manipulation.

The extent and flexibility of the planning scope, that is, how much planning ahead occurs at each stage of language production, is controversial (Martin et al., 2010; Klaus et al., 2017). The influential frame-and-slot model proposed by Garrett (1975, 1982) assumes a larger scope for abstract lexical planning compared to phonological planning, as evidenced by speech errors in the respective domains. Indeed, several studies suggest at least a phrasal scope of planning at the abstract lexical level (Smith and Wheeldon, 1999; Martin et al., 2010; Lee et al., 2013; Klaus et al., 2017). Studies testing phonological encoding during sentence planning also reported evidence for a phrasal scope of planning (Oppermann et al., 2010; Schnur, 2011), but also for a much smaller planning scope (Meyer, 1996; Wheeldon and Lahiri, 1997). One reason for such variable findings may be a

certain degree of flexibility in the planning scope. Both sentence-related factors, such as sentence complexity (Ferreira, 1991; Smith and Wheeldon, 1999; Wagner et al., 2010) and non-sentence related factors, such as concurrent cognitive load (Boiteau et al., 2014; Klaus et al., 2017) seem to impact on how far in advance lexical-semantic and phonological word forms are planned. As a link between the domain of working memory and language production, sentence repetition has not only been used as a task to probe verbal working memory, but also as a way to assess the processes which occur during sentence planning. Thus, Ferreira (1991) presented participants with sentences of different syntactic complexity and showed that it took longer to initiate the production of a syntactically complex sentence compared to a less complex one. This was taken as an indication of a grammatical planning stage in which utterances are planned at a phrasal scope. In sum, studies on sentence production support incremental planning at different production levels with a tendency for a larger planning scope for higher-order linguistic levels.

A few neurophysiological studies have tackled the production of linguistic units longer than the single word. Haller et al. (2005), for example, have shown a specific contribution of Broca's area for sentence generation from word triplets. Mere repetition of sentences has been shown to involve a network including the left hemispheric articulatory network (premotor cortex and parieto-temporal junction), semantic areas (left temporal lobe and inferior frontal cortex) as well as bilateral working-memory-related areas in the parietal and dorsolateral prefrontal cortex (cf. Majerus, 2013, for review).

There is a large number of EEG studies on language comprehension and a smaller number on language production. The initial ERP components that could be functionally related to language processing at the sentence level were the N400 component in response to semantic incongruities, discovered by Kutas and Hillyard (1980), and the P600 component in response to syntactic incongruities, discovered by Osterhout and Holcomb (1992). The N400 is a negative deflection in response to a stimulus with increased lexical or semantic processing demands and is typically related either to automatic processes at the stage of lexical access, or to later, more controlled semantic processes at the semantic integration stage (cf. Kutas and Federmeier, 2000, 2011; Lau et al., 2008, for reviews). The P600 component is a positivity typically found in response to syntactic manipulations, but also in the context of specific types of semantic violations, thus seen as an indicator of more global integration difficulties at the sentential level (Kuperberg, 2007; Friederici, 2011) or of internal monitoring of processing effort (van Herten et al., 2005; Sassenhagen et al., 2014). Electrophysiological studies on word and sentence production are fewer and have reported different effects (Pylkkänen et al., 2014; Bürki et al., 2016; Shitova et al., 2017; Blanco-Elorrieta et al., 2018). Bürki et al. (2016) reported differential ERP responses for gender congruency and phonological similarity of distractors during the production of simple determiner-noun phrases. The phonological similarity of the distractor and the noun was processed earlier than the gender congruency between distractor and target noun. This was interpreted as an indication of sequential phonological encoding, in which the encoding of the determiner follows the encoding

of the noun. Pylkkänen et al. (2014) and Blanco-Elorrieta et al. (2018) conducted several MEG studies on the production of simple two-word adjective-noun phrases and found effects for semantic composition about 200 ms after a production cue. The effects, which they related to the stage of lexical access during production, could be localized to the anterolateral temporal and to the ventro-medial prefrontal cortex. The complexity of the phrases produced as well as the need to switch between different phrase types has been found to increase the amplitude of the P3 component (Shitova et al., 2017). The P3 component is a positive potential starting at around 300 ms after stimulus onset, which has been related to domain-general processes of context updating and cortical reorientation (Nieuwenhuis et al., 2005; Polich, 2007). In sum, the typical ERP components reported in tasks that involve word production at the sentence level comprise both a negativity, related to lexical-semantic processes, and a positivity (P3), reflecting more general processing costs relating to production planning.

## THE PRESENT STUDY

Our goal was to investigate how semantic and syntactic information is used during repetition of sentences in a working memory task. We assumed that subvocal sentence repetition includes core processes of sentence production, specifically conceptual, abstract lexical and phonological planning stages in addition to the silent articulation processes. This is in alignment with widespread views on sentence repetition assuming that many different language skills relating to comprehension and production contribute to the correct repetition of sentences (Lombardi and Potter, 1992; cf. Acheson and MacDonald, 2009; Klem et al., 2015). To test this assumption, we measured ERPs as a response to unstructured word sequences vs. sentences as well as ERPs in response to more subtle linguistic violations, i.e., local semantic and syntactic anomalies. We presented participants with variants of German declarative sentences consisting of a subject, a verb, a direct object and an adverbial expression. Importantly, the violations in the semantic and syntactic anomaly condition both occurred in the same position, namely at the direct object noun. In Table 1, example strings are listed for each condition.

TABLE 1 | Examples of the stimulus material by condition.

Condition	Sentence
(A) Correct	Die Frau bindet den Schuh im Flur The woman ties the [M] shoe in the hallway
(B) Semantic	Die Frau <b>steuert</b> den <u>Schuh</u> im Flur The woman <b>navigates</b> the[M] shoe in the hallway
(C) Syntactic	*Die Frau bindet <b>das</b> [N] <u>Schuh</u> [M] im Flur *The woman ties <b>the</b> [N] shoe in the hallway
(D) Random word order	*Frau den die Flur bindet Schuh im *woman the hallway ties shoe [in the]

M, masculine; N, neuter. Violation position is underlined in each sentence except D.  
\*Ungrammatical sentence. Bold words refer "violation."



According to working memory-based models for sentence repetition, we assumed that the high working memory load incurred by unstructured word sequences would elicit long-lasting slow-waves reflecting increased verbal working memory loads (Ruchkin et al., 1997) or non-verbal domain-general memory maintenance strategies (Bosch et al., 2001) during subvocal rehearsal. Further, based on the conjecture outlined above that sentence repetition rests to a large degree on normal sentence production, we assumed more local and violation-type-specific processing costs for the semantically and syntactically manipulated sentences. Specifically, we expected processing costs reflecting different scopes of advance planning for lexical-semantic and syntactic information. As the selection of abstract lexical information and concrete determiner forms have been related to different planning stages (e.g., Bürki et al., 2016), we expected processing costs at an earlier position in the sentence for the semantic compared to the syntactic violation condition. Corresponding to a phrasal planning scope in the abstract lexical stage, we expected semantic processing costs time-locked to the verb onset in semantically anomalous verb phrases. For the syntactic condition, we expected difficulties at a later planning stage, the level of morphophonological encoding. This is based on the observation that determiners are planned together with or even after the corresponding nouns (cf. Bürki et al., 2016). Thus, we expected that the gender incongruity of the noun would modify the ERP time-locked to the rehearsal cue for the preceding determiner. Due to the explorative nature of the study, we did not have specific hypotheses about the polarity and distribution of the ERP components to be expected.

The initial reading phase served as a control condition to make sure that both our semantic and syntactic violations lead to specific processing difficulties at the same target point in the sentence, namely the direct object noun. At this position, we expected an N400 component for the semantic anomaly and a P600 component for the syntactic anomaly, reflecting the functional distinction between both types of processes.

## MATERIALS AND METHODS

### Participants

Twenty-six native German participants (university students) volunteered for the study. The participants (15 female and 11 male) were between 18 and 28 years old (mean age = 22.5 years;  $SD = 2.51$ ), all right-handed and native speakers of German. Each participant took part in two separate sessions. Two participants (one male and one female) had to be excluded from analysis, one due to technical issues during the measurement, and one because of a clinical diagnosis of dyslexia, which the experimenter was only informed about after the experiment. None of the remaining 24 participants reported any recent history of neurological or psychological disorders and none of them were subject to any medical treatments or under the influence of drugs or alcohol at the time of the experiment. All participants gave written informed consent in accordance with the declaration of Helsinki (World Medical Association, 2013) and received a written confirmation of participation.

In order to test for working memory performance, a Wechsler digit span test, forward and backward, was performed on all subjects. The mean forward span was 6.75 with a standard deviation of 1.25 (max. span = 9), while the mean score was 8.75 with  $SD = 1.82$  (max. score = 14); for the backward version, the mean span was 5.20 with  $SD = 1.28$  (max. span = 8) and the mean score was 7.25 with  $SD = 1.89$  (max. score = 14). Participants with a forward span superior to  $6 \pm 1$  are considered normal, for the backward span the typical range is  $5 \pm 1$  (Peña-Casanova et al., 2009). All participants in the study fell in the normal range.

### Stimuli

The stimulus material consisted of German declarative sentences composed of seven words each. All stimuli followed the same syntactic structure exemplified in **Table 1**. The experiment comprised four different conditions (three violation conditions and one control); each condition consisted of 90 items, resulting in a total of 360 stimuli. In addition, the experiment included 18 rehearsal check items which were used to ensure the participants were engaged in rehearsal. Those sentences followed the same pattern as the experimental material with 9 correct sentences and 3 sentences for each of the violations. As a control condition, grammatical sentences (as shown in **Table 1**) were used as a baseline for comparison with the other conditions (for a complete list of the stimuli, please refer to the **Supplementary Material**).

In the semantic-mismatch condition, the verb from the control was substituted by a verb that agreed in meaning with the subject (e.g., *Die Frau bindet den Schuh*. . . [*The woman ties the shoe*. . .]), but not with the object of the sentence (e.g., *Die Frau steuert \*den Schuh* [*The woman navigates \*the shoe*. . .]). For the morpho-syntactic violation condition, the article of the object was modified, creating a gender disagreement between article and noun (e.g., *\*das Schuh* [*the<sub>(neut.) shoe<sub>(masc.)</sub></sub>*]) as in e.g., Gunter et al. (2000). Since the female article *die* is the same as the plural article used for all grammatical genders in German, only masculine and neuter nouns were used in object position to avoid eliciting a response to number agreement mismatch rather than to the intended gender disagreement. For both the semantic and the syntactic mismatch condition, the violations occurred once the object noun of the sentence was encountered. The fourth and final condition comprised strings of words that were constructed by randomizing the word order of each individual sentence. This randomization was constrained in a way so that across the whole sentence, no more than two consecutive words appeared in a syntactically permissible sequence (i.e., the violation became apparent at the third word at the latest). In this condition, the first word presented was not capitalized unless it was a noun (since nouns are always capitalized in German). The previously mentioned 18 additional stimulus sentences for the rehearsal check were generated based on the same pattern of the four conditions outlined above (with a distribution of nine grammatical control sentences and three ungrammatical sentences per violation condition); however, each of these sentences were composed of new vocabulary, hence the ungrammatical sentences were not based on the grammatical control condition.

Each participant was presented with a total of 198 sentences (90 control sentences, 30 items per violation condition, plus the 18 rehearsal check sentences), which were divided equally into 99 stimuli per session. This distribution allowed for two versions of the same sentence to be shown to each participant, with a delay of at least 5 days between sessions, reducing the risk of potential repetition effects. The non-control stimuli were divided into three different lists using a Latin square design. Six sets of two lists (list A and B) were prepared. The lists were pseudo-randomized so that each condition would not appear more than three times in a row. Each subject was presented with one set (one list per session). List A and List B have been counterbalanced across subjects.

## Software and Hardware

Both stimulus presentation and behavioral data acquisition were performed with MATLAB (Version R2017a, Mathworks®, Natick, MA, United States) using the Psychophysics Toolbox extension (Brainard, 1997). A USB microphone was used to acquire the sound response while the button press data were acquired using a response pad from The Black Box ToolKit Ltd.

## EEG Recording and Electrodes

The EEG was recorded continuously using a TMSi 72 Refa amplifier and an EEG gel head cap by TMSi (TMSi B.V., Netherlands), using the 5% system with 64 channels (Oostenveld and Praamstra, 2001). EEG data were recorded with the TMSi Polybench software (TMSi B. V., Netherlands). The ground electrode was placed on the collar bone. The EOG was recorded using two bipolar electrophysiological inputs (BIP), the first one (EOGV) was positioned above and below the left eye, the second one (EOGH) was positioned close to the outer canthi of the left and the right eye. The impedance of all electrodes was kept below 5  $\Omega$ . The signals were acquired with a sample rate of 512 Hz with an online average reference.

## Experimental Procedure

The experiment was distributed across two sessions with identical experimental procedures. After the preparation process (30–40 min on average), participants were seated in a comfortable chair with a distance of 60 cm between the nasion and the screen. The instructions, as well as a block of five practice trials were presented to each participant, then they began the actual experiment.

In the standard trial sequence (Figure 1), the stimulus sentences were presented in a word-by-word manner, each word appearing on the screen for 500 ms with a blank screen being shown for 150 ms in between words. Subjects were instructed to silently read the words that appeared on the screen. After the final word of each sentence, a blank screen was presented for 500 ms. Participants were instructed to silently repeat each of the previously encountered words, precisely in the same order as they had been shown in the reading phase. After the pause, a fixation cross (+) appeared on the screen for 500 ms, followed by another blank screen (500 ms). This fixation cross/blank screen sequence was repeated seven times after each sentence, once for every previously presented word and as a cue, setting a rhythm

to the participant's retrieval process. After the repetition phase, a serial recognition task was used to probe sentence memory. In this task, a sequence of two words appeared on the screen and the participant had to press either one of two buttons evaluating whether this sequence had appeared in the previous sentence.

As a further measure to ensure participants followed the instructions and actually engaged in inner retrieval during the thinking phase, the standard trial sequence was slightly modified for the 18 rehearsal check stimuli. These stimuli were presented pseudo-randomly during the experiment (one for each block). For each of these sentences, two of the fixation crosses in the thinking phase in the fourth and fifth position were substituted by the image of a microphone (cf. Figure 1). This probe was shown for 600 ms and cued the participant to repeat the respective word out loud instead of silently, and their vocal response was recorded and stored in a sound file to be analyzed separately.

Each session was divided into nine blocks of eleven sentences each, with each block containing one attention check stimulus. After each block, the experiment was paused until the subject decided to resume. After the fifth block, participants were required to take a 5-min break. The duration of the whole experiment was approximately 35 min.

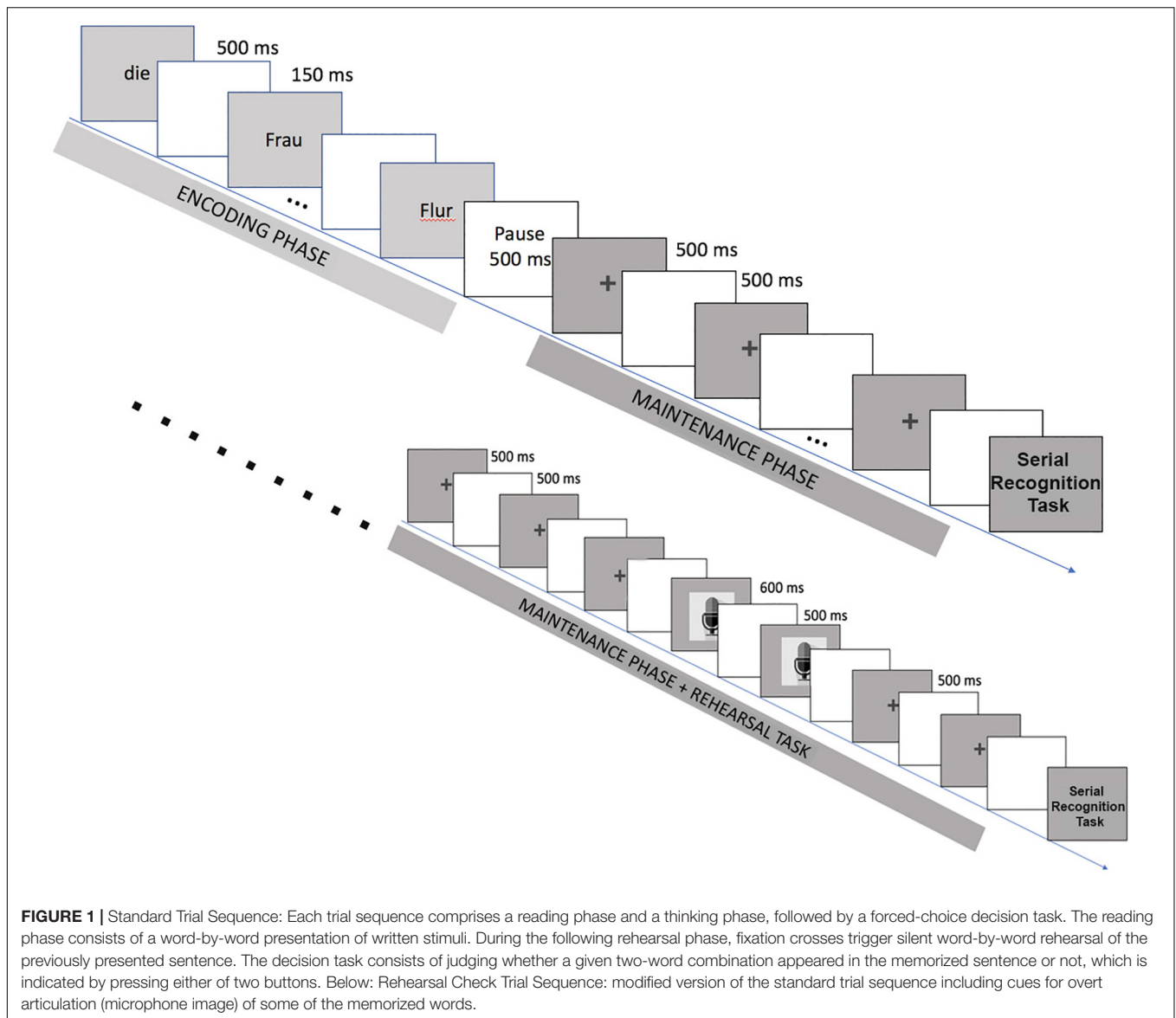
## Data Analysis

### EEG Data

The EEG data were pre-processed with MATLAB (Version R2017a, Mathworks®, Natick, MA, United States) using “EEGLAB Toolbox” version 14 (Delorme and Makeig, 2004). The data from the two sessions of each subject were merged and each file was epochized in windows of 1600 ms (600 ms before stimulus onset and 1000 ms after) in order to capture local effects (short epochs). In order to capture the expected slow waves in the rehearsal phase, epochs of 7600 ms (600 ms before stimulus onset and 7000 ms after) were selected for all conditions (long epochs).

The data were manually cleaned to remove the most evident muscular artifacts and then re-referenced to the right mastoid. FASTER analysis (Nolan et al., 2010) was then run to remove blinks and eye-movement artifacts. This included high-pass filtering before the application of an ICA during the FASTER procedure. A 0.5 Hz high-pass filter (−6 dB cut-off frequencies of 0.25 Hz) as well as a notch filter at 50 Hz (bandwidth 3 Hz) to remove line noise, were applied for the short epochs, while a 0.03 Hz high-pass filter (−6 dB cut-off frequencies of 0.015 Hz), was applied for the long epochs. All epochs were low-pass filtered at 25 Hz (−6 dB cut-off frequencies of 21.875 Hz). The data were re-referenced to averaged mastoids and resampled to 1000 Hz.

For the statistical analysis, the “Fieldtrip toolbox” was used (Oostenveld et al., 2011). ERPs of the short epochs were calculated for each condition by averaging across subjects and by applying a baseline from 0 (onset of the stimulus) to 100 ms. We chose a post-stimulus baseline for all short epochs in order to have a uniform baseline across the reading and the rehearsal phase which would not be influenced by the rehearsal of the previous word. Baseline correction for the long epochs was applied between 500 ms before the onset to 0 ms. A non-parametric cluster-based permutation analysis was applied using dependent samples *t*-tests with the threshold for alpha fixed at



0.05. The minimum number of neighbourhood channels for a defined sample to be included in the statistic was equal to 2. A permutation test based on the Monte Carlo method (Maris and Oostenveld, 2007) was used with 1000 randomizations ( $\alpha = 0.05$ ). It should be noted that the cluster-based permutation test is reliable when it comes to identifying effects in the data, but does not allow for a precise identification of latency and distribution of these effects (Sassenhagen and Draschkow, 2019). Therefore, the time-windows and distributions reported in the result section are only those of the respective clusters identified via the test and do not necessarily reflect the exact time-windows and distributions of the effect.

## Behavioral Data

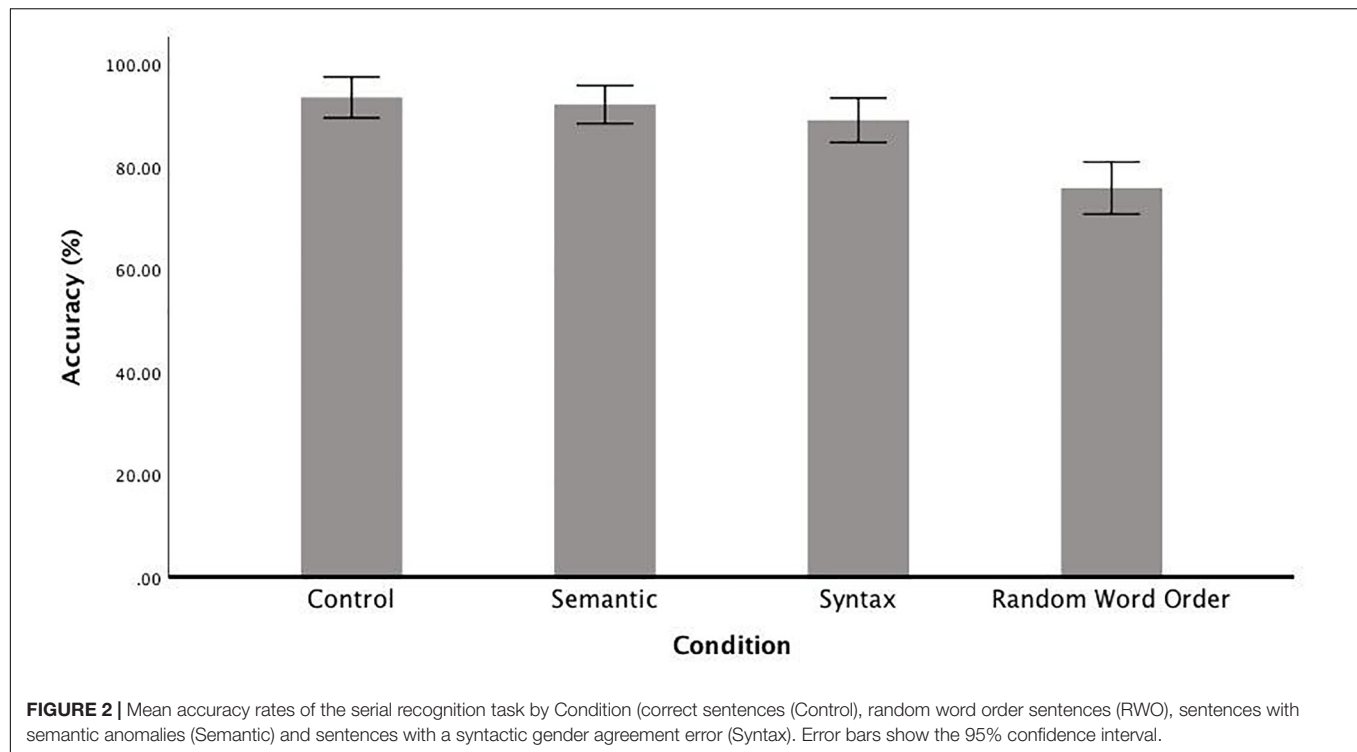
Response accuracy in the decision task was calculated for each condition and then descriptive statistics were obtained using SPSS (IBM Corp. Released 2017. IBM SPSS Statistics for MacOS,

Version 25.0. Armonk, NY, United States: IBM Corp.), and a Linear Mixed Effects (LME) analysis was carried out using the lme4 package (Bates et al., 2015) within R (R Core Team, 2018), with stimuli condition as a fixed effect and subject variability as a random effect. *P*-values were obtained by likelihood ratio tests comparing the full model against a null model. A series of *post hoc* pairwise *t*-test (Bonferroni corrected) were then completed. For the rehearsal check items, we evaluated the spoken responses of the participants and calculated the respective accuracy rates.

## RESULTS

### Behavioral

Accuracy in the serial recognition task, in which participants had to decide whether a two-word sequence had been previously presented in the identical form, was above chance level in



each of the conditions (**Figure 2**). The comparison between the full and the null models reveals that accuracy was affected by condition [ $\chi^2(1) = 77.03$ ,  $p < 0.001$ ]. To investigate this effect of condition, a series of *post hoc t*-test were carried out, which revealed no significant difference between the Control condition ( $M = 93\%$ ,  $SD = 9$ ) and the Semantic condition ( $M = 92\%$ ,  $SD = 9$ ) [ $t(23) = 1.11$ ,  $p = 1.00$ ], a significant difference between Control and Syntactic ( $M = 89\%$ ,  $SD = 10$ ) conditions [ $t(23) = 3.12$ ,  $p < 0.05$ ] and a significant difference between the RWO condition ( $M = 76\%$ ,  $SD = 12$ ) and all other conditions: Control [ $t(23) = 11.10$ ,  $p < 0.001$ ], Semantic [ $t(23) = 8.05$ ,  $p < 0.001$ ] and Syntactic [ $t(23) = 5.51$ ,  $p < 0.001$ ].

The rehearsal check items could unfortunately only be partly evaluated due to a technical error, due to which we recorded only responses up to 600 ms after production cue onset. Within that limited time window, participants produced an average of 51.3% ( $SD 15.5$ ) correct and 8.2% ( $SD 5.0$ ) incorrect answers.

## Event-Related Potentials

### Reading Phase

**Figure 3** displays the waveforms and topographic difference maps elicited by the Semantic condition compared to the Control condition in the reading phase, time-locked to the onset of the object position (e.g., *Die Frau bindet den Schuh im Flur* [*The woman ties the shoe in the hallway*]). The Semantic condition elicited a more negative-going waveform at left-central electrodes compared to the Control condition. Correspondingly, the cluster-based analysis showed a significant difference between the two conditions ( $p < 0.05$ ), originating from a negative cluster

observed at left-central sites beginning at around 360 ms and lasting until around 515 ms.

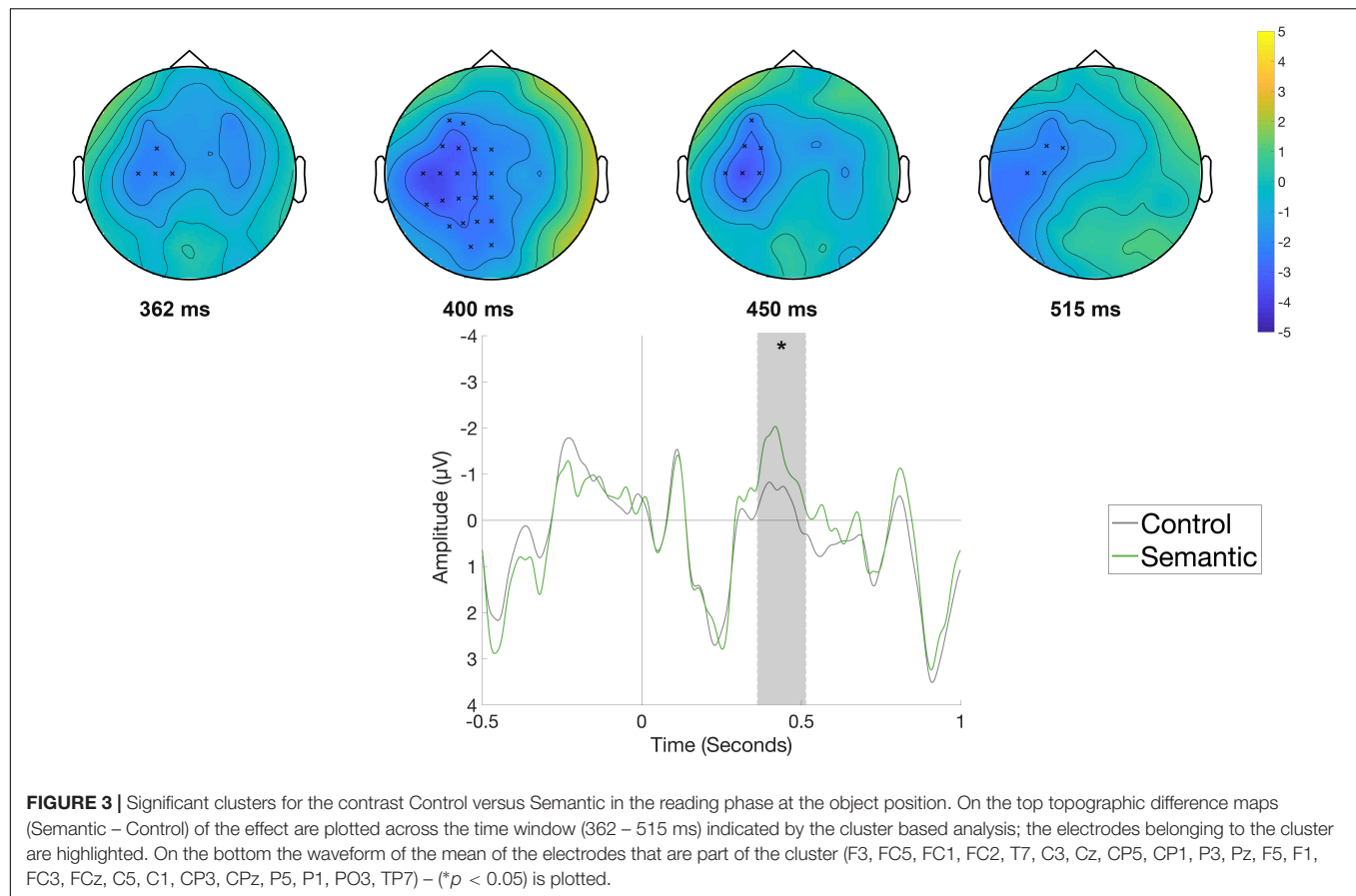
The comparison of the Syntactic and the Control condition (**Figure 4**), time-locked to the onset of the object position, indicated a more positive-going waveform at centro-posterior electrode sites for the Syntactic condition. The difference between conditions was significant ( $p < 0.01$ ), with the effect corresponding to an observed positive cluster with a centro-posterior distribution and an approximate latency of 500–1000 ms. The observed timing and the distribution of the effects in the data, with a negativity for the semantic violation and a positivity for the syntactic violation, led us to categorize the observed ERP patterns as classical N400 and P600 components.

### Rehearsal Phase

**Figure 5** displays the waveforms elicited by the Semantic and Syntactic conditions in the rehearsal phase compared with the Control condition time-locked to the onset of the object (e.g., *Die Frau bindet den Schuh im Flur* [*The woman ties the shoe in the hallway*]). The cluster-based permutation test did not reveal any significant differences between conditions. **Figure 6** shows the waveforms and topographic difference maps of the contrast of the Control versus the Semantic condition at the verb position (e.g., *Die Frau bindet den Schuh im Flur* [*The woman ties the shoe in the hallway*]) in the rehearsal phase. The Semantic condition elicited a relatively early negative deflection compared to the Control condition at fronto-central electrode sites. The cluster-based analysis thus indicated a significant negative cluster ( $p < 0.05$ ) between 114 and 214 ms with a fronto-central distribution.

The comparison of the Syntactic and the Control condition at the article position (e.g., *Die Frau bindet den Schuh im Flur*





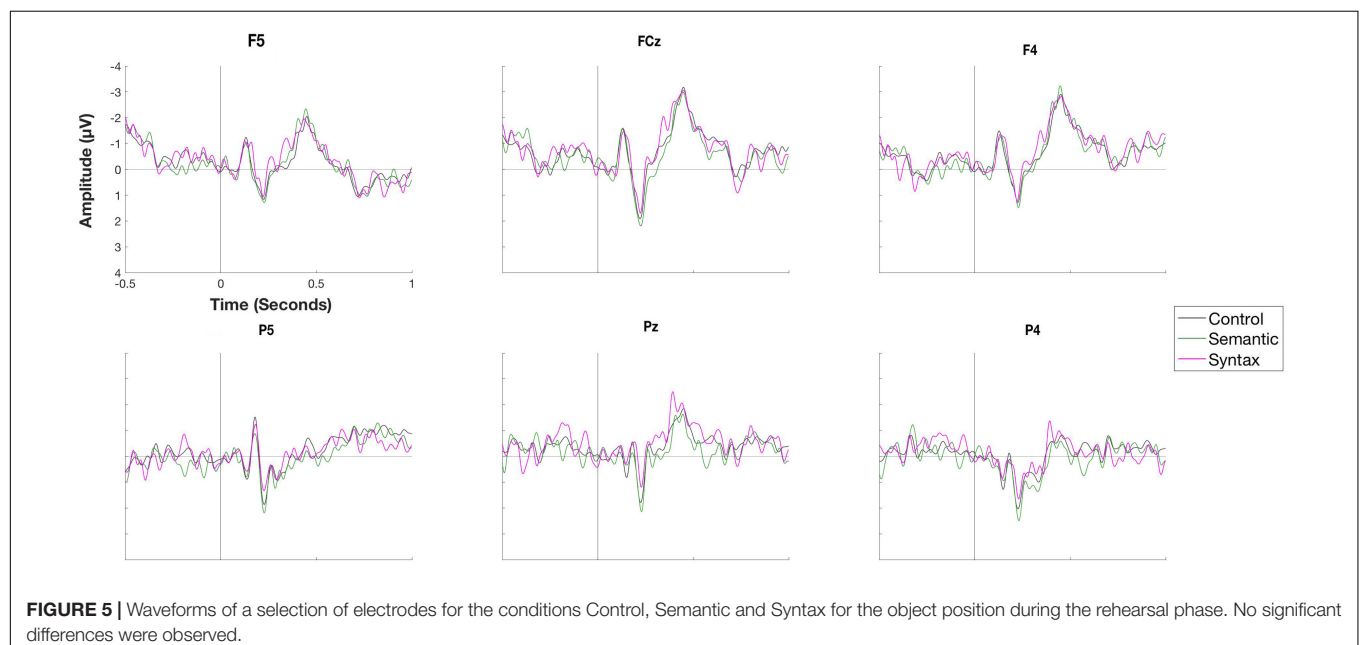
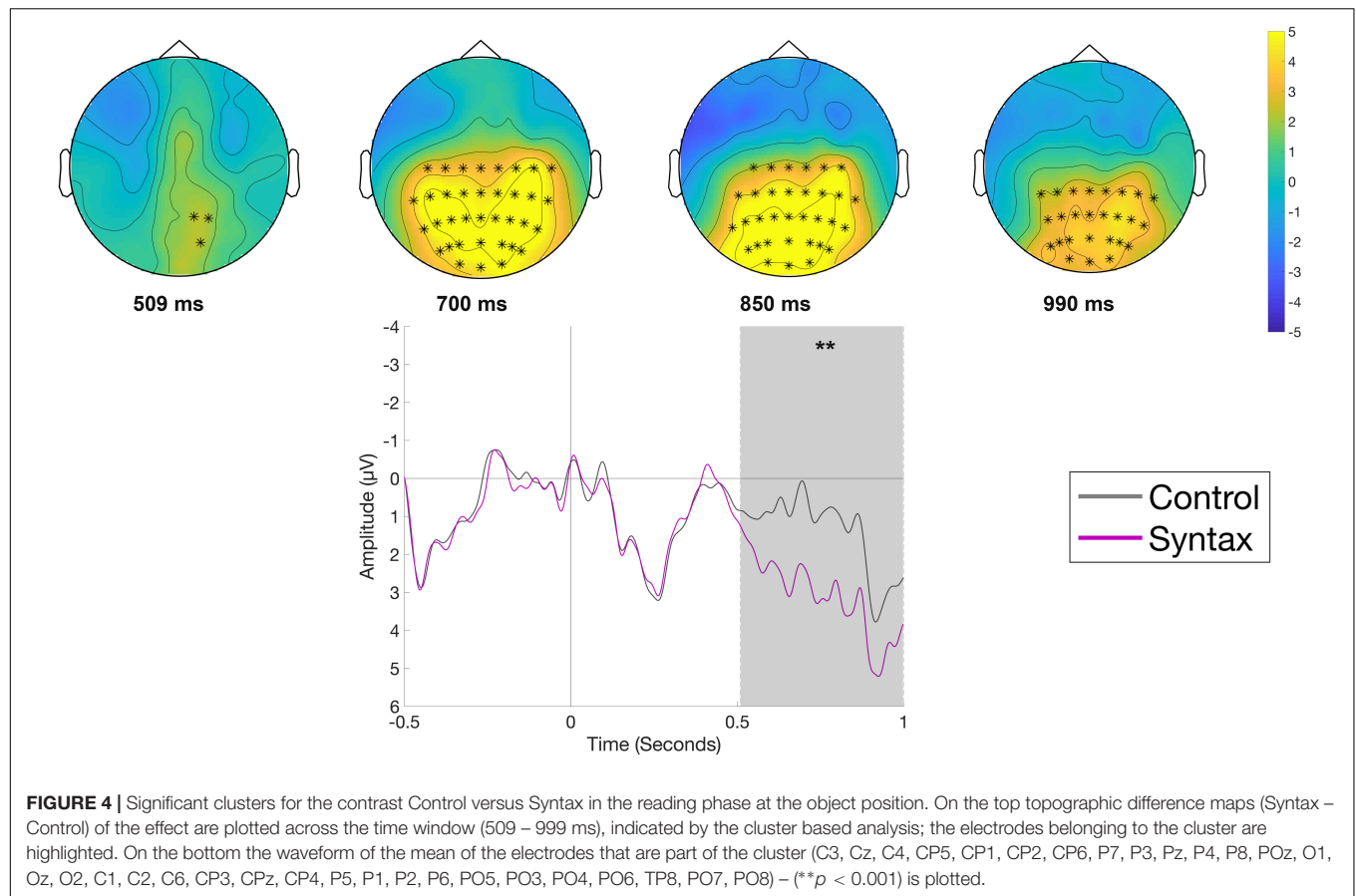
[The woman ties the shoe in the hallway.]) indicated a right frontal positivity for syntactically anomalous sentences. When the entire epoch was taken into account, no significant clusters were found. The analysis of a narrower time window between 500 and 700 ms (selected *a priori* as a time window for the P600) showed a significant positive cluster ( $p < 0.05$ ) between 580 and 674 ms with a right frontal distribution for the Syntactic condition (Figure 7). We would like to note that the significance of this effect depends on the application of high-pass filtering, which we chose in order to optimize ICA decomposition (cf. section EEG Data), while all other effects are also significant even when a more conservative filter (high-pass filter at 0.1 Hz, –6 dB, cut-off frequency 0.05 Hz) is applied.

### Rehearsal Phase – Slow Waves

For the analysis of the long epochs, given the nature of the slow waves, only clusters with significant time windows longer than 1 s (one-word time-window) will be reported. There were no significant clusters longer than 1 s for the contrasts between Control versus Semantic and Control versus Syntactic in the rehearsal phase. For the contrast Control versus Random Word Order, a significant negative cluster for the RWO condition ( $p < 0.01$ ) was identified in the time window between 1130 and 6500 ms with fronto-central distribution and a positive cluster ( $p < 0.01$ ) with right-posterior distribution in the time window between 2980 and 5900 ms, as displayed in Figure 8.

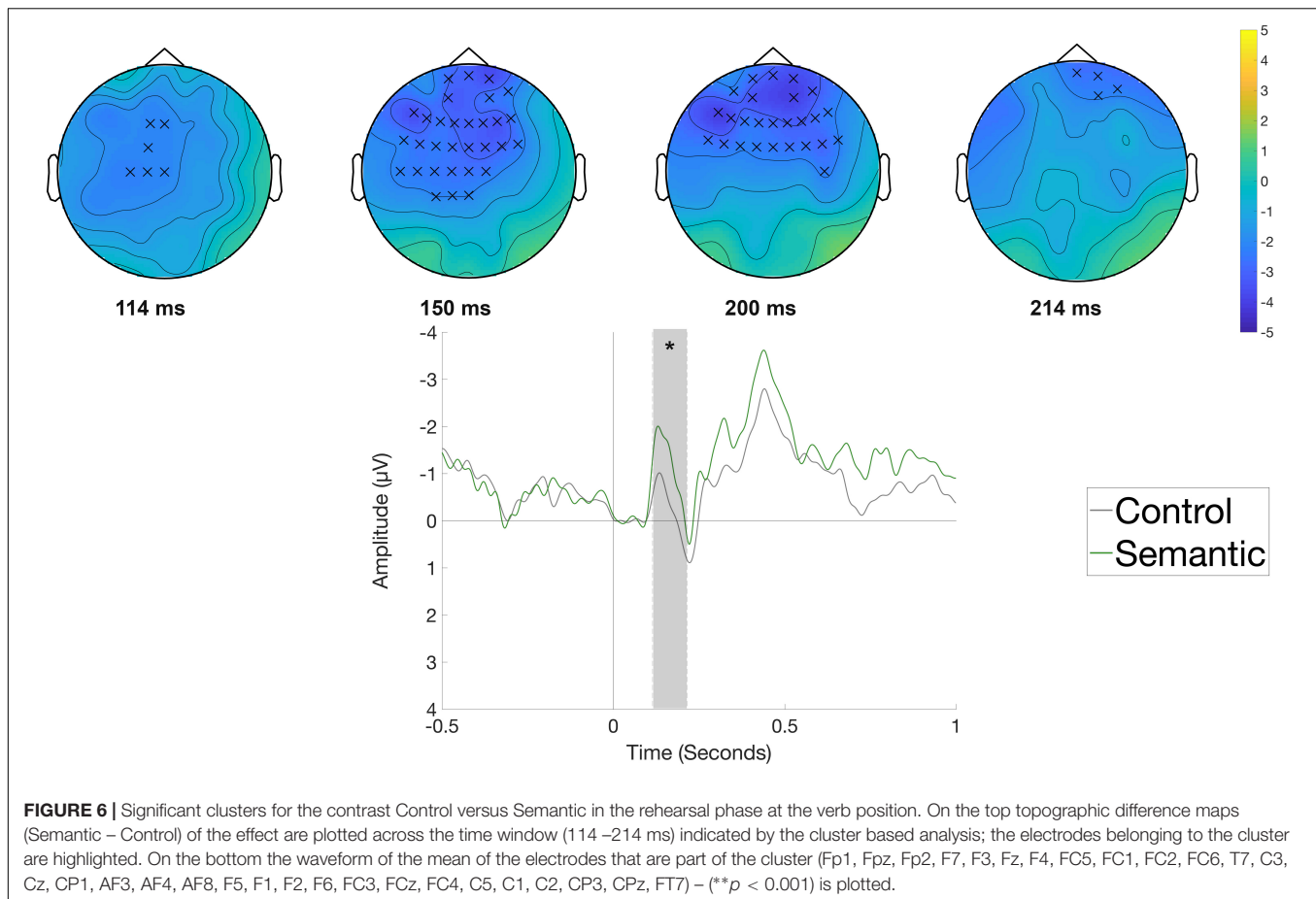
## DISCUSSION

The present study aimed to specify the language production processes supporting subvocal rehearsal of sentences in a working memory task. In order to ensure that participants engaged in subvocal rehearsal, an overt articulation cue was presented intermittently, during which participants had to produce the respective words. Importantly, the cues appeared unpredictably in the middle of sentences making sure that participants did not know in advance which words would have to be spoken out loud and which ones, silently. In more than half of the overt articulation trials, participants repeated the correct words within the first 600 ms after the articulation cue, showing that they were largely following the task. Unfortunately, technical problems precluded the analysis of responses after 600 ms from the articulation cue. This is problematic as 600 ms is the typical onset latency in many articulation tasks (Indefrey and Levelt, 2004) and thus, we probably missed many potentially correct answers. Yet, even though the final performance in the overt production trials cannot be reported, we are confident that participants engaged in subvocal rehearsal as (i) the task instruction was to do so, (ii) the overt production trials ensured commitment to the task and (iii) subvocal rehearsal was a good strategy to be able to answer the questions that followed the rehearsal (presence of a given two-word sequence).



The performance in the serial recognition task replicates the sentence superiority effect and shows that sentences, independently of the presence of semantic or syntactic anomalies,

are remembered better compared to ungrammatical word strings. Further, the ERP data show that rehearsal of unstructured word sequences compared to correct sentences was accompanied by a

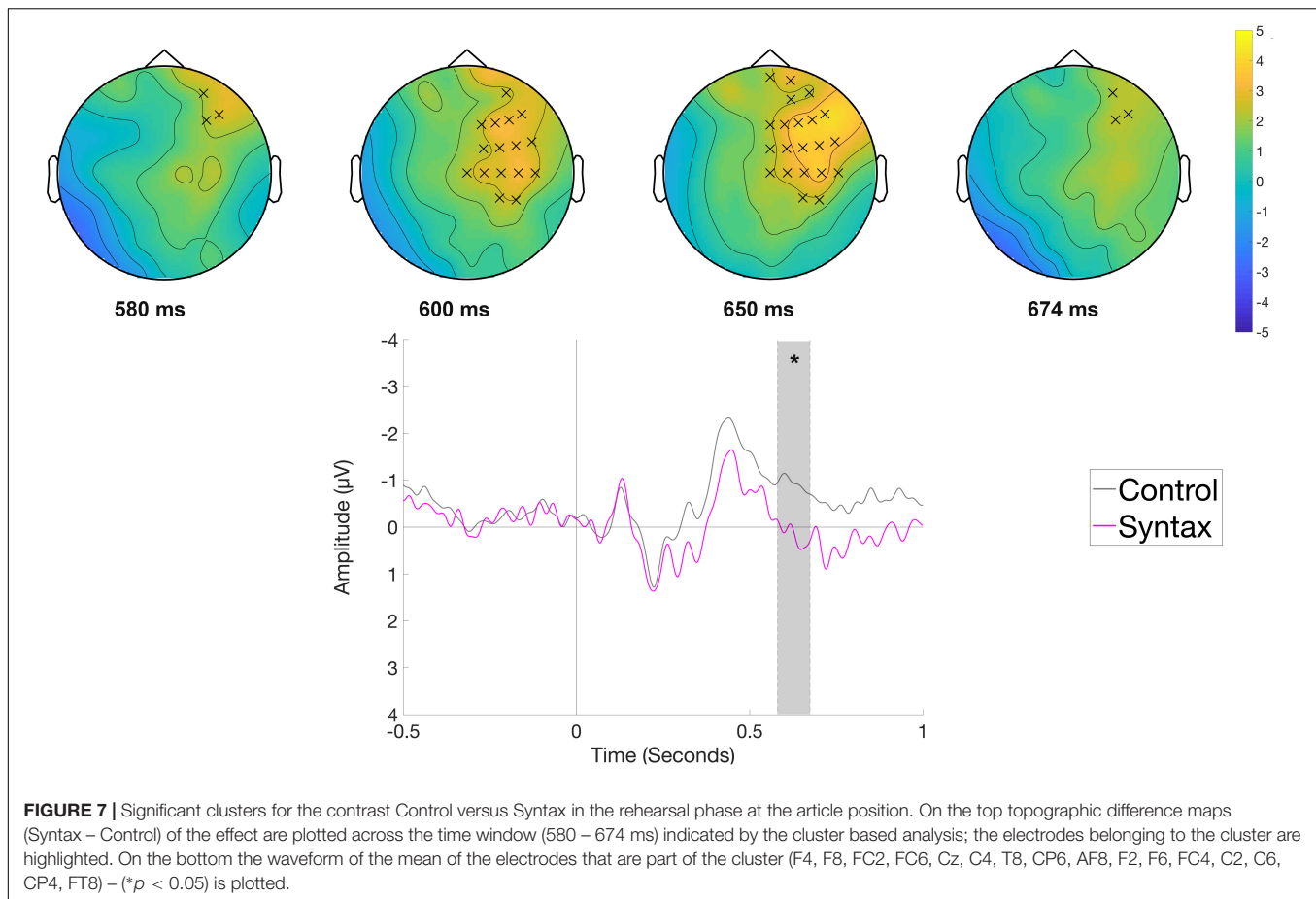


fronto-central negative shift covering 1.13 to 6.5 s and a bilateral posterior positivity between 2.98 and 5.9 s after rehearsal onset. In contrast, rehearsal of the semantic and syntactic violation conditions led to temporally and topographically different ERP responses at different sentence locations. In the semantic condition, a fronto-central negativity was found between 114 and 214 ms after the onset of the articulation cue at the position of the verb. In the syntactic condition, a positivity was found between 580 and 674 ms after the onset of the articulation cue for the syntactically incorrect determiner. In the following, we will first discuss the findings for each condition and then turn to outline the significance of the findings for conceptualizations of working memory and sentence production.

### Sentence Superiority Effect

Both behavioral and EEG data support increased processing costs for random strings of words. The accuracy in the serial recognition task was significantly better for sentences than for word sequences. ERPs were analyzed from the beginning of the repetition phase as the lack of structure and coherent meaning was assumed to induce an enhanced processing load from the start. Indeed, a long-lasting negative shift was evident from the onset of the cue for the second word until the onset of the cue for the last word. Additionally, a posteriorly distributed positive shift was observed, which started later and ended earlier compared

to the negative shift. The gradual onset of the effects might be interpreted as an indication of the attentional demands building up gradually with the first two items probably still benefiting from a primacy effect (Ebbinghaus, 1913). The negativity consisted of a frontal and a parietal portion. Previous ERP studies on verbal working memory have reported similar slow potentials, which varied depending on cognitive load and the stimulus material used (Lang et al., 1992; Ruchkin et al., 1992, 1997, 1999; Murphy et al., 2006). Initially, the frontal negative slow wave has been related to subvocal rehearsal proper (Ruchkin et al., 1992). Later studies have shown that it is also found in conditions where rehearsal is blocked and thus, it has been suggested that it is rather related to higher order cognitive control processes involved in verbal working memory (Bosch et al., 2001; Murphy et al., 2006). In the present study, articulation was manipulated in a different way than in most studies. Instead of blocking subvocal rehearsal by articulatory suppression (Murphy et al., 2006), articulation was enforced in the rehearsal phase. As participants were instructed to rehearse (and controlled for task compliance) it can be assumed that rehearsal occurred equally across all conditions. This means that the negativities in our experiment cannot be explained by assuming subvocal rehearsal in the more difficult condition and no rehearsal for the easier correct sentences. This is in line with the previous studies that showed negative shifts that were sensitive to the stimulus material, but not dependent on



the possibility of subvocal rehearsal. We suggest that the fronto-central shift represents the allocation of additional attentional resources in the light of the higher working memory demands. For example, it may reflect an upregulation of ‘multiple-demand’ cortical regions, that come on-line in response to increased task difficulty (Fedorenko et al., 2013; Geranmayeh et al., 2014; Sliwinska et al., 2017). One such region, the anterior insula, which is deeper but anatomically close to the inferior frontal gyrus, has been shown to be upregulated during sentence repetition tasks when comprehension of the sentence to be repeated is more difficult, due to degrading of the auditory signal, but not when the sentence is simple and easy to understand (Brownsett et al., 2014). The posterior slow waves, the negative and the positive shift might reflect more stimulus specific memory strategies. Posterior negative shifts have also been observed previously both for verbal and visual working memory tasks (Ruchkin et al., 1992; Bosch et al., 2001). Based on studies relating posterior slow potentials to visuo-spatial memory operations (e.g., Rösler et al., 1995a,b), Bosch et al. (2001) speculated that posterior slow potentials could be related to processes of transforming a visual to a phonological code. Bosch et al. (2001) also observed posterior positive shifts for visuo-spatial tasks in which no verbalization was possible. Thus, posterior slow waves could be related to image-based memory strategies. As in our experiment the words were presented visually, we tentatively adopt those ideas, namely that participants

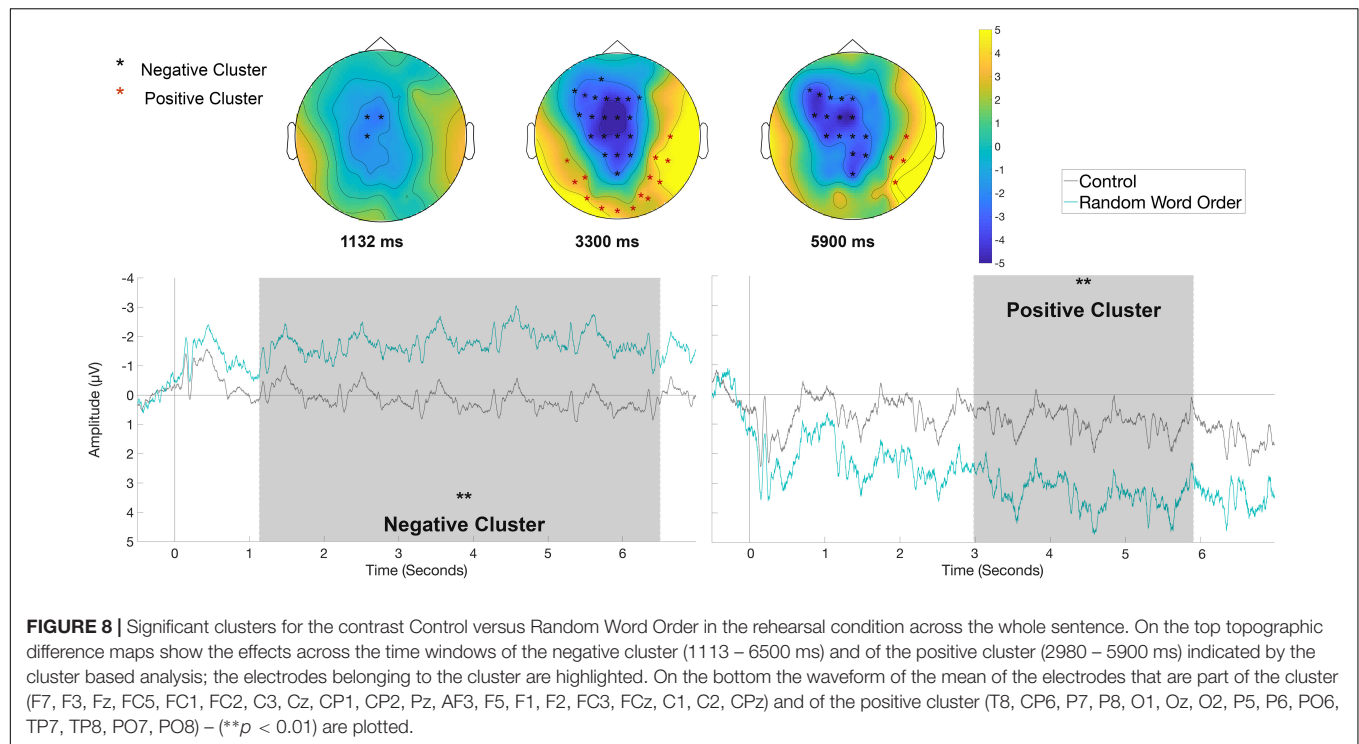
may both transform the visual code into a phonological one but that they also store the original visual code. Concerning the neural generation of slow waves, it has been suggested that signals from thalamic nuclei enhance the excitability of cortical areas, which Birbaumer et al. (1990) term “cerebral potentiality.” In this way, processing resources are allocated to specific cortical areas in preparation of a cognitive task (cf. Birbaumer et al., 1990). The interpretation of the slow waves in our study as reflecting the relatively enhanced attentional and strategic demands of the unstructured word lists is in concord with this model.

In sum, the sentence superiority effect is reflected in increased accuracy rates in a serial recognition task, and neurophysiologically, in a decrease in fronto-central and parietal slow waves which probably reflect enhanced costs in terms of cognitive control, visual and visual to phonological coding respectively.

## Lexical-Semantic Violations

While the memorization of unstructured word sequences seems to recruit domain-general networks that support working memory processing, as we argued above, the memorization of sentences that only include a semantic anomaly leads to different effects. Behaviorally, semantically anomalous sentences led to comparable accuracy rates in the serial recognition task as correct sentences. ERP analyses of the verb position revealed a significant





early fronto-central negativity peaking around 150 ms after cue onset for verbs that are later followed by a semantically unexpected noun compared to verbs from normal sentences. At the noun position, which yielded a semantic violation N400 effect during reading, no significant ERP effects were found. We take the early position of the effect as well as its early latency as an indication of advance planning of the direct object at the stage of abstract lexical planning. An extensive review by Indefrey and Levelt (2004) estimated that lexical selection for single word production occurs from about 150 to 350 ms after onset of a production cue. In our sentence repetition task, we do not know exactly when lexical selection of the verb started, but the early ERP response that largely covers the assumed time frame of that process, suggests that the semantically anomalous word that comes right after created some kind of processing cost at the lexical selection stage of the verb. This implies that the lexical planning of the verb and its arguments occurs at the same time or in fast sequence. Previous studies suggest that lexical planning in sentence production occurs at a rather large scope, spanning at least a single phrase or more (Smith and Wheeldon, 1999; Martin et al., 2010; Klaus et al., 2017). The early negativity at the verb shows that the planning scope comprises at least two words in advance, or maybe the entire phrase. Its timing as well as its distribution are inconsistent with an interpretation as an N400, which occurs at a later time window and with a different distribution. Similar effects with similar early timings have recently been reported in studies using a picture-guided noun phrase elicitation paradigm (Pykkänen et al., 2014; Blanco-Elorrieta et al., 2018). In these studies, participants composed adjective-noun combinations which were compared to the production of two single non-composable

words. The two-word combinations led to significant effects as measured with MEG starting from  $\sim 180$  ms. The effects could be localized to ventro-medial frontal cortex and to antero-lateral temporal cortex and were related to semantic composition independent of spoken or signed modality (Pykkänen et al., 2014; Blanco-Elorrieta et al., 2018). Similar effects of semantic composition were found during comprehension of equally simple noun phrases (Bemis and Pykkänen, 2011; Neufeld et al., 2016). Compared to the negativities reported in these studies, our negativity seems to occur even earlier. Note that we used a sentence repetition task, implying that the single words were already retrieved a relatively short time before the production cue. The studies on language production cited above used a picture-guided elicitation task instead, whereby participants have to first interpret the picture correctly, then retrieve the respective word without the picture being shown (Pykkänen et al., 2014; Blanco-Elorrieta et al., 2018). This task difference could lead to a shift in timing of the same or a similar effect. Note that no N400 was observed at the position of the actual semantic violation in our study. Yet, we know from the observed N400 in the reading phase that the direct object in the semantically anomalous condition induces semantic processing difficulties. This difference in the position and type of the ERP effect shows that the additional processing costs due to the semantically anomalous noun “have been paid before” during production and that there are no further integration difficulties at later stages.

## Syntactic Violations

Like with semantic anomalies, local ERP effects were found for the syntactically anomalous sentences during word-by-word silent production. Behaviorally, syntactically anomalous

sentences led to slightly lower accuracy rates in the serial recognition task than correct sentences. During the reading phase, a P600 effect was found at the noun position, but this position did not yield a significant effect during the rehearsal phase. Here, an effect was found after cue onset for determiners that are incongruent with the subsequent noun, although only when the statistical analysis was restricted to a time window of interest. The effect was a positivity with a right-frontal distribution peaking around 600 ms. We take the sentential position of the effect as well as its late latency as an indication of a mismatch between the planned or already articulated determiner form and the determiner form required by the gender of the subsequent noun.

Models of language production assume that gender information is stored in the mental lexicon linked to the noun either at the level of abstract lexical forms (Dell, 1986; Levelt, 1989) or at the level of phonological word forms (Miozzo and Caramazza, 1997). This implies that the noun has to be accessed either in its abstract or phonological form before gender information can be accessed. A previous ERP study on the production of simple determiner-noun phrases provided evidence that the phonological form of the determiner is accessed during or even after the phonological planning of the noun (Bürki et al., 2016). For the interpretation of our effect, this means that the mismatch effect we observe at the determiner indicates that the noun has become activated at that position. Further, the late onset of the ERP is consistent with the possibility that the encoding level at which the effect occurs is phonological encoding or a later process. The idea is based on the sequential nature of word production processes with phonological encoding occurring at some point after 200 ms after onset of production planning (Indefrey and Levelt, 2004). Thus, it seems plausible that the processing difficulty at the determiner occurs at that stage or later. Note that this remains speculative because in principle, ERP components can be influenced by cognitive processes occurring some time before their onset, so the nature of the difficulty experienced at that level is difficult to determine on the basis of the observed ERP pattern: it could be an effort to override a syntactic rule and thus directly related to a linguistic process; or it could be due to domain-general processing difficulties that accompany the former process. A study of Jessen et al. (2017) observed a right frontal negative deflection in response to production of regular participle forms compared to irregular participle forms in German. Although our design is different in the sense that the production of a correct noun phrase is compared to the production of an incorrect one, both experiments share the comparison of a rule-guided condition to an exceptional condition. In both cases, the exceptional condition elicited a right-frontal positivity. Thus, the positivity for the syntactic violation in our case could reflect the costs for overriding the established rule that specifies the determiner form according to the noun's gender. Another possibility would be that the effect is not related to determiner planning at all, but rather to production monitoring processes taking place in parallel. Shitova et al. (2017) reported a modulation of a relatively late occurring P300 component by complexity and task-switching during a noun phrase production task. They interpreted this

effect as related to the allocation and use of processing resources in the face of the affordances of the production task. Our late positivity could be related to similar processes, namely the additional attentional costs that come about by producing an outright grammatical error while the grammatically correct form is simultaneously activated.

Even if the exact processing level reflected by the positivity for the gender incongruent determiner is difficult to determine, the fact that the effect occurs on the determiner shows that the processing difficulties induced by the syntactically incorrect form appear at a later planning stage than the problems induced by improper lexical-semantic choices.

## Implications for Models of Language Production and Working Memory

In the present study we used sentence repetition to tackle the problem of how different types of linguistic information assist working memory. By doing this we also tapped, at least partly, into sentence production. Earlier studies with adults and many studies with developmental populations have used sentence repetition to test language production processes (Rodd and Braine, 1971; Ferreira, 1991; Brownsett et al., 2014; Klem et al., 2015). Admittedly, sentence repetition does not correspond to natural sentence production as the full form and content are already clear from the beginning and thus, working memory may be taxed much more and access and selection processes somehow less. Yet, theoretic models of sentence repetition converge in the assumption that language processing plays an important role in this task. The conceptual regeneration hypothesis, for example, posits that for the most part conceptual representations of sentences are stored and that syntactic and phonological aspects are generated during the process of repetition (Potter and Lombardi, 1990; Lombardi and Potter, 1992). Similarly, it is assumed in the context of Baddeley's multi-component model of working memory that the language processing system contributes to the advantage for memorizing sentences compared to word lists (Baddeley et al., 2009). Thus, sentence repetition may be a suitable method for assessing certain stages of language production. Obviously, due to the constrained nature of the task, the production process may not be entirely comparable to production in more natural contexts. The positions as well as the latencies of the ERP effects in response to silent rehearsal of our lexical-semantic and syntactic violations provide clear evidence that the respective types of information are accessed at different stages during the reproduction process. It is most plausible to assume that the respective violations created processing difficulties at those production levels where the critical information is in conflict with certain planning processes. The present findings are thus consistent with models of sentence production that assume different scopes of planning ahead for different types of information, as for example in the classic model of Garrett (1975, 1982), assuming at least a phrasal level of planning for abstract lexical forms and a smaller scope for concrete phonological realizations.

By integrating this evidence with models of working memory, the findings support those models that explicitly include the language production architecture in their maintenance mechanisms (Cowan, 1999; Buchsbaum and D'Esposito, 2019). The costs for maintaining sentences are dramatically decreased compared to ungrammatical word sequences, even in the presence of some embedded semantic or syntactic violations. The random word sequences induced processing costs all the way, while the ungrammatical sentence conditions showed only indications of local processing costs. This is consistent with the idea that working memory makes use of an incremental multi-staged language production process that benefits from lexical-semantic and syntactic relations between sentence parts as they are continuously integrated.

## CONCLUSION

Using a silent cued repetition task, a methodology that to our knowledge has never been used at sentence level, we found a sentence superiority effect for recognizing word sequences in sentences, compared to unstructured word sequences. Sentences with local semantic or syntactic violations were remembered comparably well, close to correct sentences, with a minor disadvantage for syntactically incorrect sentences. Electrophysiologically, a fronto-central and posterior slow wave reflected enhanced processing costs for the unstructured linguistic strings. Semantically and syntactically anomalous sentences, in contrast, yielded rather local processing costs reflecting the respective sentence planning stages at which the difficulties occurred, most likely access of abstract lexical forms and later phonological or monitoring processes. The results can be best explained by assuming that subvocal rehearsal of sentences in working memory includes typical stages of sentence planning, in line with working memory models that integrate the language architecture as a powerful supporting system. Finally, since the reported ERP effects are novel and in the case of syntactic anomalies statistically fragile, a replication of the effects would be highly desirable. In principle, the paradigm could become a valuable add-on in the toolbox for the study of the neurophysiological basis of on-line sentence production.

## REFERENCES

- Acheson, D. J., and MacDonald, M. C. (2009). Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychol. Bull.* 135, 50–68. doi: 10.1037/a0014411
- Allen, R. J., Hitch, G. J., and Baddeley, A. D. (2018). Exploring the sentence advantage in working memory: insights from serial recall and recognition. *Quar. J. Exp. Psychol.* 71, 2571–2585. doi: 10.1177/1747021817746929
- Asoulin, E. (2016). Language as an instrument of thought. *Glossa* 46, 1–23.
- Atkinson, R. C., and Shiffrin, R. M. (1968). "Human memory: a proposed system and its control processes," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, eds K. W. Spence, and J. T. Spence, (New York, NY: Academic Press), 89–195.
- Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Kommission für Forschungsethik (KFE) - University of Osnabrück. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MM conducted the research, from an original idea of RZ, created the design of the study under the supervision of JM, RZ, and FV, and performed the data analysis under the supervision of JM, FV, and TG. JM and MM wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript.

## FUNDING

We acknowledge the support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of Osnabrück University.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the precious contributions of Dr. Sonia Brownsett and Ms. Ivonne Weyers in the editing of the manuscript and Ms. Ina Hempen for the help with the data collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00412/full#supplementary-material>

- Baddeley, A. D., and Hitch, G. J. (1974). "Working memory," in *Recent Advances in Learning and Motivation*, ed. G. A. Bower, (New York, NY: Academic Press) 47–89.
- Baddeley, A. D., Hitch, G. J., and Allen, R. J. (2009). Working memory and binding in sentence recall. *J. Mem. Lang.* 61, 438–456. doi: 10.1016/j.jml.2009.05.004
- Baddeley, A. D., Lewis, V., and Vallar, G. (1984). Exploring the Articulatory Loop. *Quar. J. Exp. Psychol. Sec. A* 36, 233–252. doi: 10.1080/14640748408402157
- Baddeley, A. D., Thomson, N., and Buchanan, M. (1975). Word length and the structure of short-term memory. *J. Verb. Learn. Verb. Behav.* 14, 575–589. doi: 10.1016/S0022-5371(75)80045-4
- Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bemis, D. K., and Pytkkanen, L. (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal

- linguistic phrases. *J. Neurosci.* 31, 2801–2814. doi: 10.1523/JNEUROSCI.5003-10.2011
- Birbaumer, N., Elbert, T., Canavan, A. G., and Rockstroh, B. (1990). Slow potentials of the cerebral cortex and behavior. *Physiol. Rev.* 70, 1–41. doi: 10.1152/physrev.1990.70.1.1
- Blanco-Elorrieta, E., Kastner, I., Emmorey, K., and Pykkänen, L. (2018). Shared neural correlates for building phrases in signed and spoken language. *Sci. Rep.* 8:5492. doi: 10.1038/s41598-018-23915-0
- Boiteau, T. W., Malone, P. S., Peters, S. A., and Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *J. Exp. Psychol.* 143, 295–311. doi: 10.1037/a0031858
- Bolhuis, J. J., Tattersall, I., Chomsky, N., and Berwick, R. C. (2014). How could language have evolved? *PLoS Biol.* 12:e1001934. doi: 10.1371/journal.pbio.1001934
- Bonhage, C. E., Fiebach, C. J., Bahlmann, J., and Mueller, J. L. (2014). Brain signature of working memory for sentence structure: enriched encoding and facilitated maintenance. *J. Cogn. Neurosci.* 26, 1654–1671. doi: 10.1162/jocn\_a\_00566
- Bonhage, C. E., Meyer, L., Gruber, T., Friederici, A. D., and Mueller, J. L. (2017). Electrophysiological correlates of task-independent chunking during sentence processing. *NeuroImage* 152, 647–657. doi: 10.1016/j.neuroimage.2017.03.018
- Bor, D., Cumming, N., Scott, C. E. L., and Owen, A. M. (2004). Prefrontal cortical involvement in verbal encoding strategies. *Eur. J. Neurosci.* 19, 3365–3370. doi: 10.1111/j.1460-9568.2004.03438.x
- Bor, D., Duncan, J., Wiseman, R. J., and Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37, 361–367. doi: 10.1016/S0896-6273(02)01171-6
- Bosch, V., Mecklinger, A., and Friederici, A. D. (2001). Slow cortical potentials during retention of object, spatial, and verbal information. *Cogn. Brain Res.* 10, 219–237. doi: 10.1016/S0926-6410(00)00040-9
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Brener, R. (1940). An experimental investigation of memory span. *J. Exp. Psychol.* 26, 467–482. doi: 10.1037/h0061096
- Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., and Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain* 137, 242–254. doi: 10.1093/brain/awt289
- Buchsbaum, B. R., and D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778. doi: 10.1162/jocn.2008.20501
- Buchsbaum, B. R., and D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex* 112, 134–148. doi: 10.1016/J.CORTEX.2018.11.010
- Bürki, A., Sadat, J., Dubarry, A.-S., and Alario, F.-X. (2016). Sequential processing during noun phrase production. *Cognition* 146, 90–99. doi: 10.1016/j.cognition.2015.09.002
- Camos, V., Lagner, P., and Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *J. Mem. Lang.* 61, 457–469. doi: 10.1016/j.jml.2009.06.002
- Chein, J. M., Ravizza, S. M., and Fiez, J. A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *J. Neurolinguistics* 16, 315–339. doi: 10.1016/S0911-6044(03)00021-6
- Cowan, N. (1999). “An embedded-processes model of working memory,” in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, eds A. Miyake, and P. Shah, (New York, NY: Cambridge University Press), 62–101. doi: 10.1017/cbo9781139174909.006
- Cowan, N. (2010). The magical mystery four. *Curr. Direct. Psychol. Sci.* 19, 51–57. doi: 10.1177/0963721409359277
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321. doi: 10.1037/0033-295X.93.3.283
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Dunlosky, J., and Kane, M. J. (2007). The contributions of strategy use to working memory span: a comparison of strategy assessment methods. *Quar. J. Exp. Psychol.* 60, 1227–1245. doi: 10.1080/17470210600926075
- Ebbinghaus, H. (1913). *On Memory: a Contribution to Experimental Psychology*. New York, NY: Teachers College.
- Fedorenko, E., Duncan, J., and Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16616–16621. doi: 10.1073/pnas.1315235110
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *J. Mem. Lang.* 30, 210–233. doi: 10.1016/0749-596X(91)90004-4
- Fiebach, C., Schlesewsky, M., and Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *J. Mem. Lang.* 47, 250–272. doi: 10.1016/S0749-596X(02)00004-9
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011
- Garrett, M. F. (1975). “The analysis of sentence production,” in *The Psychology of Learning and Motivation*, ed. G. H. Bower, (New York, NY: Academic Press), 133–177. doi: 10.1016/S0079-7421(08)60270-4
- Garrett, M. F. (1982). “Production of speech: observations from normal and pathological language use,” in *Normality and Pathology in Cognitive Functions*, ed. A. Ellis, (London: Academic Press), 19–76.
- Geranmayeh, F., Brownsett, S. L. E., and Wise, R. J. S. (2014). Task-induced brain activity in aphasic stroke patients: what is driving recovery? *Brain* 137(Pt 10), 2632–2648. doi: 10.1093/brain/awu163
- Griesmayr, B., Gruber, W. R., Klimesch, W., and Sauseng, P. (2010). Human frontal midline theta and its synchronization to gamma during a verbal delayed match to sample task. *Neurobiol. Learn. Mem.* 93, 208–215. doi: 10.1016/j.nlm.2009.09.013
- Gunter, T. C., Friederici, A. D., and Schriefers, H. J. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568. doi: 10.1162/089892900562336
- Haller, S., Radue, E. W., Erb, M., Grodd, W., and Kircher, T. (2005). Overt sentence production in event-related fMRI. *Neuropsychologia* 43, 807–814. doi: 10.1016/j.neuropsychologia.2004.09.007
- Indefrey, P., and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/J.COgnITION.2002.06.001
- Jefferies, E., Ralph, M. A. L., and Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: the role of long-term and working memory. *J. Mem. Lang.* 51, 623–643. doi: 10.1016/j.jml.2004.07.005
- Jensen, O., and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Neuroscience* 15, 1395–1399. doi: 10.1046/j.1460-9568.2002.01975.x
- Jessen, A., Fleischhauer, E., and Clahsen, H. (2017). Morphological encoding in German children's language production: evidence from event-related brain potentials. *J. Child Lang.* 44, 427–456. doi: 10.1017/S0305000916000118
- King, J. W., and Kutas, M. (1995). Who did what and when – using word level and clause-level ERPs to monitor working-memory usage in reading. *J. Cogn. Neurosci.* 7, 376–395. doi: 10.1162/jocn.1995.7.3.376
- Klaus, J., Mädebach, A., Oppermann, F., and Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: effects of concurrent verbal and visuospatial working memory load. *Quar. J. Exp. Psychol.* 70, 811–831. doi: 10.1080/17470218.2016.1167926
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J.-E., and Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Dev. Sci.* 18, 146–154. doi: 10.1111/desc.12202
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063
- Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Neurosci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657
- Lang, W., Starr, A., Lang, V., Lindinger, G., and Deecke, L. (1992). Cortical DC potential shifts accompanying auditory and visual short-term memory.



- Electroencephalogr. Clin. Neurophysiol.* 82, 285–295. doi: 10.1016/0013-4694(92)90108-T
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532
- Lee, E.-K., Brown-Schmidt, S., and Watson, D. G. (2013). Ways of looking ahead: hierarchical planning in language production. *Cognition* 129, 544–562. doi: 10.1016/j.COgnITION.2013.08.007
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: The MIT Press.
- Levelt, W. J. M. (1994). “The skill of speaking,” in *International Perspectives on Psychological Science*, eds P. Bertelson, P. Eelen, and G. d’Ydewalle, (Hillsdale, NJ: 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
- Lombardi, L., and Potter, M. C. (1992). The regeneration of syntax in short term memory. *J. Mem. Lang.* 31, 713–733. doi: 10.1016/0749-596X(92)90036-W
- Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Front. Hum. Neurosci.* 7:357. doi: 10.3389/fnhum.2013.00357
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Marks, L. E., and Miller, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *J. Verb. Learn. Verb. Behav.* 3, 1–5. doi: 10.1016/S0022-5371(64)80052-9
- Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P. II, and Yang, C.-L. (2010). Planning in sentence production: evidence for the phrase as a default planning scope. *Cognition* 116, 177–192. doi: 10.1016/j.cognition.2010.04.010
- Meltzer, J. A., Kiehl, A., Panamsky, L., Links, K. A., Deschamps, T., and Leigh, R. C. (2017). Electrophysiological signatures of phonological and semantic maintenance in sentence repetition. *NeuroImage* 156, 302–314. doi: 10.1016/j.neuroimage.2017.05.030
- Meltzer, J. A., Rose, N. S., Deschamps, T., Leigh, R. C., Panamsky, L., Silberberg, A., et al. (2016). Semantic and phonological contributions to short-term repetition and long-term cued sentence recall. *Mem. Cogn.* 44, 307–329. doi: 10.3758/s13421-015-0554-y
- Meyer, A. S. (1996). Lexical access in phrase and sentence production: results from picture–word interference experiments. *J. Mem. Lang.* 35, 477–496. doi: 10.1006/jmla.1996.0026
- Miller, G. A., and Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *Am. J. Psychol.* 63:176. doi: 10.2307/1418920
- Miozzo, M., and Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the tongue states. *J. Exp. Psychol.* 23, 1410–1423. doi: 10.1037/0278-7393.23.6.1410
- Morrison, A. B., Rosenbaum, G. M., Fair, D., and Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Mem. Cogn.* 44, 922–936. doi: 10.3758/s13421-016-0608-9
- Murphy, K., Roodenrys, S., and Fox, A. (2006). Event-related potential correlates of the word length effect in working memory. *Brain Res.* 1112, 179–190. doi: 10.1016/J.BRAINRES.2006.07.007
- Murray, D. J. (1967). The role of speech responses in short-term memory. *Can. J. Psychol.* 21, 263–276. doi: 10.1037/h0082978
- Neufeld, C., Kramer, S. E., Lapinskaya, N., Heffner, C. C., Malko, A., and Lau, E. F. (2016). The electrophysiology of basic phrase building. *PLoS One* 11:e0158446. doi: 10.1371/journal.pone.0158446
- Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychol. Bull.* 131, 510–532. doi: 10.1037/0033-2909.131.4.510
- Nolan, H., Whelan, R., and Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *J. Neurosci. Methods* 192, 152–162. doi: 10.1016/j.jneumeth.2010.07.015
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/S1388-2457(00)00527-7
- Oppermann, F., Jescheniak, J. D., and Schriefers, H. (2010). Phonological advance planning in sentence production. *J. Mem. Lang.* 63, 526–540. doi: 10.1016/j.jml.2010.07.004
- Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z
- Peña-Casanova, J., Quiñones-Úbeda, S., Quintana-Aparicio, M., Aguilar, M., Badenes, D., Molinuevo, J. L., et al. (2009). Spanish Multicenter Normative Studies (NEURONORMA Project): norms for verbal span, visuospatial span, letter and number sequencing, trail making test, and symbol digit modalities test. *Arch. Clin. Neuropsychol.* 24, 321–341. doi: 10.1093/arclin/acp038
- Perham, N., Marsh, J. E., and Jones, D. M. (2009). Short article: syntax and serial recall: how language supports short-term memory for order. *Quar. J. Exp. Psychol.* 62, 1285–1293. doi: 10.1080/17470210802635599
- Pinker, S., and Bloom, P. (1990). Natural language and natural selection. *Behav. Brain Sci.* 13, 707–784. doi: 10.1017/s0140525x00081061
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Potter, M. C., and Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *J. Mem. Lang.* 29, 633–654. doi: 10.1016/0749-596X(90)90042-X
- Pylkkänen, L., Bemis, D. K., and Blanco Elorrieta, E. (2014). Building phrases in language production: an MEG study of simple composition. *Cognition* 133, 371–384. doi: 10.1016/j.cognition.2014.07.001
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rodd, L. J., and Braine, M. D. S. (1971). Children’s imitations of syntactic constructions as a measure of linguistic competence. *J. Verb. Learn. Verb. Behav.* 10, 430–443. doi: 10.1016/S0022-5371(71)80043-9
- Rösler, F., Heil, M., Bajric, J., Pauls, A. C., and Hennigshausen, E. (1995a). Patterns of cerebral activation while mental images are rotated and changed in size. *Psychophysiology* 32, 135–149. doi: 10.1111/j.1469-8986.1995.tb03305.x
- Rösler, F., Heil, M., and Hennigshausen, E. (1995b). Distinct cortical activation patterns during long-term memory retrieval of verbal, spatial, and color information. *J. Cogn. Neurosci.* 7, 51–65. doi: 10.1162/jocn.1995.7.1.51
- Ruchkin, D. S., Berndt, R. S., Johnson, R., Grafman, J., Ritter, W., and Canoune, H. L. (1999). Lexical contributions to retention of verbal information in working memory: event-related brain potential evidence. *J. Mem. Lang.* 41, 345–364. doi: 10.1006/jmla.1999.2644
- Ruchkin, D. S., Berndt, R. S., Johnson, R., Ritter, W., Grafman, J., and Canoune, H. L. (1997). Modality-specific processing streams in verbal working memory: evidence from spatio-temporal patterns of brain activity. *Cogn. Brain Res.* 6, 95–113. doi: 10.1016/S0926-6410(97)00021-9
- Ruchkin, D. S., Johnson, J. R., Grafman, J., Canoune, H., and Ritter, W. (1992). Distinctions and similarities among working memory processes: an event-related potential study. *Cogn. Brain Res.* 1, 53–66. doi: 10.1016/0926-6410(92)90005-c
- Rummer, R., and Schweppe, J. (2005). Evidence for a modality effect in sentence retention. *Psychon. Bull. Rev.* 12, 1094–1099. doi: 10.3758/BF03206449
- Sassenhagen, J., and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* 56:e13335. doi: 10.1111/psyp.13335
- Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain Lang.* 137, 29–39. doi: 10.1016/j.bandl.2014.07.010
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040. doi: 10.1073/pnas.0805234105
- Schnur, T. T. (2011). Phonological planning during sentence production: beyond the verb. *Front. Psychol.* 2:319. doi: 10.3389/fpsyg.2011.00319
- Schweppe, J., and Rummer, R. (2007). Shared representations in language processing and verbal short-term memory: the case of grammatical gender. *J. Mem. Lang.* 56, 336–356. doi: 10.1016/j.jml.2006.03.005
- Schweppe, J., Rummer, R., Bormann, T., and Martin, R. C. (2011). Semantic and phonological information in sentence recall: converging psycholinguistic and neuropsychological evidence. *Cogn. Neuropsychol.* 28, 521–545. doi: 10.1080/02643294.2012.689759
- Shitova, N., Roelofs, A., Coughler, C., and Schriefers, H. (2017). P3 event-related brain potential reflects allocation and use of central processing

- capacity in language production. *Neuropsychologia* 106, 138–145. doi: 10.1016/j.neuropsychologia.2017.09.024
- Sliwiska, M. W., Violante, I. R., Wise, R. J. S., Leech, R., Devlin, J. T., Geranmayeh, F., et al. (2017). Stimulating multiple-demand cortex enhances vocabulary learning. *J. Neurosci.* 37, 7606–7618. doi: 10.1523/jneurosci.3857-16.2017
- Smedt, K., and Kempen, G. (1987). “Incremental sentence production, self-correction and coordination,” in *Natural Language Generation*, ed. G. Kempen, (Dordrecht: Springer), 365–376. doi: 10.1007/978-94-009-3645-4\_23
- Smith, M., and Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition* 73, 205–246. doi: 10.1016/S0010-0277(99)00053-0
- Souza, A. S., and Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cogn. Psychol.* 107, 1–21. doi: 10.1016/j.cogpsych.2018.09.002
- Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychol.* 27, 285–292. doi: 10.1016/0001-6918(67)90070-4
- Turley-Ames, K. (2003). Strategy training and working memory task performance. *J. Mem. Lang.* 49, 446–468. doi: 10.1016/S0749-596X(03)00095-0
- van Herten, M., Kolk, H. H. J., and Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cogn. Brain Res.* 22, 241–255. doi: 10.1016/j.cogbrainres.2004.09.002
- Wagner, V., Jescheniak, J. D., and Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *J. Exp. Psychol.* 36, 423–440. doi: 10.1037/a0018619
- Waugh, N. C., and Norman, D. A. (1965). Primary memory. *Psychol. Rev.* 72, 89–104. doi: 10.1037/h0021797
- Wheeldon, L., and Lahiri, A. (1997). Prosodic units in speech production. *J. Mem. Lang.* 37, 356–381. doi: 10.1006/jmla.1997.2517
- World Medical Association (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *JAMA* 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.

Copyright © 2019 Mascelloni, Zamparelli, Vespignani, Gruber and Mueller. 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.



# Cross-Situational Statistical Learning of New Words Despite Bilateral Hippocampal Damage and Severe Amnesia

David E. Warren<sup>1\*</sup>, Tanja C. Roembke<sup>2</sup>, Natalie V. Covington<sup>3</sup>, Bob McMurray<sup>4</sup> and Melissa C. Duff<sup>3</sup>

<sup>1</sup>Department of Neurological Sciences, University of Nebraska Medical Center, Omaha, NE, United States, <sup>2</sup>Institute of Psychology, RWTH Aachen University, Aachen, Germany, <sup>3</sup>Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, United States, <sup>4</sup>Psychological and Brain Sciences, University of Iowa, Iowa, IA, United States

## OPEN ACCESS

### Edited by:

Joshua Oon Soo Goh,  
National Taiwan University, Taiwan

### Reviewed by:

Gerard Nisal Bischof,  
Julich Research Centre, Germany  
Itamar Lerner,  
Rutgers University, United States

### \*Correspondence:

David E. Warren  
david.warren@unmc.edu

### Specialty section:

This article was submitted to  
Cognitive Neuroscience, a section of  
the journal *Frontiers in Human  
Neuroscience*

**Received:** 30 June 2019

**Accepted:** 05 December 2019

**Published:** 14 January 2020

### Citation:

Warren DE, Roembke TC,  
Covington NV, McMurray B and Duff  
MC (2020) Cross-Situational  
Statistical Learning of New Words  
Despite Bilateral Hippocampal  
Damage and Severe Amnesia.  
*Front. Hum. Neurosci.* 13:448.  
doi: 10.3389/fnhum.2019.00448

Word learning requires learners to bind together arbitrarily-related phonological, visual, and conceptual information. Prior work suggests that this binding can be robustly achieved via incidental cross-situational statistical exposure to words and referents. When cross-situational statistical learning (CSSL) is tested in the laboratory, there is no information on any given trial to identify the referent of a novel word. However, by tracking which objects co-occur with each word across trials, learners may acquire mappings through statistical association. While CSSL behavior is well-characterized, its brain correlates are not. The arbitrary nature of CSSL mappings suggests hippocampal involvement, but the incremental, statistical nature of the learning raises the possibility of neocortical or procedural learning systems. Prior studies have shown that neurological patients with hippocampal pathology have word-learning impairments, but this has not been tested in a statistical learning paradigm. Here, we used a neuropsychological approach to test whether patients with bilateral hippocampal pathology ( $N = 3$ ) could learn new words in a CSSL paradigm. In the task, patients and healthy comparison participants completed a CSSL word-learning task in which they acquired eight word/object mappings. During each trial of the CSSL task, participants saw two objects on a computer display, heard one novel word, and selected the most likely referent. Across trials, words were 100% likely to co-occur with their referent, but only 14.3% likely with non-referents. Two of three amnesic patients learned the associations between objects and word forms, although performance was impaired relative to healthy comparison participants. Our findings show that the hippocampus is not strictly necessary for CSSL for words, although it may facilitate such learning. This is consistent with a hybrid account of CSSL supported by implicit and explicit memory systems, and may have translational applications for remediation of (word-) learning deficits in neurological populations with hippocampal pathology.

**Keywords:** word learning, amnesia, hippocampus, cross-situational statistical learning, statistical learning, declarative memory, relational memory

## INTRODUCTION

Statistical learning is the ability to learn from repeated (often incidental) exposure to probabilistic associations among elements of the input (Frost et al., 2019). This form of learning has been long-studied in the language literature, and it is posited to be particularly important for very early cognitive development of language (Saffran et al., 1996; Smith and Yu, 2008) as well as other domains. In language development, substantial learning occurs in complex environments that require segmentation of continuous input based on repeated exposure and probabilistic associations (Saffran et al., 1996; Karuza et al., 2013). Studies of infants, children, and adults suggest that statistical learning can occur at multiple developmental stages and can support learning at multiple levels of language (speech perception, word recognition, syntax, etc.; Saffran et al., 1996; Conway and Christiansen, 2005; Yu and Smith, 2007; Baldwin et al., 2008; Schapiro et al., 2016).

Recently, statistical learning has received attention in the memory literature (Schapiro et al., 2014; Covington et al., 2018). This attention has prompted new descriptions of the empirical phenomenon of statistical learning in terminology of *multiple memory systems*. The multiple memory systems perspective suggests that several unique brain systems support different types and rates of learning (Cohen and Squire, 1980; McClelland et al., 1995; Eichenbaum and Cohen, 2001; Norman and O'Reilly, 2003; Ranganath, 2010). Which of these systems support statistical learning? Novel findings from neuropsychological investigations indicate that certain domain-specific forms of statistical learning may (or may not) rely on memory processes associated with the medial temporal lobe and hippocampus (Schapiro et al., 2014; Covington et al., 2018). However, prior neuropsychological investigations have not tested statistical learning of multimodal associations. This is important because learning multimodal associations such as the mappings between new words and their referents (i.e., word learning) may span multiple learning systems. Further, the necessity of specific memory systems (and associated brain regions) for statistical learning of linguistic information such as words has not been evaluated.

## Statistical Learning and Multiple Memory Systems

Until very recently, statistical learning has been primarily an empirical phenomenon with an ambiguous relationship to theories of memory systems. Learning in a statistical context requires learners to extract consistent regularities (statistical associations) from repeated exposure to complex input which contains more than one element. Statistical learning has import for memory theory because the learned representations cannot be trivially categorized into a single type of memory representation described by theories of multiple memory systems.

Theories positing multiple memory systems were developed in part to address findings from neuropsychological studies of amnesic patients with damage to the medial temporal lobe (Scoville and Milner, 1957; Cohen and Squire, 1980).

These theories suggest that (at least) two types of memory representations are supported by unique brain correlates. Under this framework, *procedural* (or *non-relational*) memory stores information about individual elements of prior experience incrementally and in a manner that supports future expression under primarily *implicit* conditions (e.g., faster response times, increased sensitivity, or experience-dependent response bias). *Declarative* (or *relational-declarative*) memory stores information about relations between elements of prior experience rapidly and in a manner that supports future expression primarily under *explicit* conditions (e.g., free recall, old/new recognition, or multiple-choice recognition). Critically, neuropsychological studies indicate that the medial temporal lobes—including the hippocampus—are necessary for normal declarative-relational memory but not procedural memory (Scoville and Milner, 1957; Cohen and Squire, 1980; McClelland et al., 1995; Poldrack et al., 2001).

We note that the term “relational” has been used in psychology and neuroscience to describe various forms of representation (Eichenbaum and Cohen, 2001; Hummel and Holyoak, 2003; Cleland et al., 2007). Here, we use “relational” as discussed by Eichenbaum et al. (1992) who observed that the “... the critical property of declarative [relational] memory ... is the encoding of memories in terms of the relations among multiple items ...” (p. 3). In describing laboratory tests of relational memory, those authors noted that “[i]n some formal tests of memory, such as paired associate learning, demands for relational representation and/or representational flexibility—and hence declarative [relational] memory—are immediately evident” (p. 7; emphasis added).

In statistical learning, the incremental and incidental (i.e., implicit) acquisition of statistical associations between items strongly resembles the pace and function of procedural learning and representations. At the same time, statistical learning has also recently been studied in the context of learning mappings between words and objects (Yu and Smith, 2007; Smith and Yu, 2008; Roembke and McMurray, 2016; Roembke et al., 2018). This type of mapping requires that participants learn and express arbitrary relations (e.g., between faces and scenes, among sets of novel objects, or associations between words and referents), and relational representation is thought to rely on hippocampal-dependent relational-declarative representations (Eichenbaum et al., 1994; Eichenbaum and Cohen, 2001; Davachi and Dobbins, 2008; Ranganath, 2010). Because statistical word learning involves the incremental acquisition of arbitrary relations, describing the phenomenon using the terminology of multiple memory systems is challenging. This suggests that statistical learning paradigms requiring acquisition of arbitrary relations—such as word-referent learning—may provide novel opportunities to test and extend theories of multiple memory systems.

## Cross-Situational Statistical Learning and Multiple Memory Systems

Evidence for statistical forms of word learning comes from the cross-situational statistical learning (CSSL) paradigm. In



this paradigm, participants see an array of unfamiliar objects while hearing one or more novel word forms. Initially, the word-referent mapping appears completely random—there is no information to lead the learner to the correct referent. However, across trials, a given word is more likely to be heard with its referent than other objects. Hence, by tracking the co-occurrence between word forms and referents (objects), learners can acquire the mappings. This simple manipulation of statistical co-occurrence is sufficient to drive robust memory for word-referent pairings. Laboratory studies using this paradigm suggest that infants and adults can learn word-referent pairings from their environment through purely implicit statistical exposure (Saffran et al., 1996; Yu and Smith, 2007; Smith and Yu, 2008; Roembke and McMurray, 2016; Roembke et al., 2018).

CSSL of word-referent mappings has been hypothesized to be supported by various cognitive mechanisms. One hypothetical mechanism is gradual and associative: learners track associations between each word and multiple referents, and these associations reflect the relative evidence for a given mapping (Roembke and McMurray, 2016). An alternative instead relies on “single informative exposures”; here, learners form only a single hypothesis for a word’s referent, and update or reject this hypothesis during subsequent trials using inferential processes (Trueswell et al., 2013). Importantly, these hypothesized mechanisms need not be mutually exclusive and could function in parallel (Yurovsky and Frank, 2015). These cognitive mechanisms could be roughly mapped to components of the multiple memory systems framework. That is, the more gradual associative form of learning could be primarily mediated by non-hippocampal/non-relational systems, whereas the more inferential hypothesis-testing scenario could be supported by the hippocampal-relational system.

Note that in the CSSL paradigm, the association between the sound of a word and its referent is overwhelmingly an arbitrary relation. Yet, thousands of these arbitrary mappings are mastered by children during healthy language development (Bloom, 1973). According to one multiple memory systems perspective, learning about arbitrary associations between items is exclusively the domain of hippocampal-dependent relational-declarative memory (Eichenbaum and Cohen, 2001; Davachi and Dobbins, 2008; Ranganath, 2010). Under this theory—which challenges both the gradual associative account of learning and the hybrid account—a reasonable hypothesis would be that CSSL of new word-referent associations requires the hippocampus and will, therefore, be impaired in patients with hippocampal pathology. However, another possibility is that word-referent associations can be learned at least in part *via* statistical mechanisms with non-hippocampal brain correlates, and this would yield spared learning in patients with hippocampal pathology. Thus, the arbitrary nature of the mapping problem makes cross-situational statistical word learning a unique paradigm in which the contributions of multiple memory systems to statistical learning can be evaluated.

## Brain Correlates of Statistical Learning

Evidence from neuroimaging and neuropsychology is mixed regarding potential contributions of medial temporal lobe

regions to any form of statistical learning. Functional neuroimaging with fMRI has shown hippocampal activation during statistical learning of sequential dependencies in healthy young adults (Turk-Browne et al., 2009; Schapiro et al., 2016). Consistent with this, Schapiro et al. (2014) used a neuropsychological approach to study statistical learning of sequential dependencies in a patient with extensive medial temporal lobe damage (including the hippocampus). The patient performed at chance, and her performance was impaired relative to healthy comparison participants, suggesting that the medial temporal lobe may be necessary for statistical learning. Interpretation of these findings must be tempered, however, by results from a larger group of amnesic patients including some with focal hippocampal pathology (Covington et al., 2018). In that study, healthy comparison participants showed greater statistical learning than patients with focal hippocampal pathology. However, the patients still showed evidence of statistical learning that was above chance and often within the lower extent of the healthy range. Taken together, findings from previous studies suggest that the hippocampus may contribute to—but not be necessary for—statistical learning.

Previous neuropsychological studies of statistical learning have principally focused on sequential temporal dependencies among unimodal elements (syllables, tones, symbols, etc.). CSSL for words has not been examined in patients with hippocampal pathology. Critically, this form of feedback-free learning is arbitrary, temporally spaced, and multimodal—properties that may be consistent with hippocampus-dependent relational representations.

The current study is the first to explicitly test whether the hippocampus is necessary for CSSL. A role for the hippocampus in CSSL may have special relevance in early life (e.g., healthy development of language and vocabulary) and late-life (e.g., word-learning impairments in healthy and pathological aging). Relevant to this point, the hippocampus changes throughout life and both early development and late-life are periods when the hippocampus functions differently than in healthy maturity (i.e., young adulthood; Raz et al., 2004; Ghetti and Bunge, 2012; Ofen, 2012; Fjell et al., 2013). Prior work has established that the hippocampus is necessary for normal learning of word-referent mappings under certain explicit and implicit instructional regimes (Smith et al., 2014; Warren and Duff, 2014; but see Sharon et al., 2011). In contrast, Vargha-Khadem et al. (1997) reported results from children with perinatal or childhood hippocampal pathology who “... attained levels of speech and language competence, literacy, and factual knowledge ... within the low average to average range.” This suggests that hippocampus is not strictly necessary for ecological word learning. A recent study by Berens et al. (2018) studied CSSL in neurotypical adults using functional MRI. They found evidence for a quick learning mechanism that is consistent with rapid pattern separation processes in the hippocampus. However, CSSL for words has not been tested neuropsychologically in adults with bilateral hippocampal pathology. This is essential for understanding the role of the hippocampus in

different types of statistical learning (and for word learning more broadly).

Previous findings could support predictions for or against a hippocampal contribution to CSSL. Hippocampal amnesia has a profound negative impact on relational-declarative memory in general (Scoville and Milner, 1957; Cohen and Squire, 1980; Ryan et al., 2000; Hannula et al., 2006) and word learning specifically (Gabrieli et al., 1988; Postle and Corkin, 1998; Warren and Duff, 2014). Moreover, words exemplify the type of highly relational stimuli that require hippocampus (Warren and Duff, 2014). Prior studies of word learning by patients with hippocampal pathology have demonstrated that patients learn words more slowly and less successfully than healthy comparison participants under a variety of instructional conditions [e.g., explicit encoding (EE) and fast mapping; e.g., Warren and Duff, 2014]. These points suggest that hippocampus is necessary for normal CSSL.

However, CSSL paradigms are frequently implicit, and some studies have reported that the hippocampus is not necessary for word learning in implicit tasks (Sharon et al., 2011). Further, CSSL paradigms often employ a style of frequent repetition of stimuli that partly resembles procedural/non-declarative learning paradigms (e.g., errorless learning) in which patients with MTL or hippocampal pathology can learn as well as healthy comparison participants (Scoville and Milner, 1957; Cohen and Squire, 1980; Glisky et al., 1986; Glisky, 1992).

These contrasting perspectives illustrate the ambiguous state of the current literature, and they motivate a targeted study of hippocampal necessity for word learning in a CSSL paradigm. Further, by studying a unique form of statistical learning, such findings might expand and inform debates over hippocampal necessity for statistical learning more generally.

## Current Study

Here, we used a neuropsychological approach to test the necessity of the hippocampus for CSSL in the domain of word learning. We adapted a CSSL task that has been previously reported (Roembke and McMurray, 2016) for our study. In this task, participants learn statistical associations between phonological and visual information across many presentations. On each trial, participants hear one novel phonological word form and view two novel objects. Each word form is consistently presented (across trials) with a specific object; the other object is a randomly-selected competitor (itself associated with a different word). Thus,

the task requires learning a set of arbitrary relationships between phonological and visual stimuli in the presence of potentially interfering competitors. Damage to the hippocampus would be predicted to impair the relational memory abilities needed to learn such arbitrary relations. An EE task was also administered separately. The rationale for the EE task, which involved sequential exposure to word-referent associations without competitor items, was to measure simple (non-statistical) learning of arbitrary relations. Patient performance on each task was compared to healthy normal comparison participants.

Statistical learning of multi-modal associations is novel in this patient population. However, this study also expands on previous research studying other forms of statistical learning in patients with hippocampal pathology. In prior work (e.g., studying statistical learning of temporally adjacent dependencies), learning is assessed at a single time-point, in a two-alternative forced-choice (2AFC) post-test after the exposure phase (Schapiro et al., 2014; Covington et al., 2018). In contrast, we assess learning over time, and this will contribute to understanding the trajectory of statistical learning in the absence of hippocampal contributions.

## MATERIALS AND METHODS

### Participants

Three groups of participants were recruited. First, we recruited a group of patients ( $N = 3$ ) with hippocampal pathology (Table 1 and next paragraphs). All patients had participated in a prior study of hippocampal necessity for statistical learning (but not CSSL; Covington et al., 2018). Patients completed both a CSSL task and an EE task. Second, we recruited a group of healthy normal comparison participants (NC;  $N = 12$ ) with no history of neurological or psychiatric disease. These were used as comparisons for the novel CSSL task. Each NC participant was matched to one of the patients for sex, handedness, age ( $\pm 5$  years), and education ( $\pm 2$  years); in total, four NC participants were matched to each patient. This matching strategy was selected to provide sufficient statistical power to detect deficits in performance in the patient group based on prior research in healthy adults (Roembke and McMurray, 2016). Finally, another smaller group of healthy normal comparison participants ( $N = 4$ ) was recruited to complete the EE task (see below). As with the previous NC group, these participants were demographically matched to

**TABLE 1 |** Demographic and neuropsychological data characterizing participants with amnesia.

ID	Age	Sex	Edu.	Eti.	Chr.	Hand	FSIQ	VIQ	PIQ	DS	BNT	GMI	AVLT	CFT C/R	HcV
1846	52	F	14	An./SE	22	100	84	88	86	10	43	57	7/3	28/6	-4.23*
2363	59	M	18	An.	17	100	98	112	91	8	58	73	8/0	26/5	-2.64*
2563	61	M	16	An.	16	-80	94	91	98	14	52	63	10/4	36/7	NA

Individual scores are presented for each patient with hippocampal pathology. The significant memory impairment of the amnesic group is evident in several neuropsychological measures. Abbreviations: Age, years; Edu., education, years; Chr., Chronicity, years since injury; Hand, handedness (+100 = fully right-handed, -100 = fully left-handed); Eti., Etiology; Anoxia/An., anoxic/ischemic episode, SE, status epilepticus; FSIQ, WAIS-III full-scale IQ (Wechsler, 1997); VIQ, verbal IQ; PIQ, performance IQ; DS, WAIS 3/4 Digit Span; WMS-III GMI, general memory index (The Psychological Corporation, 1997); AVLT, Rey Auditory Verbal Learning Task, trial 5/30-min. delay; CFT, Complex Figure Task copy/recall (Rey, 1941; Osterrieth, 1944); BNT, Boston Naming Test (Goodglass et al., 1983); HcV, bilateral hippocampal volumes per Allen et al. (2006). Volumes are expressed in Studentized residuals relative to normative expectations: NA, volumetric measurements unavailable due to contraindication for MRI.

the patients (here, one-to-one). We had a strong *a priori* expectation that the massed practice of the EE condition would yield ceiling performance in NC participants (which was confirmed), so this second NC group was recruited principally for proof-of-method.

Patients had severe, selective deficits in declarative memory according to neuropsychological assessments (**Table 1**). Impairment of declarative memory (including visual and verbal domains) was evident in patients' profoundly impaired performance ( $\geq 2$  SD below normal) on the WMS-III General Memory Index, Rey Auditory-Verbal Learning Task, and Rey-Osterrieth Complex Figure Task. Other cognitive abilities were generally preserved and in the normal range. Because naming abilities may be of special importance for word learning and CSSL, we also considered a neuropsychological measure of naming. Results of the Boston Naming Test indicated that naming performance was normal for patients 2363 and 2563 but impaired for patient 1846 (43/60, first percentile). However, 1846 performs normally when naming animals, fruits, and vegetables (Warren et al., 2012, p. 347). We interpreted 1846's pattern of performance on naming tasks as evidence that her naming abilities were sufficiently well-preserved for her to participate in this study.

Patients had pathological bilateral atrophy of the hippocampus as confirmed by neuroimaging studies. Two patients (1846 and 2363) had substantial atrophy of the hippocampus confirmed with high-resolution T1-weighted MRI (Allen et al., 2006). In that report, the authors used previously established estimates of adult hippocampal volume (measured through manual tracing) from T1-weighted MRI data of healthy adults age 22–88 (Allen et al., 2005). Adjusted for age and sex based on a regression model fit to the normative data, the hippocampal volume of patient 1846 was 4.23 standard deviations below normal expectations (53% reduction); for patient 2363, hippocampal volume was 2.64 standard deviations below normal expectations (28% reduction). Patient 1846 was later studied with ultra-high-resolution T2-weighted MRI (Warren et al., 2012). Analysis of those data confirmed the earlier findings of hippocampal atrophy greatly exceeding expectations for age. The remaining patient (2563) wears a pacemaker and is contraindicated for MRI studies. His anatomy was instead visualized with computerized tomography and atrophy of the hippocampal region was reported (but not quantified) by an expert rater (Hannula et al., 2006).

Patients were recruited from the Iowa Registry of Neurological Patients. Comparison participants were recruited from Iowa City and surrounding communities. This research was approved by the University of Iowa Human Subjects Office and by the Biomedical Institutional Review Board, and the study was conducted according to the principles expressed in the Declaration of Helsinki. Informed consent was obtained from all participants prior to their first experimental session. Consent documents described the study's purpose as follows: "... to investigate whether certain regions of the brain participate in the learning and expression of names." All participants were remunerated at \$15/h.

## Stimuli

Materials were auditory and visual stimuli that have been previously described (Roembke and McMurray, 2016). Visual stimuli were novel visual objects superimposed on a black background (**Figure 1**). Auditory stimuli were two-syllable, consonant-vowel-consonant-vowel (CVCV) pseudowords which were phonologically legal in English. There was no phonological overlap among any words at the onset. Words were recorded by a native speaker of English, and five tokens of each word were used to include natural variability in the phonological representation of the word. All materials were pre-experimentally unfamiliar to participants.

## Equipment

Visual stimuli were presented on a 21-in LCD monitor (Multi-Sync 2190UXi, NEC Corporation of America, Irving, TX, USA) at a distance of 550 mm. Behavioral responses were made with a computer mouse. During the tasks, subjects placed their head in a padded chinrest/headrest apparatus, and eye movements were monitored at a sampling rate of 1,000 Hz using an EyeLink 1000 remote infrared camera system (SR Research Limited, Kanata, ON, Canada). Calibration procedures were conducted every 30 trials and ensured that gaze position was accurate to within 1° of visual angle.

## Procedure

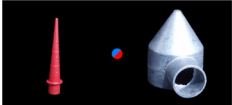
### Cross-Situational Statistical Learning

Participants completed a set of tasks designed to test CSSL of words (**Figure 1**). Our procedure was similar to that of Roembke and McMurray (2016). There were three phases. First, participants completed a learning phase in which visual and auditory stimuli were presented; each learning trial required a response for learning assessment. Second, memory for the auditory word forms was tested using a 2AFC format. Third, memory for the visual stimuli was tested using a 2AFC format. Visual and auditory stimuli were unfamiliar to the participants prior to the experiment.

During the CSSL phase, the participant was told that their task was to learn which visual stimulus ("object") was paired with which auditory stimulus ("word"). During each trial, two objects were presented along with one word (**Figure 1A**); the participant was instructed to select the object associated with the word. Participants were told that initially their selections would be guesses, but they should learn the associations over time. The experimenter ensured that all participants understood the instructions before testing began.

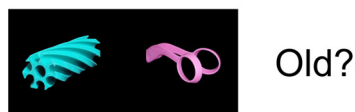
During each trial, two objects were presented on the left and right sides of the display with a blue dot in the center. Participants were required to fixate the dot to continue. After 1,050 ms, the blue dot turned red signaling the participant to click the dot. Clicking the red dot then triggered the presentation of the word. After hearing the word, the participant clicked on one of the objects to advance to the next trial. No feedback was provided following the response. The referent associated with the word was presented equally often in the left and right positions across trials. For the patients, the experimenter checked between blocks and as

**A Cross-situational statistical learning**

Trial	Objects	Word
1		<i>jifei</i>
2		<i>zati</i>
3		<i>bure</i>
4		<i>jifei</i>
5		<i>tagu</i>
6		<i>zati</i>
	...	

**B Post-CSSL auditory recognition**

*jifei*   *naida*   Old?

**C Post-CSSL visual recognition**

**FIGURE 1 |** Procedure for cross-situational statistical learning (CSSL) and recognition testing. Our procedure adopted the approach of a previous study (Roembke and McMurray, 2016) to implement and test CSSL. **(A)** The procedure for cross-situational learning involved studying (auditory-visual) word-object pairs accompanied by a competitor object. The association of a specific word with a specific object (e.g., word *jifei* with the spiral blue object) was invariant across trials but not immediately obvious to participants because of the competitor object. Participants selected the object they believed was associated with the word to advance to the next trial. Eight word-object pairs were presented 14 times per block; three blocks were completed. **(B)** After the CSSL task, memory for the auditory word stimuli was tested using a two-alternative forced-choice (2AFC) recognition test. Two words (one studied, one novel) were presented auditorily in sequence, and the participant decided which had been studied. **(C)** Memory for the visual object stimuli was also tested using a 2AFC recognition test. Two objects (one studied, one novel) were presented on the display, and the participant decided which had been studied.

needed to ensure that patients' understanding of the instruction set was maintained throughout testing.

Within the CSSL phase, eight word-image pairs were presented 14 times each per block; three blocks were administered. Word-image pairs were unique for each patient

(and their matched NC participants). By design, the word presented during each trial was uniquely associated with one of the objects, and the association was deterministic (i.e., a given word was exclusively presented in the presence of its paired object). The second, competitor object was selected at random from the non-paired objects. The random selection was made without replacement to avoid unintentional statistical association with a word; thus, the co-occurrence of each word with each non-paired object was one in seven (14.3%) vs. 100% for the paired object.

After the learning phase, two recognition tests were administered. In the auditory recognition test, the participant was asked to identify which of two words had been presented during the learning phase. The target was a word from the learning phase; the competitor item was intraexperimentally novel but otherwise had similar stimulus properties to words from the CSSL phase. The two words were presented sequentially separated by a short, silent pause. Simultaneously with each word, a colored square (orange and blue for the two words, respectively) appeared. The participant was instructed to tell the experimenter which item was studied, and the experimenter recorded the response. The interactive display allowed the participant to replay either word *ad-lib* (by clicking the orange or blue square) until a decision had been made. Once the participant's response was recorded, the trial was advanced to another auditory recognition trial until all words from the learning phase had been tested. The target word was presented equally often in the first and second (and thus, left and right) positions.

The visual recognition phase followed a similar logic. The participant was asked to identify which one of two objects had been presented during the learning phase. The target was an object from the learning phase; the competitor item was intraexperimentally novel but had similar stimulus properties. The two objects were presented simultaneously on the display at the left and right sides of the display. The participant observed the test display, then responded by clicking on one object (the studied object) with the mouse, thus advancing the trial. All objects from the learning phase were tested using this approach. The target item was presented equally often in the left and right positions.

All patients with hippocampal pathology completed the CSSL task along with 12 NC participants (matched 4:1 as described above).

### Explicit Encoding

To contrast with the CSSL task, we also administered an EE task. In the EE task, each trial presented a single unfamiliar visual stimulus (object) along with a single unfamiliar auditory stimulus (word). To advance to the next trial, participants clicked the mouse on the (single) object. As in the CSSL condition, no feedback was provided during learning exposures. After 28 exposures to each of eight word-object pairs (14 presentations/block  $\times$  2 blocks), participants completed a 2AFC recognition test which matched the format of CSSL learning blocks (8 word-object pairs  $\times$  7 presentations = 56 trials). No feedback was provided



during the 2AFC recognition test. All patients with hippocampal pathology completed the EE task along with four task-naïve NC participants (matched 1:1 as described above). Importantly, stimuli presented during the EE and CSSL tasks were unique and not overlapping (thus, word-referent learning during the CSSL task did not influence EE performance). For patients, the EE task was always administered after the CSSL task and in a separate test session.

## Analysis

Data were aggregated using Python 3.6 and Python's *panda*'s module. Data were analyzed and visualized using R 3.5.1 and the *lme4*, *afex*, *psycho*, and *ggplot2* libraries. All statistical tests used  $\alpha = 0.05$  to determine statistical significance. This value was corrected for multiple comparisons in cases described below, and this correction is indicated with  $p_c$  in the Results.

Data from the CSSL learning phase were analyzed to assess group and individual trends in accuracy across learning exposures. Specifically, the 336 total trials (8 word-object pairs  $\times$  14 presentations/block  $\times$  3 blocks) for each participant were divided into six sequential epochs of 56 trials each. Performance during each epoch was operationalized as the proportion of trials in which the participant selected the object paired with the word in our experimental design. This proportion correct (prop. correct) measure for each participant in each epoch is plotted to illustrate performance by block.

First, we tested whether the NC group showed a learning trend across time and whether NC participants matched to different patients performed differently. This was analyzed using a generalized linear mixed-effects model with a binomial link function as implemented in the R package/function *afex::mixed*. The model had fixed effects for learning epoch (factor, levels: 1–6) and matched patient (factor, levels: 1846, 2363, and 2563); the participant was a random effect (intercept).  $\beta$  weights

for factor levels were tested for statistical significance with the likelihood ratio method; statistically significant differences among factor levels were tested with a chi-squared ( $\chi^2$ ) statistic.

Second, we tested whether patient performance differed from NC performance using a Bayesian implementation of Crawford's modified *T*-test (Crawford and Garthwaite, 2007). This was applied at each epoch; we corrected  $\alpha$  for this test using Bonferroni's method with a correction factor of six (i.e.,  $\alpha = 0.05/6 = 0.0083$ ); Bonferroni-corrected tests are indicated with  $p_c$ .

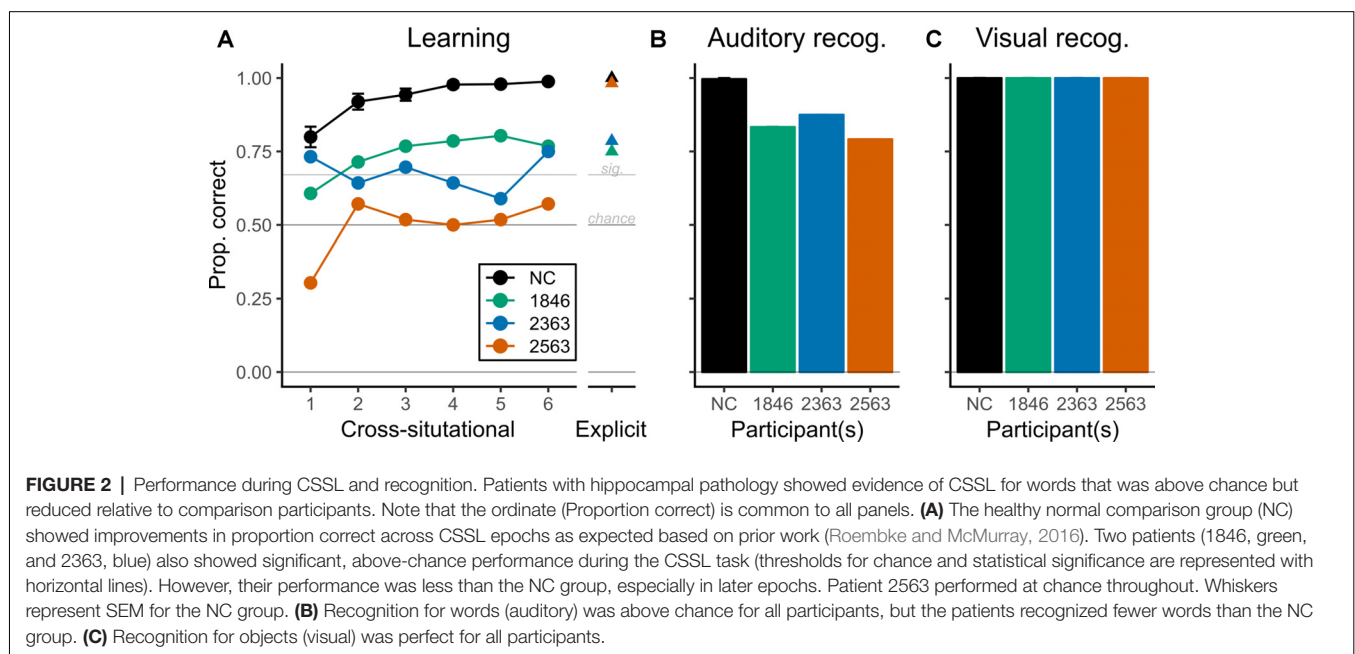
Third, we tested whether the patient performance was greater than chance (prop. correct = 0.5) using a one-sided binomial test; this test was corrected for multiple comparisons as before.

Data from the auditory and visual recognition tests were analyzed to test group differences in recognition performance after CSSL exposure. When sufficient variance was available in the NC group, we used Crawford's modified *T*-test to assess whether patients performed in the NC range. Also, we tested whether the patient performance was greater than chance (prop. correct = 0.5) using a one-sided binomial test.

Data from the EE task were analyzed to test group differences in learning after EE exposure. When sufficient variance was available in the NC group, we used Crawford's modified *T*-test to assess whether patients performed in the NC range. Also, we tested whether the patient performance was greater than from chance (prop. correct = 0.5) using a one-sided binomial test.

## RESULTS

In the CSSL task, two of three patients with hippocampal pathology showed above-chance performance for word-object associations by the final epoch, but performed less well than the NC group (Figure 2A); the third patient did not show evidence of learning.



Specifically, the NC group showed no differences by matched patient,  $\chi^2_{(2)} = 3.889$ ,  $p = 0.143$ , but performed above chance in each epoch (each  $T_{(11)} > 8.5$ , each  $p_c < 0.001$ ) and showed differences in accuracy across epochs,  $\chi^2_{(5)} = 253$ ,  $p < 0.001$ , such that performance increased monotonically (**Figure 2A**, black line). With no evident differences by matched patient, the NC group was combined for all subsequent analyses.

Similar to the learning trend observed in the NC group, patient 1846 showed monotonically increasing performance across the first five epochs and performed statistically better than a chance for epochs 2–6 (each  $p_c < 0.001$ ). Although her performance was not statistically different from the NC group in epochs 1–3 (each  $p_c > 0.01$ ), her performance was significantly less than the NC group in epochs 4–6 (each  $p_c < 0.001$ ). Patient 2363 also showed learning but presented a less consistent pattern of performance. He performed statistically above chance in epochs 1, 3, and 6 (each  $p_c < 0.0025$ ) and had performance statistically less than the NC group in all but the first epoch (each  $p_c < 0.01$  for epochs 2–6). Finally, patient 2563 showed no significant evidence of any learning during the CSSL task: he never performed above chance (each  $p_c > 0.175$ ), and his performance was always less than the NC group (each  $p_c < 0.0025$ ). To reiterate, two of three patients showed evidence of learning word-object associations during the CSSL task while the third did not.

Auditory and visual recognition performance after CSSL exposure suggested that patients retained the knowledge of the individual studied stimuli, although recognition performance relative to the NC group diverged by modality. In the auditory recognition task, the NC group was effectively at the ceiling—11 of 12 NC participants performed without error. All patients performed well below ceiling but also significantly above chance (prop. correct: 1846 = 0.833; 2363 = 0.875; 2563 = 0.792; each  $p < 0.01$ ; **Figure 2B**). Visual recognition performance was perfect for all participants which suggested good retention by patients but prevented formal between-group statistical tests (**Figure 2C**).

In the EE task, two of three patients (1846 and 2363) performed almost identically to their last-epoch CSSL performance (prop. correct: 1846, CSSL = 0.77 vs. EE = 0.75; 2363, CSSL = 0.75 vs. EE = 0.79) while patient 2563 showed a marked improvement (CSSL = 0.57 vs. EE = 0.98; **Figure 2A**, rightmost points). All patients performed significantly above chance (each  $p < 0.001$ ). The secondary NC group ( $N = 4$ ) performed without error. Thus, after EE exposure all patients had above-chance learning which was less than NC performance (albeit only slightly for 2563).

## DISCUSSION

We found that two of three patients with bilateral hippocampal pathology were able to learn new word-object associations in a CSSL paradigm. This is consistent with the suggestion that the hippocampus is not strictly necessary for statistical learning (Covington et al., 2018)—and conversely, that non-hippocampal brain regions can support statistical learning. Our findings are

also consistent with neuropsychological findings indicating that patients with amnesia due to MTL or focal hippocampal pathology can sometimes learn new word-object associations (Duff et al., 2006). Critically, our work extends those earlier findings by demonstrating that patients with hippocampal pathology can simultaneously learn multiple arbitrary, multimodal word-object associations even when potentially interfering information is presented (Roembke and McMurray, 2016). This novel finding addresses a key question regarding the necessity of the hippocampus for CSSL, and it makes contact with several current theories of hippocampal contributions to learning and memory as well as theories of statistical word learning.

## Interindividual Differences in Task Performance Within the Current Patient Group

Our finding that patients with hippocampal pathology acquired new words from CSSL includes certain caveats. Of the three patients, two (1846 and 2363) showed robust learning while the third (2563) did not. This individual difference was not obviously attributable to the degree of memory impairment, etiology, or neuroanatomy. Notably, the patient who did not show evidence of learning, 2563, adopted and later informally described a tactical approach to the task (alternating left-right responses) that did not benefit his performance. Because unsupervised learning paradigms (including our CSSL task) present no feedback, the ineffectiveness of a given tactic may never become evident to the learner. Such tactics have been found to affect performance in both human and non-human animal learning (Wasserman et al., 2015; Roembke and McMurray, 2016). We speculate that 2563's tactic during CSSL may have interfered with the residual capacity for statistical learning shown by the two remaining patients.

This account may also address 2563's excellent performance in the EE task. In that condition, the lack of response selection during the learning phase meant that there was no opportunity to develop or apply a tactical approach. Alternatively, the substantial differences in 2563's performance across conditions could be attributed to an unusual vulnerability to interference, but the two other patients did not exhibit a similar susceptibility. Finally, we note that 2563's poor performance in the CSSL task reported here was qualitatively similar to his poor statistical learning performance in Covington et al. (2018) where he showed less evidence of statistical learning than 1846 and 2363.

Regarding the CSSL exhibited by patients 1846 and 2363, both showed significant evidence of acquiring word-object associations during the task. However, their learning was less rapid and less robust than the comparison group. As with a prior report (Covington et al., 2018), we interpret the learning shown by 1846 and 2363 as a reflection of contributions from a broad network of (non-hippocampal) brain regions to statistical learning performance that has been implied by prior neuroimaging studies (Bischoff-Grethe et al., 2000; McNealy et al., 2006; Turk-Browne et al., 2009; Karuza et al., 2013).

## Word Learning in Patients With Hippocampal Pathology

Patients in our study showed evidence of learning multimodal, auditory-visual word-referent representations. While near-normal recognition performance for single items has been reported for patients with hippocampal amnesia (Ryan et al., 2000; Konkel et al., 2008), residual learning of inter-item relational information is unusual (Giovanello et al., 2003; Mayes et al., 2004; Turriziani et al., 2004; Hannula et al., 2007; Konkel et al., 2008) but not unprecedented, even in the context of word-referent representations. For example, Duff et al. (2006) tested EE of picture-word pairs in a control condition. After 24 exposures, patients showed some learning of picture-word pairs in a cued recall test (mean = 35% correct) although it was much less than that of comparison participants (who were 100% correct after only four exposures). In contrast, Warren and Duff (2014) tested word-referent learning in two conditions (EE and fast mapping (FM)) and observed no evidence of above-chance learning after two exposures. Other studies contrasting EE and FM word learning have also reported little evidence of learning multimodal relational from small numbers of exposures (for review see Cooper et al., 2019).

A key difference between studies of word-referent pairs that observed no learning and those that observed some learning may lie in the number of stimulus presentations. Duff et al. (2006) found evidence of limited but measurable multimodal relational learning after 24 presentations; here, we observed impaired but measurable learning across 42 CSSL presentations per word-object pair; and studies that reported little or no learning typically provided many fewer presentations (e.g., Warren and Duff, 2014). This suggests that the massed practice which characterizes CSSL paradigms may allow slower, non-hippocampal brain systems to learn multimodal, relational information. As with prior studies that demonstrated evidence of inefficient but measurable learning by patients with amnesia (Scoville and Milner, 1957; Cohen and Squire, 1980; Glisky et al., 1986; Glisky, 1992), our finding that multimodal word-referent representations can be learned despite hippocampal pathology suggests an intriguing translational potential for CSSL methods. However, subsequent investigations should also address why laboratory evidence for CSSL does not necessarily generalize to the ecological learning of word-referent pairs by patients with hippocampal pathology.

## Statistical Learning and the Hippocampus

Our findings contribute to a growing literature describing hippocampal contributions to statistical learning. Prior work first suggested that medial temporal lobe and/or hippocampus might make necessary contributions to statistical learning: functional neuroimaging indicated that hippocampal activation can be related to statistical learning (Turk-Browne et al., 2009; Schapiro et al., 2016); and a neuropsychological study indicated that the medial temporal lobe might be necessary for statistical learning (Schapiro et al., 2014). However, more recent work indicates that while the hippocampus may contribute to

statistical learning, learning through statistical exposure is still possible despite hippocampal pathology (Covington et al., 2018) albeit reduced relative to normal performance. Our observations are consistent with the latter account, that is, the hippocampus is not strictly necessary for statistical learning—even when the statistics describe arbitrary relations between elements. Our findings also converge with neuroimaging results from Berens et al. (2018) which indicated that rapid binding of representations in the hippocampus may enhance CSSL in healthy adults.

We suggest that the nature of hippocampal contributions to statistical learning is informed by our finding that patients with hippocampal damage learned less efficiently than healthy comparison participants (see also Covington et al., 2018). Our observations are consistent with a role for the hippocampus in which it can contribute to statistical learning indirectly by supporting the rapid binding of independent and arbitrarily associated pieces of information. This familiar contribution is predicted by relational memory theory (Eichenbaum et al., 1994; Eichenbaum and Cohen, 2001), and it is consistent with our finding that healthy participants showed rapid learning of the arbitrary relations during the task. Absent contributions of the hippocampus, two patients learned some information, but that learning was slower (i.e., less learning from identical exposure; patient 1846) and/or potentially less stable (patient 2363) than healthy comparisons.

To the extent that deficits in relational memory limited performance of patients in the current word-referent learning task, we would predict that tasks that required memory for additional relations (e.g., between more items) would show similar or greater deficits in performance. A reasonable question might be, is it possible that this outcome could also obtain if the nature of the deficit was qualitatively different? If the deficits in CSSL that we observed were (for example) exclusively attributable to a hippocampus-dependent impairment in incremental learning of associations from statistical exposure, would the same outcome be observed? While not impossible, this explanation would not be consistent with substantial prior evidence that patients with hippocampal pathology can often show incremental learning as efficient as that of healthy comparison participants in a variety of laboratory tasks (Milner, 1968; Cohen and Squire, 1980; Duff et al., 2006). An important caveat is that such tasks have typically used explicit, deterministic exposure rather than incidental, statistical exposure. Our approach intentionally replicated typical CSSL methods to align with the existing literature, but future studies might be expressly designed to probe this issue. Testing the nature of the CSSL representations for hallmark features of relational representation (part-cued retrieval, flexibility, etc.) in patients and healthy comparisons would be especially informative.

## Statistical Learning and Non-hippocampal Brain Regions

Although our design was not intended to exhaustively probe patient memory representations, we speculate that patient memory representations would have hallmark

features of non-hippocampal memory including contextual dependence, lack of generalizability, and inflexibility (Cohen and Squire, 1980; Glisky et al., 1986; Duff et al., 2006; Warren et al., 2012). Alternatively, our findings could also be interpreted through the lens of complementary learning systems models (McClelland et al., 1995; Norman and O'Reilly, 2003). Under this interpretation, the availability of enhanced pattern separation and completion supported by the hippocampus may have enhanced the speed of statistical learning for healthy participants by sharpening representations of studied associations (Schapiro et al., 2016). Meanwhile, non-hippocampal MTL (and other brain regions) would support slower learning that is more prone to interference because of relatively poor pattern separation (McCloskey and Cohen, 1989; Norman and O'Reilly, 2003). While we believe that the relational memory account is especially informative, our observations of less efficient learning by patients with hippocampal pathology are consistent with either perspective.

### CSSL and Hippocampus: On-Line Processing of Representations

Encoding durable memory representations is a hallmark of hippocampal function, but the hippocampus is also increasingly understood to contribute to ongoing cognitive processes ("memory at the moment") in ways that may influence CSSL performance. Patients with hippocampal pathology have been shown to perform more poorly than healthy comparisons in a variety of tasks which do not put obvious demands on long-term memory representations such as visual search tasks (Barens et al., 2007; Voss et al., 2011; Warren et al., 2011, 2012).

This is highly relevant to CSSL because it has been hypothesized that two distinct processes comprise CSSL (Roembke and McMurray, 2016; Roembke et al., 2018): (1) a gradual associative process which incrementally updates word-referent weights; and (2) a rapid, real-time inference process employed during referent selection based on the current weightings (McMurray et al., 2012). In this framework, the second process does not reflect learning. Instead, it describes real-time processing which allows participants to combine any available evidence of (statistical) associations with the current context to make a more accurate decision. One effect of this processing may be the temporary amplification of relatively weak mappings to achieve better accuracy in the moment (Yurovsky et al., 2014).

This latter process may benefit from hippocampal-dependent processing of information in the moment. Conversely, degraded hippocampal function could contribute to impairments in the inferential process and impair CSSL performance. Our findings would be consistent with this account. Further still, statistical learning may not be simply based on the observed statistics of the input. Rather, elements of the input that receive more attention may become more strongly associated (McMurray et al., 2012; Yurovsky and Frank, 2015). From this perspective, a contribution of the hippocampus might be to strengthen associations between input elements that were preferentially attended. This

would be consistent with the well-characterized roles of the hippocampus in encoding new relational representations (Eichenbaum and Cohen, 2001) and/or pattern separation (Norman and O'Reilly, 2003).

Targeted experimental designs should assess whether failures of real-time inferential processing uniquely contribute to impaired performance in patients with hippocampal pathology.

### CSSL, Hippocampus, and Language Development

Our findings are also relevant to understanding word learning during language development. We observed that the hippocampus is not necessary for learning of word-referent pairs in a CSSL paradigm. This suggests that an extended network of (non-hippocampal) language-related brain regions could support CSSL in infants and young children (Smith and Yu, 2008; Suanda et al., 2014; Fitneva and Christiansen, 2017; Vlach and DeBrock, 2017; Roembke et al., 2018). Significant word learning occurs before 36 months, a time when the hippocampus and MTL are still developing (Gogtay et al., 2006; Ghetti and Bunge, 2012; Ofen, 2012). Thus, our evidence for non-hippocampal learning suggests that other brain regions may support early developmental language milestones. This is consistent with findings from developmentally amnesic individuals with perinatal hippocampal pathology (Vargha-Khadem et al., 1997; Elward and Vargha-Khadem, 2018), as those individuals showed relatively preserved vocabulary acquisition (low-normal) despite severe deficits in declarative memory. New studies in developmental populations could test the implications of a greater childhood reliance on non-hippocampal learning by comparing the efficiency and quality of CSSL in children and adults.

### LIMITATIONS

Our study had some limitations. First, as with many neuropsychological studies, our sample size of three patients was small. This limitation did not prevent our design from capturing important information from the behavior of our sample. However, it did limit our ability to address certain questions such as a putative relationship between volume of preserved hippocampal tissue and CSSL performance. Second, the MRI exam of the hippocampus for volumetric assessment was not possible for one of the patients, but CT evaluation suggested bilateral hippocampal atrophy. While there was no evidence of atrophy of other brain regions in the patient's CT imaging data, it is possible that his unusual pattern of chance performance on the CSSL task could be attributed to subtle neuroanatomical changes. However, this would not be consistent with his performance on standard neuropsychological tests. Third, our design could not assess the resilience or persistence of new word knowledge, although we speculate that patients would have impaired retention of new word learning over time. Retention (and consolidation) could be addressed in future research by testing learned information again after a delay. Fourth, although the CSSL and EE tasks used unique stimuli and were administered in separate sessions, the order of



administration to the patient group was fixed (CSSL then EE). Meanwhile, the healthy comparison groups completed either the CSSL or EE tasks, but not both. Because healthy comparison performance was perfect in the EE task, we do not believe that they were selectively disadvantaged by the relative novelty of the task. Similarly, it is not clear how prior exposure to a different (CSSL) task would influence the EE task performance of the patient group. However, counterbalancing of task order could be used in future studies to address concerns regarding any potential confound of this nature. Finally, the CSSL task used here was subject to certain design constraints. The number of studied items was limited, the number of competitor items was fixed and small, and the word-referent pairings were deterministic (vs. stochastic). These elements of our design were deliberate and intended to provide sufficient power for our novel investigation of CSSL in patients with hippocampal pathology. Future investigations seeking to extend our findings should parametrically vary design parameters with the goal of refining the field's understanding of hippocampal contributions to CSSL.

## CONCLUSIONS

In conclusion, our findings are consistent with the suggestion that the hippocampus is not strictly necessary for statistical learning (Covington et al., 2018). Rather, the hippocampus may contribute to CSSL by: (1) providing an additional route for faster learning; and/or (2) supporting real-time processing to improve performance at the moment. Critically, this supports accounts of CSSL that include the incremental accumulation of statistics or the gradual building of associations (in addition to more rapid forms of learning or inference; Frank et al., 2009; McMurray et al., 2012).

We speculate that non-hippocampal brain regions or structures that contribute to statistical learning may include medial temporal lobe neocortex (McClelland et al., 1995; Norman and O'Reilly, 2003) and basal ganglia (Poldrack et al., 2001; Poldrack and Packard, 2003; Poldrack and Rodriguez, 2004) among others (Bischoff-Grethe et al., 2000; McNealy et al., 2006; Turk-Browne et al., 2009; Karuza et al., 2013). CSSL may also benefit more specifically from contributions by a network of language-related brain regions including anterior and lateral temporal lobes (McNealy et al., 2006; Davis et al., 2009; Karuza et al., 2013; Warren et al., 2016). Additional

functional neuroimaging and neuropsychological investigations might address this hypothesis.

Importantly, if our findings generalize to other populations with memory deficits due to hippocampal damage or dysfunction (e.g., Alzheimer's disease, medial temporal lobe epilepsy, anti-NMDA receptor encephalitis), then those individuals should be able to learn new word-referent mappings under conditions promoting statistical learning. It remains to be determined whether the durability of information learned in this manner is different from more traditional explicit learning formats, but the translational potential of learning in a simple cross-situational statistical format is exciting. Finally, our work highlights the utility of multidisciplinary studies which combine methods and theoretical perspectives from the literature of language and memory (Duff and Brown-Schmidt, 2012) and the unique capacity of neuropsychological methods to inform the necessity of key brain regions for processes supporting memory, language, or both.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Iowa Human Subjects Office and by the Biomedical Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

All authors designed the research. DW, TR, and NC conducted the research. DW and TR analyzed the data. All authors prepared and approved the manuscript.

## FUNDING

This research was supported by Delta Center Interdisciplinary Research Grant (DW) and National Institute on Deafness and Other Communication Disorders, Grant Nos. R01-DC008089 (BM), and R01-DC011755 (MD).

## REFERENCES

- Allen, J. S., Bruss, J., Brown, C. K., and Damasio, H. (2005). Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. *Neurobiol. Aging* 26, 1245–1260. doi: 10.1016/j.neurobiolaging.2005.05.023
- Allen, J. S., Tranel, D., Bruss, J., and Damasio, H. (2006). Correlations between regional brain volumes and memory performance in amnesia. *J. Clin. Exp. Neuropsychol.* 28, 457–476. doi: 10.1080/13803390590949287
- Baldwin, D., Andersson, A., Saffran, J., and Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition* 106, 1382–1407. doi: 10.1016/j.cognition.2007.07.005
- Barense, M. D., Gaffan, D., and Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia* 45, 2963–2974. doi: 10.1016/j.neuropsychologia.2007.05.023
- Berens, S. C., Horst, J. S., and Bird, C. M. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Curr. Biol.* 28, 1132.e5–1136.e5. doi: 10.1016/j.cub.2018.02.042
- Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., and Berns, G. S. (2000). Conscious and unconscious processing of nonverbal predictability in wernicke's area. *J. Neurosci.* 20, 1975–1981. doi: 10.1523/jneurosci.20-05-01975.2000
- Bloom, L. (1973). *One Word At a Time: The Use of Single Word Utterances Before Syntax*. The Hague: Mouton.

- Clelland, T. A., Johnson, B. A., Leon, M., and Linster, C. (2007). Relational representation in the olfactory system. *Proc. Natl. Acad. Sci. U S A* 104, 1953–1958. doi: 10.1073/pnas.0608564104
- Cohen, N. J., and Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* 210, 207–210. doi: 10.1126/science.7414331
- Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39. doi: 10.1037/0278-7393.31.1.24
- Cooper, E., Greve, A., and Henson, R. N. (2019). Little evidence for Fast Mapping (FM) in adults: a review and discussion. *Cogn. Neurosci.* 10, 196–209. doi: 10.1080/17588928.2018.1542376
- Covington, N. V., Brown-Schmidt, S., and Duff, M. C. (2018). The necessity of the hippocampus for statistical learning. *J. Cogn. Neurosci.* 30, 680–697. doi: 10.1162/jocn\_a\_01228
- Crawford, J. R., and Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: development of a Bayesian approach. *Cogn. Neuropsychol.* 24, 343–372. doi: 10.1080/02643290701290146
- Davachi, L., and Dobbins, I. G. (2008). Declarative memory. *Curr. Dir. Psychol. Sci.* 17, 112–118. doi: 10.1111/j.1467-8721.2008.00559.x
- Davis, M. H., Di Betta, A. M., Macdonald, M. J., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J. Cogn. Neurosci.* 21, 803–820. doi: 10.1162/jocn.2009.21059
- Duff, M. C., and Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Front. Hum. Neurosci.* 6:69. doi: 10.3389/fnhum.2012.00069
- Duff, M. C., Hengst, J., Tranel, D., and Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nat. Neurosci.* 9, 140–146. doi: 10.1038/nn1601
- Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. New York, NY: Oxford University Press.
- Eichenbaum, H., Otto, T., and Cohen, N. J. (1992). The hippocampus—what does it do? *Behav. Neural Biol.* 57, 2–36. doi: 10.1016/0163-1047(92)90724-I
- Eichenbaum, H., Otto, T., and Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behav. Brain Sci.* 17, 449–472. doi: 10.1017/s0140525x00035391
- Elward, R. L., and Vargha-Khadem, F. (2018). Semantic memory in developmental amnesia. *Neurosci. Lett.* 680, 23–30. doi: 10.1016/j.neulet.2018.04.040
- Fitneva, S. A., and Christiansen, M. H. (2017). Developmental changes in cross-situational word learning: the inverse effect of initial accuracy. *Cogn. Sci.* 41, 141–161. doi: 10.1111/cogs.12322
- Fjell, A. M., Westlye, L. T., Grydeland, H., Amlie, I., Espeseth, T., Reinvang, I., et al. (2013). Critical ages in the life course of the adult brain: nonlinear subcortical aging. *Neurobiol. Aging* 34, 2239–2247. doi: 10.1016/j.neurobiolaging.2013.04.006
- Frank, M. C., Goodman, N. D., and Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychol. Sci.* 20, 578–585. doi: 10.1111/j.1467-9280.2009.02335.x
- Frost, R., Armstrong, B. C., and Christiansen, M. H. (2019). Statistical learning research: a critical review and possible new directions. *Psychol. Bull.* 145, 1128–1153. doi: 10.1037/bul0000210
- Gabrieli, J. D., Cohen, N. J., and Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain Cogn.* 7, 157–177. doi: 10.1016/0278-2626(88)90027-9
- Ghetti, S., and Bunge, S. A. (2012). Neural changes underlying the development of episodic memory during middle childhood. *Dev. Cogn. Neurosci.* 2, 381–395. doi: 10.1016/j.dcn.2012.05.002
- Giovanello, K. S., Verfaellie, M., and Keane, M. M. (2003). Disproportionate deficit in associative recognition relative to item recognition in global amnesia. *Cogn. Affect. Behav. Neurosci.* 3, 186–194. doi: 10.3758/cabn.3.3.186
- Glisky, E. L. (1992). Acquisition and transfer of declarative and procedural knowledge by memory-impaired patients: a computer data-entry task. *Neuropsychologia* 30, 899–910. doi: 10.1016/0028-3932(92)90034-j
- Glisky, E. L., Schacter, D. L., and Tulving, E. (1986). Computer learning by memory-impaired patients: acquisition and retention of complex knowledge. *Neuropsychologia* 24, 313–328. doi: 10.1016/0028-3932(86)90017-5
- Gogtay, N., Nugent, T. F., Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., et al. (2006). Dynamic mapping of normal human hippocampal development. *Hippocampus* 16, 664–672. doi: 10.1002/hipo.20193
- Goodglass, H., Kaplan, E., and Weintraub, S. (1983). *Boston Naming Test*. Philadelphia, PA: Lea & Febiger.
- Hannula, D. E., Ryan, J. D., Tranel, D., and Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *J. Cogn. Neurosci.* 19, 1690–1705. doi: 10.1162/jocn.2007.19.10.1690
- Hannula, D. E., Tranel, D., and Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26, 8352–8359. doi: 10.1523/JNEUROSCI.5222-05.2006
- Hummel, J. E., and Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* 110, 220–264. doi: 10.1037/0033-295x.110.2.220
- Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., and Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang.* 127, 46–54. doi: 10.1016/j.bandl.2012.11.007
- Konkel, A., Warren, D. E., Duff, M. C., Tranel, D. N., and Cohen, N. J. (2008). Hippocampal amnesia impairs all manner of relational memory. *Front. Hum. Neurosci.* 2:15. doi: 10.3389/neuro.09.015.2008
- Mayes, A. R., Holdstock, J. S., Isaac, C. L., Montaldi, D., Grigor, J., Gummer, A., et al. (2004). Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus* 14, 763–784. doi: 10.1002/hipo.10211
- 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
- McCloskey, M., and Cohen, N. J. (1989). “Catastrophic interference in connectionist networks: the sequential learning problem,” in *Psychology of Learning and Motivation*, (Vol. 24) ed. G. H. Bower (New York, NY: Academic Press), 109–165.
- McMurray, B., Horst, J. S., and Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychol. Rev.* 119, 831–877. doi: 10.1037/a0029872
- McNealy, K., Mazziotta, J. C., and Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* 26, 7629–7639. doi: 10.1523/JNEUROSCI.5501-05.2006
- Milner, B. (1968). Disorders of memory after brain lesions in man. *Neuropsychologia* 6, 175–179. doi: 10.1016/0028-3932(68)90017-1
- Norman, K. A., and O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* 110, 611–646. doi: 10.1037/0033-295x.110.4.611
- Ofen, N. (2012). The development of neural correlates for memory formation. *Neurosci. Biobehav. Rev.* 36, 1708–1717. doi: 10.1016/j.neubiorev.2012.02.016
- Osterrieth, P. A. (1944). Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. *Arch. Psychol.* 30, 206–356.
- Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature* 414, 546–550. doi: 10.1038/35107080
- Poldrack, R. A., and Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia* 41, 245–251. doi: 10.1016/s0028-3932(02)00157-4
- Poldrack, R. A., and Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiol. Learn. Mem.* 82, 324–332. doi: 10.1016/j.nlm.2004.05.003
- Postle, B. R., and Corkin, S. (1998). Impaired word-stem completion priming but intact perceptual identification priming with novel words: evidence from

- the amnesic patient H.M. *Neuropsychologia* 36, 421–440. doi: 10.1016/s0028-3932(97)00155-3
- Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus* 20, 1263–1290. doi: 10.1002/hipo.20852
- Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K. M., Williamson, A., and Acker, J. D. (2004). Aging, sexual dimorphism and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. *Neurobiol. Aging* 25, 377–396. doi: 10.1016/s0197-4580(03)00118-0
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique (*Les problèmes*). *Arch. Psychol.* 28, 215–285.
- Roembke, T. C., and McMurray, B. (2016). Observational word learning: beyond propose-but-verify and associative bean counting. *J. Mem. Lang.* 87, 105–127. doi: 10.1016/j.jml.2015.09.005
- Roembke, T. C., Wiggs, K. K., and McMurray, B. (2018). Symbolic flexibility during unsupervised word learning in children and adults. *J. Exp. Child Psychol.* 175, 17–36. doi: 10.1016/j.jecp.2018.05.016
- Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychol. Sci.* 11, 454–461. doi: 10.1111/1467-9280.00288
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., and Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* 26, 1736–1747. doi: 10.1162/jocn\_a\_00578
- Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., and Norman, K. A. (2016). Complementary learning systems within the hippocampus: a neural network modeling approach to reconciling episodic memory with statistical learning *bioRxiv* 051870v1. [Preprint]. doi: 10.1101/051870
- Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., and Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus* 26, 3–8. doi: 10.1002/hipo.22523
- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21. doi: 10.1136/jnnp.20.1.11
- Sharon, T., Moscovitch, M., and Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proc. Natl. Acad. Sci. U S A* 108, 1146–1151. doi: 10.1073/pnas.1005238108
- Smith, C. N., Urgolites, Z. J., Hopkins, R. O., and Squire, L. R. (2014). Comparison of explicit and incidental learning strategies in memory-impaired patients. *Proc. Natl. Acad. Sci. U S A* 111, 475–479. doi: 10.1073/pnas.1322263111
- Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010
- Suanda, S. H., Mugwanya, N., and Namy, L. L. (2014). Cross-situational statistical word learning in young children. *J. Exp. Child Psychol.* 126, 395–411. doi: 10.1016/j.jecp.2014.06.003
- The Psychological Corporation. (1997). *WAIS-III and WMS-III Technical Manual*. San Antonio, TX: The Psychological Corporation.
- Trueswell, J. C., Medina, T. N., Hafri, A., and Gleitman, L. R. (2013). Propose but verify: fast mapping meets cross-situational word learning. *Cogn. Psychol.* 66, 126–156. doi: 10.1016/j.cogpsych.2012.10.001
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., and Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* 21, 1934–1945. doi: 10.1162/jocn.2009.21131
- Turriziani, P., Fadda, L., Caltagirone, C., and Carlesimo, G. A. (2004). Recognition memory for single items and for associations in amnesic patients. *Neuropsychologia* 42, 426–433. doi: 10.1016/j.neuropsychologia.2003.10.003
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277, 376–380. doi: 10.1126/science.277.5324.376
- Vlach, H. A., and DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory and language abilities. *J. Mem. Lang.* 93, 217–230. doi: 10.1016/j.jml.2016.10.001
- Voss, J. L., Warren, D. E., Gonsalves, B. D., Federmeier, K. D., Tranel, D., and Cohen, N. J. (2011). Spontaneous revisitation during visual exploration as a link among strategic behavior, learning and the hippocampus. *Proc. Natl. Acad. Sci. U S A* 108, E402–E409. doi: 10.1073/pnas.1100225108
- Warren, D. E., and Duff, M. C. (2014). Not so fast: hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus* 24, 920–933. doi: 10.1002/hipo.22279
- Warren, D. E., Duff, M. C., Jensen, U., Tranel, D., and Cohen, N. J. (2012). Hiding in plain view: lesions of the medial temporal lobe impair online representation. *Hippocampus* 22, 1577–1588. doi: 10.1002/hipo.21000
- Warren, D. E., Duff, M. C., Magnotta, V., Capizzano, A. A., Cassell, M. D., and Tranel, D. (2012). Long-term neuropsychological, neuroanatomical and life outcome in hippocampal amnesia. *Clin. Neuropsychol.* 26, 335–369. doi: 10.1080/13854046.2012.655781
- Warren, D. E., Duff, M. C., Tranel, D., and Cohen, N. J. (2011). Observing degradation of visual representations over short intervals when medial temporal lobe is damaged. *J. Cogn. Neurosci.* 23, 3862–3873. doi: 10.1162/jocn\_a\_00089
- Warren, D. E., Tranel, D., and Duff, M. C. (2016). Impaired acquisition of new words after left temporal lobectomy despite normal fast-mapping behavior. *Neuropsychologia* 80, 165–175. doi: 10.1016/j.neuropsychologia.2015.11.016
- Wasserman, E. A., Brooks, D. I., and McMurray, B. (2015). Pigeons acquire multiple categories in parallel via associative learning: a parallel to human word learning? *Cognition* 136, 99–122. doi: 10.1016/j.cognition.2014.11.020
- Weschler, D. (1997). *Weschler Adult Intelligence Scale*. San Antonio, TX: The Psychological Corporation.
- Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x
- Yurovsky, D., and Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition* 145, 53–62. doi: 10.1016/j.cognition.2015.07.013
- Yurovsky, D., Fricker, D. C., Yu, C., and Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychon. Bull. Rev.* 21, 1–22. doi: 10.3758/s13423-013-0443-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 © 2020 Warren, Roembke, Covington, McMurray and Duff. 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.



# Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship

Melissa C. Duff<sup>1\*</sup>, Natalie V. Covington<sup>1</sup>, Caitlin Hilverman<sup>1</sup> and Neal J. Cohen<sup>2</sup>

<sup>1</sup>Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, United States,

<sup>2</sup>Department of Psychology, Beckman Institute, University of Illinois, Champaign, IL, United States

## OPEN ACCESS

### Edited by:

Tom Verguts,  
Ghent University, Belgium

### Reviewed by:

Thanujeni Pathman,  
York University, Canada  
Christine Bastin,  
University of Liège, Belgium

### \*Correspondence:

Melissa C. Duff  
melissa.c.duff@vumc.org

### Specialty section:

This article was submitted to  
Cognitive Neuroscience, a section of  
the journal *Frontiers in Human  
Neuroscience*

**Received:** 01 August 2019

**Accepted:** 23 December 2019

**Published:** 24 January 2020

### Citation:

Duff MC, Covington NV, Hilverman C  
and Cohen NJ (2020) Semantic  
Memory and the Hippocampus:  
Revisiting, Reaffirming, and Extending  
the Reach of Their  
Critical Relationship.  
*Front. Hum. Neurosci.* 13:471.  
doi: 10.3389/fnhum.2019.00471

Since Tulving proposed a distinction in memory between semantic and episodic memory, considerable effort has been directed towards understanding their similar and unique features. Of particular interest has been the extent to which semantic and episodic memory have a shared dependence on the hippocampus. In contrast to the definitive evidence for the link between hippocampus and episodic memory, the role of the hippocampus in semantic memory has been a topic of considerable debate. This debate stems, in part, from highly variable reports of new semantic memory learning in amnesia ranging from profound impairment to full preservation, and various degrees of deficit and ability in between. More recently, a number of significant advances in experimental methods have occurred, alongside new provocative data on the role of the hippocampus in semantic memory, making this an ideal moment to revisit this debate, to re-evaluate data, methods, and theories, and to synthesize new findings. In line with these advances, this review has two primary goals. First, we provide a historical lens with which to reevaluate and contextualize the literature on semantic memory and the hippocampus. The second goal of this review is to provide a synthesis of new findings on the role of the hippocampus and semantic memory. With the perspective of time and this critical review, we arrive at the interpretation that the hippocampus does indeed make necessary contributions to semantic memory. We argue that semantic memory, like episodic memory, is a highly flexible, (re)constructive, relational and multimodal system, and that there is value in developing methods and materials that fully capture this depth and richness to facilitate comparisons to episodic memory. Such efforts will be critical in addressing questions regarding the cognitive and neural (inter)dependencies among forms of memory, and the role that these forms of memory play in support of cognition more broadly. Such efforts also promise to advance our understanding of how words, concepts, and meaning, as well as episodes and events, are instantiated and maintained in memory and will yield new insights into our two most quintessentially human abilities: memory and language.

**Keywords:** semantic, episodic, memory, language, hippocampus, methods



## INTRODUCTION

Nearly 50 years ago, Tulving (1972) proposed that memory research may benefit from observing a distinction between episodic and semantic memory. In distinguishing episodic and semantic memory, Tulving stated that episodic memory referred to knowledge “*about temporally dated episodes or events, and temporal-spatial relations among these events*” and noted that such memory is stored “*in terms of its autobiographical reference to the already existing contents of the episodic memory store*” (Tulving, 1972, p. 385). Semantic memory was defined as the “*memory necessary for the use of language. It is a mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning, and referents, about relations among them, and about the rules, formulas, and algorithms for the manipulation of these symbols, concepts, and relations*” (Tulving, 1972, p. 386). This distinction was offered by Tulving as something of a thought experiment, one that he proposed might have utility in understanding, and accounting for, the broader range of memory phenomena and experimental findings of the time. Indeed, Tulving stated, “*I will refer to both kinds of memory as two stores or as two systems, but I do this primarily for the convenience of communication, rather than as an expression of any profound belief about the structural or functional separation of the two. Nothing very much is lost at this stage of our deliberations if the reality of the separation lies solely in the experimenter’s and the theorist’s, and not the subject’s mind*” (Tulving, 1972, p. 384).

Despite Tulving’s own ambivalence, at least in his early writings, about the reality of the distinction between episodic and semantic memory, this distinction has persisted and has formed the foundation for decades of theoretical and experimental work in the cognitive neuroscience of memory. Considerable effort has been directed towards understanding the similar and unique features of episodic and semantic memory as part of a broader effort to characterize the neurobiology of memory, its functional divisions, and neuroanatomical correlates (e.g., Cohen and Squire, 1980; Squire, 1992; Tulving and Markowitsch, 1998; Thompson-Schill, 2003; Ryan et al., 2008; Greenberg and Verfaellie, 2010; Henke, 2010; Ranganath, 2010; Hannula and Duff, 2017). A key finding, and area of broad consensus, is that the hippocampus, and surrounding medial temporal lobe (MTL) structures, play a critical role in the encoding and subsequent retrieval of new long-term episodic memories (Cohen, 1984; Squire, 1992; Cohen and Eichenbaum, 1993; Gabrieli, 1998; Davachi, 2006; Eichenbaum et al., 2007; Rugg et al., 2015). A key source of evidence for the link between episodic memory and the hippocampus came from studies of patients with hippocampal damage who had profound deficits in acquiring new information about their daily lives and experiences (Scoville and Milner, 1957; Damasio et al., 1989; Corkin, 2002; Rosenbaum et al., 2005). This observed deficit was in contrast to the seemingly preserved ability of these patients to recount episodes from the remote past (or at least relative to events experienced since the onset of amnesia) and the ability to acquire new skills and habits (non-declarative, or procedural, memory).

But, what was the status of semantic memory? Was semantic memory, like episodic memory, also critically dependent on the hippocampus? And, given hippocampal damage, are deficits in episodic and semantic memory observed in tandem? This was a central question in the field. One prominent proposal was that semantic and episodic memory comprise, or depend upon, a unitary memory system, the declarative memory system, and that hippocampal damage would yield similar deficits (Cohen, 1984; Cohen and Eichenbaum, 1993; Squire and Zola, 1996; Cohen et al., 1997; Eichenbaum, 1998). An alternative proposal was that episodic and semantic memory formation was independent and could be acquired or damaged in isolation (Kinsbourne and Wood, 1975). Yet another proposal suggested that all memories start out as episodic and that over time some become semantic through processes of semantization or decontextualization (i.e., whereas episodic memories are bound to temporal and spatial contexts, the absence or loss of this specific context makes such memories semantic; for discussion and review, see Meeter and Murre, 2004).

In contrast to the clear and definitive evidence for the link between the hippocampus and episodic memory, the role of the hippocampus in semantic memory has been a topic of considerable debate. This debate stems from highly variable results from studies of new semantic memory learning in amnesia (as measured by different groups, in different patient populations, with different paradigms) ranging from reports of profound impairment to full preservation, and various degrees of deficit and ability in between. While interest in the (in)dependence of semantic memory and the hippocampus remained high, as evidenced by a number of reviews and commentaries (e.g., Mishkin et al., 1998; Squire and Zola, 1998; Manns et al., 2003; Manns, 2004; Moscovitch et al., 2006), research over the intervening decades did not produce sufficient data to form a core of consistent findings that could definitively adjudicate between competing views or resolve the debate.

More recently, a number of significant advances in the field have occurred, resulting in new provocative data on a robust role of the hippocampus in semantic memory. Thus, this is an ideal moment to revisit this debate, to re-evaluate the data and methods that informed traditional views on this topic, and to synthesize new findings. In line with these advances, our review has two primary goals. First, we provide a historical lens with which to evaluate, update, and contextualize the literature on semantic memory and the hippocampus. In doing so, we look back on this body of work and note a shift in the framing of the research questions, hypotheses, and levels of evidence that altered the trajectory of this line of research away from the original question on the extent to which semantic and episodic memory depends on the hippocampus in parallel and instead moved towards studies on new semantic learning in amnesia largely in isolation from episodic memory. While this “hypothesis drift” was likely unintentional, it seems to have gone unnoticed or at least not discussed in the literature. One consequence is that more recent researchers have inferred an answer to the original question (do episodic and semantic memory have shared dependence on the hippocampus?) based on evidence that was generated in response to the new reframed (drifted)

question (can any new semantic learning be accomplished in amnesia?). We note that during this same time period, the number of investigations into the role of the hippocampus in episodic memory grew exponentially relative to those on semantic memory, based on powerful methods and techniques capable of measuring and quantifying episodic memory, and its perceptual, temporal, and spatial richness. Likewise, advances in theoretical proposals for understanding the nature and function of episodic memory have outpaced those related to semantic memory. We conclude that as time passed, researchers not only moved further away from the question originally posed about (in)dependence of episodic and semantic memory vis a vis the hippocampus but were also increasingly ill-equipped (methodologically and theoretically) to address it. The second goal of this review is to provide a critical reporting and synthesis of new findings on the role of the hippocampus in semantic memory. These advances have significant implications for understanding the role the hippocampus may play in the various stages of acquisition, maintenance, activation, and use of semantic memory in processing, paralleling what we have learned about the role of the hippocampus in the acquisition, maintenance, activation, and use of episodic memory.

We will argue here that the hippocampus is critical to both episodic and semantic memory. With the theoretical and empirical advances in the study of semantic memory and its neural bases, we can see that the depth and richness of semantic compare favorably to that of episodic memory and that they are both highly flexible, (re)constructive, relational and multimodal systems reliant upon the properties of the hippocampus. Such advances promise to illuminate our understanding of how words, concepts, and meaning are instantiated and maintained in memory, and then activated and used on-demand, just as well as, and in the same ways as are episodes and events.

Before we begin, we should acknowledge that our focus in this review is on semantic memory and that our approach is from the specific vantage point of the debate in the cognitive neuroscience literature on the extent to which semantic and episodic memory depends in tandem on the hippocampus. We place special emphasis on work with neurological patients as it has figured prominently in the history of this literature and it speaks to issues of necessity. Thus, our review does not cover semantic theory or its history (e.g., Grice, Locke, Searle) and we do not review the neuroimaging literature on semantic memory (although see these reviews: Martin and Chao, 2001; Thompson-Schill, 2003; Binder et al., 2009; Binder and Desai, 2011). Our review also places a special focus on the hippocampus. While the cortices of the MTL (e.g., perirhinal, parahippocampal, entorhinal) have been shown to contribute to episodic and semantic memory (e.g., Davachi et al., 2003; Davies et al., 2004; Eichenbaum et al., 2007; Clarke and Tyler, 2015), it is the shared, and often focal, hippocampal damage across patient studies that offer the most compelling evidence for the role of the hippocampus in semantic memory.

We start this review by reexamining and providing a critical context for the historical literature on the ability of individuals with hippocampal damage and resulting anterograde amnesia to acquire new semantic memory and on the integrity of their remote semantic memory, and show how it directly connects

to current understanding of the role of the hippocampus in episodic memory.

## NEW SEMANTIC LEARNING AND REMOTE SEMANTIC MEMORY IN AMNESIA

### New Semantic Learning in Amnesia

The neuropsychological and neuroanatomical description of the seminal case of HM provided significant insight into the organization and neural correlates of human memory (Scoville and Milner, 1957; Corkin, 2002). It also provided the early testing ground for the question of whether hippocampal damage produced commensurate deficits in episodic and semantic memory. The emphasis of this work was on new learning. Empirical testing and behavioral observation revealed that HM had a profound deficit in the encoding and subsequent retrieval of new episodic memory while his ability to recall and recount detailed events and experiences from his remote past appeared intact. It also appeared that HM's remote semantic memory was intact. He did not present with aphasia, was able to name objects, hold conversations, and answer questions about remote facts and knowledge acquired long before the onset of his amnesia. The open question then was whether the deficit in acquiring new semantic memory mirrored his deficit in acquiring new episodic memory.

Before examining this literature, it is important to consider what the shared dependence of episodic memory and semantic memory on the hippocampus might look like. Because this review largely focuses on the abilities and deficits of patients with hippocampal amnesia, let's consider various outcomes and standards for evaluating the data. One standard for confirming that episodic and semantic memory depend on the hippocampus in tandem might be to require equivalent levels of performance, ability, or deficit, in both episodic and semantic memory, in patients with amnesia. Another standard might be to require impairment in both systems but accept variable degrees of a deficit. In contrast, if the two systems are independent, then one might expect a dissociation, with impaired ability in one area and preserved ability in the other. Irrespective of the standard applied, addressing this question has proven difficult due to challenges in equating task demands and characteristics of to-be-learned stimuli across memory systems, and in quantifying lesion extent and residual abilities across patients with amnesia. Thus, a more common approach has been to examine the ability of patients with hippocampal amnesia to learn new semantic information and compare their performance to healthy comparison participants (to establish the existence of a deficit), and then to compare (often in relative rather than quantitative terms) the magnitude of these deficits across systems. Here, one standard might be to require that patients with hippocampal amnesia and healthy comparison participants perform similarly on all aspects of semantic learning (i.e., amount of information acquired, learning rate, generalization). Another standard might be to accept any level of patient learning even if it differs significantly from what healthy individuals are able to acquire, so long as this learning seems different or better than patients' episodic memory ability. As, we will see below,

each of these approaches has yielded variable levels of evidence and different groups have applied different standards that have shifted over time.

### New Semantic Learning in Amnesia: None, or at Least Not Much

Gabrieli et al. (1988) were among the first to examine new semantic memory in HM. They tested the ability of HM and seven healthy comparison participants to acquire the meanings and synonyms of eight low-frequency words (e.g., *quotidian*, *manumit*, *hegira*) under formal laboratory conditions (i.e., each word presented individually with its definition, participants read each word and definition aloud). Knowledge was tested without asking for recall or recognition of any explicit, episodic aspect of prior experience with the words. Gabrieli et al. (1988) reported that HM did not learn any of the new words, or their synonyms, failing to ever reach criterion with experimental sessions aborted after 20 trials. In contrast, controls rapidly acquired the meanings of the new words and their synonyms, and were able to generalize these word meanings to new semantic contexts (e.g., in a sentence). While controls reached criterion in 7.3 trials, on average, it was estimated that HM would have required 335 trials to do so. That HM failed to learn the meaning of even a single word was taken as strong evidence for a profound impairment in semantic memory. The authors reported that “*HM could not learn, in a laboratory setting, the meaning of any word that he did not already know*” (Gabrieli et al., 1988, p. 161). The interpretation was that the impairment in new semantic learning was so severe, it seemed commensurate with that seen in the episodic domain; therefore, both episodic and semantic memory appeared to depend in common upon the hippocampus.

Future studies provided additional evidence for a deficit in learning new semantic information in HM (Postle and Corkin, 1998) and studies with other MTL patients provided more evidence that patients with amnesia were impaired on both semantic and episodic memory to a similar degree (Hamann and Squire, 1995). Hamann and Squire asked a group of amnesic patients to learn new facts (40 three-word sentences such as “MEDICINE cured HICCUP”) and tested their knowledge by presenting them with a sentence fragment to complete (e.g., MEDICINE cured \_\_\_\_\_). The amnesic patients learned at an abnormally slow rate (progressing from 0% to 19% correct vs. better than 75% for controls) and acquired a few exemplars relative to controls. Patient EP, a severely amnesic patient who is reported to have no detectable episodic memory, participated in this study. Like HM in the Gabrieli et al. (1988) study, EP exhibited no semantic learning at all. In recounting these data later, Squire and Zola (1998) commented that “*in a patient with no detectable capacity for episodic memory, there was also no detectable capacity for acquiring semantic knowledge*” (p. 208). Studies like these provided strong evidence for a deficit in new semantic learning in hippocampal amnesia, suggesting commensurate deficits in semantic and episodic memory and providing support for their shared dependence on the hippocampus for normal functioning. As, we will see below, however, the emphasis researchers (including ourselves) placed

on *zero* semantic learning and *no detectable capacity* for semantic memory likely shifted the null hypothesis for subsequent studies.

### New Semantic Learning in Amnesia: Some

Numerous groups have now shown that under some conditions, individuals with amnesia can acquire some new semantic memory. The majority of these studies used tasks and manipulations that attempted to promote new learning by reducing errors or interference (e.g., prevent incorrect information from interfering with recall of correct information; Glisky, 1992), and increasing the meaningfulness (e.g., embedding word lists in high-imagery narratives; Kovner et al., 1983) or semantic relatedness (e.g., table-chair; Shimamura and Squire, 1984) of the to-be-learned stimuli rather than traditional learning (study-test) methods. An approach popularized by Glisky et al. (1986) was to teach new semantic information to memory-impaired individuals using a technique called vanishing cues, a learning strategy under the umbrella approach of errorless learning. The general motivation for using errorless learning strategies to teach new information to individuals with memory impairment came from a growing body of work showing more success in approaches that compensate for specific memory problems compared to those aimed at restoring memory ability (Wilson and Moffat, 1983). Glisky et al. (1986) taught amnesic patients to associate computer terminology (e.g., save, run, boot) with their definitions. Consistent with the premise of reducing opportunities for patients to make errors, when patients could correctly produce the correct answer following a particular cue, they were then trained to respond to reduced cues (cues with letters removed). If participants made an error, letters were added to the cues until correct answers were remembered. In the Glisky et al.’s (1986) study, this technique was successful in teaching four patients with severe amnesia to learn some new computer vocabulary. Using similar learning strategies, patients with hippocampal amnesia can acquire some new semantic information (e.g., Tulving et al., 1991; Gordon Hayman et al., 1993; Baddeley and Wilson, 1994; Bayley and Squire, 2002; Skotko et al., 2004; Stark et al., 2005; Dewar et al., 2009; Hilverman et al., 2016). Across all of these studies, however, irrespective of method or technique, while the patients with amnesia do show some new learning, the learning is impaired and performance is far below what healthy participants can or would be expected to achieve. Patients with hippocampal amnesia acquire only a fraction of what controls learn, their rate of learning is abnormally slow [e.g., in Bayley and Squire (2002) a patient required 48 trials instead of the four trials required by controls], and, unless variability is built into the training procedure, the information they acquire is often rigid and inflexible (Stark et al., 2005).

Building on previous studies showing evidence for some new semantic learning in hippocampal amnesia, O’Kane et al. (2004) returned to HM, who is considered the gold standard case of amnesia as he was the first and most extensively studied case of amnesia in the literature. O’Kane et al. (2004) tested HM on his incidental learning of the names of individuals who had become famous after the onset of his amnesia using a



2-alternative forced-choice (AFC) recognition of famous names design and free recall of associated semantic information. They noted that, “*Until recently, it seemed unlikely that any semantic knowledge could be acquired following extensive bilateral damage to the MTLs. . . and stated that “whether the hippocampus proper is necessary for all semantic learning, or whether some degree of semantic learning can occur in the absence of a functioning hippocampus”* was an open question (O’Kane et al., 2004, p. 417). HM’s performance on the task was above zero indicating he had acquired new semantic memory since the onset of his amnesia. But, this learning was certainly not normal or in line with the performance of healthy participants. HM generated semantic knowledge about only a fraction of the famous people known to the comparison participants and what knowledge he had was sparse and highly variable and inconsistent, particularly relative to his knowledge of pre-morbidly acquired famous people (e.g., HM might correctly identify someone as famous but not know their sex). The conclusion was that, “*Although HM’s semantic learning was clearly impaired, the results provide robust, unambiguous evidence that some new semantic learning can be supported by structures beyond the hippocampus proper*” (O’Kane et al., 2004, p. 417).

The study by O’Kane et al. (2004) represents, and is reflective of, a significant turning point in the literature. Looking back on this literature, the findings of, and emphasis on, zero learning or floor level performance on tests of new semantic learning in amnesia by Gabrieli et al. (1988) and Hamann and Squire (1995) likely resulted in “hypothesis drift.” We borrow the term hypothesis drift from Nadel (1991) to reference the phenomena of recasting the hypothesis to accommodate new, often contradictory, data. We can see this drift represented in how O’Kane et al. (2004) framed the question for their study. Whereas the earlier studies were asking if episodic and semantic memory each had a dependence on the hippocampus, O’Kane et al. (2004) were asking a different question: Can any new semantic learning be accomplished in amnesia and can semantic learning occur independent of the hippocampus? This hypothesis drift was likely unintentional and went largely unnoticed, such that the bar for demonstrating new learning remained the same, despite the change in the research question. As a result of earlier studies with HM and EP, the bar for demonstrating “new learning” was set so low that any performance better than zero would be noteworthy.

Indeed, taken together with the growing body of studies documenting *some* new semantic learning in amnesia, HM’s “*clearly impaired*” learning was interpreted as a viable challenge to the notion of commensurate deficits in episodic and semantic memory in amnesia and as evidence for the independence of semantic memory from the hippocampus. Some authors even argued that the semantic learning observed in amnesia was “*partially or perhaps even wholly preserved*” although the experiments contained no control group or direct comparison to experimental episodic memory performance (Tulving et al., 1991, p. 614).

These studies reflect another, perhaps more subtle, drift in framing the hypothesis: that the hippocampus alone supports semantic memory. Returning to the original

proposals on the shared dependence of episodic and semantic memory on the hippocampus, the hypothesis was never that dependence on the hippocampus was exclusive, just that it was necessary (e.g., Squire and Zola, 1996; Cohen et al., 1997; Baddeley et al., 2001).

In our view, given the similarities between semantic and episodic memory representations (e.g., both require relational binding of multimodal information, expressed flexibly in novel contexts), a shared dependence on the hippocampus across memory systems makes intuitive sense. Further, just as we have come to understand that the full capacity and expression of episodic memory depends critically on a network of brain structures, including but not limited to the hippocampus (e.g., Buckner and Carroll, 2007; Ritchey et al., 2015; Wang et al., 2015; Moscovitch et al., 2016), so too, semantic memory, in its full capacity, relies on a network that includes, but goes beyond the hippocampus (Binder et al., 2009; Binder and Desai, 2011). In fact, there is considerable neuroanatomical overlap in the semantic network and the default-mode network, which supports episodic memory (Binder and Desai, 2011; Irish et al., 2016; Renoult et al., 2019).

A common interpretation across studies of new semantic learning in amnesia was that, even if fully normal semantic learning could not be obtained in the presence of hippocampal damage, some degree of semantic learning could be supported by structures beyond the hippocampus, specifically those associated with the non-declarative memory system. The connection between the limited semantic memory ability in adults with amnesia and their preserved non-declarative memory ability fits well with the properties of the non-declarative memory system (e.g., slow, inflexible, experience-dependent; Reber et al., 1996). Furthermore, a role for non-declarative memory processes in semantic memory acquisition, in concert with hippocampal-dependent memory processes, also fits well with its proposed role in normal word learning ability in healthy individuals (Davis and Gaskell, 2009; Gupta, 2012). Viewed from the perspective that non-declarative memory processes are part of normal word learning, it becomes less surprising that such processes are used to support semantic learning in amnesia and more striking how impoverished and difficult new semantic learning is without the contribution of the hippocampus.

Acknowledging the hypothesis drift and reframing of the research questions that occurred in the literature, and its impact is important for several reasons. To our knowledge, there has been no explicit discussion of it in the literature. One consequence is that readers and researchers alike have inferred an answer to the original question (do episodic and semantic memory have shared dependence on the hippocampus?) based on evidence that was generated in response to the new reframed (drifted) question (can any new semantic learning be accomplished in amnesia?). As we will discuss in more detail below, this hypothesis drift likely changed the types of data, and levels of evidence, that have accumulated over the intervening decades. We propose this has had cascading effects on the direction the field has gone and the pace of theoretical and methodological development in the area of semantic memory.



## New Semantic Learning: Normal, or at Least a Lot, but...

Several groups have now reported normal semantic memory in the context of severe deficits in episodic memory (e.g., Sharon et al., 2011; but, see Warren and Duff, 2014, 2019; Elward et al., 2019). The work on semantic learning in patients with developmental amnesia by Vargha-Khadem et al. (1997) is the most highly cited on this topic and is considered the most compelling evidence for the dissociation in new learning of episodic and semantic memory in the literature (Vargha-Khadem et al., 1997). They reported on three cases of developmental amnesia, individuals who sustained selective hippocampal damage early in life; at birth for one case, and at ages 4 and 9 in the other two cases. At the time of the report, these three individuals were in their teens and early twenties. Neuroimaging assessment revealed hippocampal volumes between 43% and 61% of the mean values of a healthy comparison group but showed surrounding MTL cortices to be unaffected. It is important to note that while the neuroimaging assessment indicates that there is still residual hippocampal tissue present, it has been suggested that a reduction in hippocampal volume of approximately 40% likely represents a near-complete loss of hippocampal neurons (Gold and Squire, 2005). Neuropsychological data showed severe deficits in episodic memory across a battery of tests (e.g., the logical memory and visual memory subtests of the Wechsler Memory Scales (WMS), Children's Auditory Verbal Learning Test (AVLT), Rey-Osterreith Complex Figures Test). The participants also displayed significant difficulty with episodic memory in their day-to-day lives. Yet, despite these severe episodic memory deficits, these three individuals acquired language, semantic knowledge and factual information that placed them in the low-average to average range on standardized assessments, and were able to attend mainstream school. The authors concluded that developmental amnesia “*produces a severe loss of episodic memory but leaves general cognitive development, based mainly on semantic memory functions, relatively intact*” (Vargha-Khadem et al., 1997, p. 376). Furthermore, given the level of semantic learning achieved in the context of significant episodic memory deficits and hippocampal pathology, the authors argued that normal levels of semantic learning can be achieved independent of the hippocampus. These data were remarkable on many levels. Prior to this publication, the prediction was that early hippocampal pathology would produce widespread and devastating cognitive and intellectual deficits. The amount of semantic learning acquired in these cases far exceeded what was predicted. Furthermore, the level of semantic memory acquired in developmental amnesia seemed strikingly superior to that achieved in adult cases.

There are well-acknowledged challenges in comparing data from developmental and adult-onset populations (Squire and Zola, 1998; Elward and Vargha-Khadem, 2018). One critique of the developmental amnesia work has been that semantic memory was not tested as directly, or formally, in laboratory settings, as was episodic memory, in contrast, for example, in the way it was tested in patient HM (Gabrieli et al., 1988). This makes it difficult to compare quantitative measures of performance

on standardized tests of episodic memory (where individuals encode and recall newly acquired information in the same testing session) with extensive, repeated real-world exposure to semantic memory across time and naturalistic contexts. However, note that standardized episodic and standardized semantic memory tests are not well equated either. Episodic tests (e.g., AVLT, WMS) examine what an individual acquires in the testing session and semantic tests (picture vocabulary tests like the Boston Naming Test or Pyramids and Palm Trees Test) examine vocabulary and semantic knowledge acquired and reinforced over a lifetime.

There are now more formal, laboratory studies of new semantic learning in cases of developmental amnesia in the literature (Elward and Vargha-Khadem, 2018). When examined using laboratory tasks that more closely mirror those used in the adult-onset literature, the pattern of deficit in developmental amnesia seems remarkably similar to the adult-onset cases: the learning rate is slower (Gardiner et al., 2008; Elward and Vargha-Khadem, 2018), less information is acquired (Baddeley et al., 2001) and there is less evidence of generalization relative to controls. The learning deficit is most striking in tasks that require rapid learning and free recall, supporting the notion that the hippocampus is critical for rapid and efficient semantic learning, whereas performance is significantly better, or even similar to controls, when additional learning trials are provided and when learning is measured with recognition or cued recall (Elward and Vargha-Khadem, 2018). Additional evidence for a semantic memory deficit in developmental amnesia comes from Blumenthal et al. (2017) who asked a patient to generate semantic features for object concepts. They reported abnormal patterns of feature generation and typicality ratings in the patient with developmental amnesia relative to controls. The authors attributed these semantic memory deficits to impairments in hippocampal binding mechanisms and suggested that the dissociation between semantic and episodic memory in developmental amnesia may not be as complete as previously conceptualized (Blumenthal et al., 2017).

Duff et al. (2006) have also reported an intact rate of learning for semantic information in adults with hippocampal amnesia. In their study, four patients with hippocampal amnesia completed a referential communication task with a familiar partner (spouse, friend). The patients sat across from their partner and each had a board with 12 numbered spaces and a set of 12 cards displaying Chinese tangrams (i.e., abstract black and white figures with no established names but which could be perceived as people, animals, or objects). A low barrier was between them preventing a view of each others' cards but allowing them to see each other's facial expressions and gestures. The patients with amnesia were the directors and communicated to their partner (always the matcher) how to complete the board with the cards so that at the end of a trial the two boards looked alike (i.e., their cards were in the same numbered spaces on each board). The task was presented as a game and pairs were instructed to communicate freely and have fun. Despite severe episodic memory impairments, the amnesic participants developed and used unique labels for the cards. Across trials, these labels became increasingly concise

and simplified. Most strikingly, the rate of learning exhibited by amnesic participants, measured by the reduction in time and words necessary to complete each trial, did not differ from that of healthy participants. The long-term retention of this new learning at 30 min, 6 months, and even 2 years for one participant did not differ between groups. These results were the first to show an intact rate of new semantic learning in adult-onset amnesia in a social-communicative learning paradigm. The results also have significant implications for rehabilitation and highlight the role of social interaction as a means of facilitating new learning in individuals with memory impairment.

Yet, there is a caveat: the learning did not require the acquisition of new arbitrary relations, an ability that relies critically on the hippocampus and that is part of what normal semantic learning typically demands. The patients with amnesia negotiated meaningful labels for the tangrams using pre-existing semantic information (e.g., “siesta man” for a figure that could be viewed as a person lying against a tree). When patients with hippocampal amnesia are the matchers, and their partners are the directors (i.e., the ones generating the perceptual and linguistic perspectives), the patients show little learning, likely because the to-be-learned labels generated by their partners are, in the minds of the patients, arbitrarily related to the tangram figures (Gupta Gordon et al., 2018). Thus, patients with hippocampal amnesia can be successful at learning new semantic information when the task does not demand hippocampal mediated learning (e.g., arbitrary relational binding) and, in the context of real-world social communication, this learning can even be achieved at a normal rate. The role of social interaction and communication in new semantic learning warrants further consideration. Not only is social interaction the canonical context for semantic learning in development and language acquisition, but it is also the context for the most impressive examples of new semantic learning in amnesia, even if not fully normal, whether in developmental or adult-onset cases of amnesia (Koutstaal, 2019). This is particularly true for individuals with developmental amnesia who have learned a wealth of semantic information outside the laboratory.

Looking back on all the evidence of new semantic learning in amnesia, there is yet to be a replicable example of fully normal semantic learning (i.e., where the rate and amount of learning between amnesic patients and controls are similar and where the to-be-learned information encompasses the full range of demands (arbitrary binding) that are inherent to semantic memory). While there are learning conditions and formats that promote new learning in amnesia (e.g., errorless learning), when evaluated together and with a fixed standard, the empirical evidence shows that patients with dense amnesia following hippocampal damage fail to show normal acquisition of new semantic information, and thus supports the conclusion that the hippocampus plays a necessary role in the acquisition of new semantic memory. Taken altogether, although over time semantic and episodic memory have largely been studied separately, and increasingly apart from the early question of whether both forms of memory share a common neural substrate, the evidence is compelling that new

semantic learning, like new episodic learning, relies critically on the hippocampus.

## Remote Semantic Memory in Amnesia

There has been an overwhelming consensus that remote semantic knowledge, acquired long before the onset of hippocampal pathology, becomes independent of the hippocampus *via* neocortical consolidation (McClelland et al., 1995) and is intact in amnesia. This view has been supported by data from patients with hippocampal amnesia on tests of linguistic knowledge: patients with amnesia do not have aphasia or semantic dementia, and they perform within normal limits on neuropsychological measures of vocabulary knowledge and naming (Kensinger et al., 2001). Further, patients with amnesia perform similarly to healthy participants on measures thought to assess remote word knowledge, like naming or matching a label with a phrase, definition, or sentence (Gabrieli et al., 1988; Verfaellie et al., 2000; Manns et al., 2003). Together, these data have been taken as evidence that patients with amnesia have intact remote semantic memory.

Perhaps the methods used in these studies are not fine-grained enough to detect impairment in patients with amnesia. Many of the tasks used in these studies were originally designed to detect aphasia or semantic dementia. As such, they capture differences in naming or linguistic ability at a coarse level. Examples of the procedures used include showing participants a picture of a common object, like an apple, and prompting patients to name it; matching the label *apple* to a definition like, *a sweet, red fruit*; and determining whether A-P-P-L-E is a real English word. While tests such as these are certainly useful in identifying a deficit in people with severe semantic or naming impairment, they do not capture more subtle deficits that may be evident in the remote semantic memory of patients with amnesia.

The same can be said of clinical tools commonly used for detection of deficits in people with semantic dementia or Alzheimer’s disease. Two such tools are the Semantic Memory Test Battery and the Boston Naming Test. These tests tend to be implemented with relatively few naming trials. When these tests are used in people with semantic dementia, naming impairment is evident. For example, studies with this population using just 28 items (Lambon Ralph et al., 2007) and 48 items (Schmolck et al., 2002) found deficits in naming. When these tests are used in patients with hippocampal amnesia, no naming impairment is found. Kensinger et al. (2001) tested patient HM using the Boston Naming Test—which included 42 black-and-white line drawings—and developed two picture naming tasks. One task had 96 colored pictures of objects and the other had 105 black and white drawings. HM performed similarly to controls on these tasks, leading to the interpretation that his remote semantic knowledge was intact.

More recently, researchers have sought to examine remote semantic memory in patients with amnesia using more sensitive measures that align more closely with approaches to study semantic richness (see below). Klooster and Duff (2015) examined how much information is associated with highly familiar words that were previously acquired in patients with amnesia and healthy and brain-damaged comparison

participants. The tasks included a word associates test (identifying synonyms and common collocates), a word senses task (name all the senses of a word; e.g., lemon can be a fruit, a color, a defective automobile) and a word features task (name all of the features of a word; e.g., lemon tastes sour, is native to Asia, used in tea). Patients with amnesia performed significantly worse than healthy and brain-damaged comparison groups (i.e., patients with ventromedial prefrontal cortex damage), on all three measures of word knowledge. For example, patients with amnesia generated, on average, only half the number of features for common words (e.g., shirt) as comparison participants. The deficit in remote semantic memory was even evident on tasks where all the information was in view of the participants. For example, when provided with a word (e.g., sudden) and asked to endorse possible synonyms (e.g., beautiful, quick, surprising, thirsty), all of which were written on paper in view of the participants, individuals with amnesia were significantly less likely to identify the correct responses. Furthermore, this deficit was evident despite showing no differences from comparison participants on self-reported rates of familiarity (scoring familiarity on a 9-point scale) of words used in the word features and senses tasks. Importantly, the fact that the patients knew these words (i.e., had high familiarity ratings), suggests that they likely would have performed like comparison participants with traditional measures (e.g., naming) that only assess surface level semantic knowledge. Using tasks and measures that assess semantic richness, or depth of semantic knowledge, patients with hippocampal amnesia perform significantly worse than comparison groups suggesting impoverished remote semantic memory. These findings also raise the possibility that the hippocampus plays a long-term role in maintaining semantic representations across the lifetime.

Returning to studies of naming, deficits in remote semantic knowledge in amnesia are also evident when a more extensive set of items are probed. Dawood et al. (2018) conducted a naming task similar to previous studies in which patients with amnesia and comparison participants viewed color photographs of items and were instructed to provide a name for the picture. Unlike previous naming studies that all contained fewer than 100 images, this study used 1,458 items from the Bank of Standardized Stimuli (BOSS) database (Brodeur et al., 2010, 2014) that varied across a range of word features such as imageability, frequency, and familiarity. By using a wide range of image-word pairs, even subtle differences between patients with amnesia and comparisons in naming may be detected. Unlike previous tests of naming in this population, Dawood et al. (2018) found that patients with amnesia were less likely than comparison participants to correctly name the objects that they viewed. Furthermore, patients with amnesia were more likely to provide a general label for an object (e.g., *bird* for a *cardinal*) than healthy participants. Using a wider range of materials and a detailed analysis of error type provides further evidence of the impoverishment of remote semantic memory in amnesia.

Closer examination of language production also reveals group differences where patients with amnesia use words rated as

less semantically rich relative to controls. Hilverman et al. (2017) analyzed the features of words used when patients with amnesia and healthy participants described events, both past and imagined. Features of words reflect characteristics of what the word describes (e.g., a word's imageability measures the degree to which the word invokes an image in one's mind). Although patients with amnesia are known to produce significantly fewer episodic details in their descriptions of events (Race et al., 2011), the specific words that are used are not necessarily related to the number of episodic details; similar representations can be communicated with the same amount of episodic details but using words that vary considerably in their imageability and concreteness. For example, one could say, "I was on a jetski on a nice summer day and water was hitting my face as I went across the lake" or "I was riding a jet ski on a bright summer day and water was spraying my face as I sped across the lake". In both cases, the number of episodic details is the same, but the imageability and concreteness of the words used are much greater in the second account. Hilverman et al. (2017) found that patients with hippocampal amnesia used words that were significantly less imageable than healthy comparison participants. This was found even when controlling for number of overall features in the narrative and word frequency. This finding fits with data from Heyworth and Squire (2019) who found that in narrative recollections of a guided walk, patients with amnesia used higher-frequency and less concrete words than controls. Thus, even in semi-naturalistic speaking contexts, patients with amnesia demonstrate language use that is semantically impoverished.

These deficits in remote semantic memory are not present only in fine-grained aspects of language. Similar findings have been demonstrated in patients with amnesia when describing semantic knowledge acquired long before the onset of their amnesia. When prompted to recount fairy tales and bible stories, patients with amnesia produce fewer details than controls (Rosenbaum et al., 2009; Verfaellie et al., 2014). Patients with MTL lesions also show impairment in the general details and in the ordering of the main steps (Verfaellie et al., 2014). Further, a review of neuropsychological research on autobiographical knowledge demonstrated that patients with MTL damage were impaired on measures of autobiographical fact knowledge—a type of personal semantic memory—relative to comparison participants (Grilli and Verfaellie, 2014). Finally, patients with MTL damage are impaired relative to healthy participants at generating hypothetical meanings for novel word compounds (e.g., cactus carpet) suggesting that the hippocampus plays a role in relational and combinatorial semantic processing even when remote knowledge of the individual words appeared intact (Keane et al., 2019).

There is growing evidence of remote semantic memory impairment in amnesia. These impairments may mirror deficits in remote episodic memory in amnesia. Close examination of remote episodic memory in amnesia reveals a lack of specificity, detail, and richness relative to healthy participants (e.g., Rosenbaum et al., 2008; St-Laurent et al., 2014; Robin et al., 2019) and support the proposal that the hippocampus plays a long-term or permanent role in the maintenance of

episodic memory representations (Nadel and Moscovitch, 1997). To test the notion that hippocampus plays a long-term or permanent role in the maintenance of both episodic and semantic memory, researchers will need to develop/apply methodological approaches to the study of semantic memory that mirror those used to study episodic memory in terms of their ability to capture the breadth and richness of the multimodal and relational features that are inherent to both forms of memory.

## METHODOLOGICAL AND THEORETICAL APPROACHES TO STUDYING EPISODIC AND SEMANTIC MEMORY

One challenge of testing the shared dependence of episodic and semantic memory on the hippocampus has been equating task demands and characteristics of the to-be-learned stimuli across memory systems. A consequence of the early confirmation and consensus on the role of the hippocampus in episodic memory (while the early data on semantic memory were more equivocal) is that the number of investigations and highly sophisticated experimental designs to study episodic memory have significantly outpaced those on semantic memory. Consistent with proposals that view the hippocampus as playing a critical role in relational binding and in the flexible (re)construction and (re)combination of rich multimodal features of events and experiences (Eichenbaum and Cohen, 2001; Schacter and Addis, 2007; Ranganath, 2010; Yonelinas, 2013; Rubin et al., 2017), the field now has a diverse set of methods for capturing and quantifying the relational features and contextual richness of episodic memory. For example, to study episodic memory, we have coding schemes for rating and quantifying the spatial, temporal, and perceptual vividness and richness of event narratives (e.g., Levine et al., 2002), experimental designs for examining how episodes are (re)constructed, (re)combined, and integrated across time, space, and people (e.g., Zacks and Swallow, 2007; Schacter et al., 2008; Schlichting and Preston, 2015; Eichenbaum, 2017) from photographs, text, and movie clips (e.g., Staresina and Davachi, 2009; Zacks et al., 2009; St-Laurent et al., 2014), and techniques like eye-tracking (e.g., Ryan et al., 2000) and entropy analyses (e.g., Lucas et al., 2019) that allow us to study episodic encoding and recall, and its organization, without asking participants to explicitly study or remember. In contrast, particularly in patient studies, the study of semantic memory still largely involves asking individuals to label pictures of famous faces and to learn facts or word-meaning pairings (Manns et al., 2003; Sharon et al., 2011). Our methods and techniques for measuring episodic and semantic memory, and equating task demands and stimuli, are further apart than they were decades ago.

This lack of methodological depth and breadth in the study of semantic memory (and therefore the lack of substantive data) has made it difficult for researchers to offer complete and comprehensive theories across distinct forms of memory. For example, Nadel and Moscovitch (1997) note in their seminal paper laying out the points of similarity and divergence between standard consolidation models and their multiple trace theory that most studies of remote general semantic knowledge do not

include detailed tests sensitive enough to detect deficits, which limits the comparison to other forms of memory. More recently, Yonelinas et al. (2019) proposed an alternative to standard systems consolidation theory called contextual binding theory which focuses nearly exclusively on the role of the hippocampus in episodic memory. Discussion of semantic memory was cursory, with the authors simply stating that whether or not contextual binding theory might be applied to semantic memory is an open question. Indeed, given the dearth of semantic memory studies with sufficient depth and sensitivity, this is all that can be said. This lack of data and methods may also make it more attractive, or tractable, to test hypotheses for which there are more established data and tools (as is the case in the area of episodic memory). Thus, over the past several decades, not only have researchers moved further away from testing if episodic and semantic memory has shared neural correlates, but, as a field, we are ill-equipped (methodologically and theoretically) to do so.

Other disciplines (e.g., psycholinguistics, semiotics, cognitive science) however, have conceptualized semantic memory as a knowledge system that is as rich, relational, and multifaceted as we have come to view episodic memory. From these fields come a set of tools and methods with increased sensitivity to capture a wider breadth of semantic memory phenomena than used in the memory literature to date. These methods may also have utility in attempts to equate task demands and stimuli across memory systems. In the next section, we review some of these broader approaches to demonstrate the similarities between episodic and semantic memory and to highlight their application to recent studies of hippocampal contributions to semantic memory.

## SEMANTIC MEMORY AS A FLEXIBLE, CONSTRUCTIVE, RELATIONAL, AND MULTIMODAL SYSTEM

Episodic memory is often described as a dynamic system capable of reconstructive and combinational processes that allow us to recollect about our past and simulate future events (Buckner and Carroll, 2007; Schacter and Addis, 2007). While the study of semantic memory in amnesia has often been reduced to word-definition pairs or recognition of famous faces or facts, other perspectives view semantic memory as a highly flexible, (re)constructive, relational and multimodal system that we use to create, represent, and extract meaning as we navigate our most fundamental interactions with the environment and each other (Rogers et al., 2004; Reilly et al., 2016). Like episodic memory, semantic knowledge is not a static repository of information. Rather, it grows and changes as we continuously acquire, integrate, and reinforce rich representations of the relations between words, their referents, and their relations with associated referents (Zettersten et al., 2018; Klooster et al., 2019). Indeed, it is estimated that the average English-speaking adult has acquired 12.5 million bits of information, the majority of which is lexical-semantic knowledge (Mollica and Piantadosi, 2019). These millions of bits of information are not isolated, but rather are interconnected and combined in both familiar and novel ways to represent and act in the world.



The acquisition of richly interleaved semantic knowledge is facilitated by the dynamic contexts in which words are learned and used. For example, single words are seldom learned or presented in isolation. Rather, words appear in rich contexts in which related words and concepts are also present, facilitating the development of interrelated semantic representations that can be flexibly deployed (Wojcik and Saffran, 2013, 2015; Wojcik, 2018). In addition to representing the relations between words and their referents, while adding increasing layers of nuance to the meanings of words over time (Ellis and Ogden, 2017), learners also represent relationships *among* lexical items, based on their co-occurrence in the ambient language (Arnon and Christiansen, 2017). That is, many sequences of words repeatedly co-occur in language and we encode those relations in addition to our knowledge of individual words (Pawley and Syder, 1980).

Like episodic memory, which is often characterized, and measured, in terms of its richness (e.g., episodic richness is the amount of multimodal information that is associated with a given event or experience; Levine et al., 2002; St-Laurent et al., 2014), semantic memory is also characterized, and measured, by richness. Semantic richness refers to the amount of information contained within or associated with a word or concept and it influences the speed and accuracy of behavioral responses (e.g., greater semantic richness is associated with faster and more accurate naming, lexical decision, categorization; Pexman et al., 2002, 2003; Duñabeitia et al., 2008; Grondin et al., 2009). Words and concepts that are richer, or associated with more information, are also better remembered (Hargreaves et al., 2012).

Semantic richness can be indexed or measured in a number of ways. It can be a metric of how many concepts, words, or features are associated with a specific word. Words with denser semantic neighborhoods—or words that are associated with many different words or concepts—are processed more quickly in naming, lexical retrieval, and lexical decision tasks (e.g., it is easier to retrieve the word “nurse” after viewing the word “doctor” than it would be having just viewed the word “grass;” Hargreaves and Pexman, 2012; Yap et al., 2012; Taler et al., 2013). Semantic richness can also be represented by how many sensory and perceptual features are associated with a particular word or concept. Indeed, words that are higher in imageability (can readily generate a mental image) and concreteness (can be imagined with the senses) are typically processed more quickly; it is easier to retrieve the word “banana”—something that can be seen, touched and tasted—than it is to retrieve the word “government”—a concept that is more abstract (e.g., Bennett et al., 2011). Semantic richness can also be a reflection of how many contexts a word or concept is associated with or can be successfully used in, typically measured across print sources (Adelman et al., 2006) but may also extend to distinct physical settings and speakers. Words that appear across more diverse contexts facilitate faster word naming and lexical decision times than do words that are just more frequently occurring. From the perspective of richness, there are obvious parallels between semantic and episodic memory. Manipulating semantic richness may be one way to help equate stimuli and task demands

across memory systems. For example, work by Klooster and Duff (2015) and Hilverman et al. (2017) documenting deficits in semantic richness (e.g., the amount of information associated with a word) in patients with hippocampal damage highlights the shared role of the hippocampus in both episodic and semantic richness. Manipulating context as a form of semantic richness may also provide an opportunity to expand on, or test, existing memory theory. For example, contextual diversity is an interesting measure as it seems to capture the interaction of semantic representation and episodic experience rather than the extraction or decontextualization of semantics from episodes (e.g., semantization).

Rich semantic representations allow us to go beyond the literal meanings of words themselves, combining and integrating across concepts to communicate meanings that might otherwise be inexpressible (Katz, 1989). For example, the use of metaphor in human communication and thought is widespread (Lakoff and Johnson, 1980). To generate and comprehend metaphors (e.g., “my job is a jail”), language users create or identify *relations* between the metaphor topic (“job”) and vehicle (“jail”). Metaphor comprehension requires rapid processing of novel relations between seemingly disparate lexical items, and may, therefore, place high demands on the MTL relational memory system. Use of a metaphor is also inherently creative. Metaphors are thought to be a primary device driving lexical innovation (McGlone et al., 1994; Makkai et al., 1995). Metaphors help to fill lexical “gaps” in a language by extending existing words to describe novel categories and concepts. Another example is a conceptual combination. Speakers leverage the relations among lexical items to create new concepts and meaning by combining words and concepts from pre-existing knowledge stores (e.g., elephant-ferry; these words can be processed individually or as an integrated concept, an elephant ferry; Coutanche et al., in press; Lucas et al., 2017).

Metaphor and conceptual combination would seem to require the same compositionality and representational flexibility inherent in characterizations of episodic memory. That is, relational representations (semantic and episodic) can be broken down into constituent elements, which can then be combined and recombined in novel ways (Cohen and Eichenbaum, 1993; Cohen et al., 1997). Metaphor generation and conceptual combination clearly involve the combination of far-reaching mental representations and results in the generation of a verbal expression that creatively combines disparate concepts to provide the listener with novel insight. These creative combinatorial and constructive features of semantic memory processing and use are highly reminiscent of the flexible and creative (re)construction and (re)combination of episodic memory representations for remembering the past and imagining the future (Buckner and Carroll, 2007; Schacter and Addis, 2007). Indeed, individuals with hippocampal pathology are impaired in creative uses of language (Duff et al., 2009) including metaphor comprehension (Covington et al., 2017). Furthermore, work by Keane et al. (2019) on generating novel meanings for word combinations (e.g., cactus carpet) highlights the shared role of the hippocampus in both relational episodic processing and relational semantic processing.

Viewed through a broader interdisciplinary lens, episodic and semantic memory have many shared features including the depth and breadth of multimodal relational information they encompass and the constructive and flexible nature of their expression and use across contexts. While these shared features align closely with the processing capabilities of the hippocampus (e.g., relational binding, representational flexibility, compositionality; Cohen et al., 1997; Eichenbaum and Cohen, 2001), in the core memory literature, these broader semantic paradigms, and their (in)dependence to the hippocampal memory system, have, until recently been understudied. We next review recent developments in our understanding of the hippocampus that further align, and demonstrate, the capacity of the hippocampus to meet the processing demands of semantic memory use and processing.

## EXTENDING THE REACH OF THE HIPPOCAMPUS AND ITS ROLE IN SEMANTIC MEMORY PROCESSING

The hippocampus has long been associated with long-term memory. Converging evidence has challenged the traditional view that the hippocampus exclusively supports long-term memory, showing that the hippocampus plays a critical role in memory for relations over very short delays, and even when there are no delays at all, on the timescale of short-term or working memory (Hannula et al., 2006, 2017; Olson et al., 2006; Hannula and Ranganath, 2008). These findings suggest that new hippocampus-dependent representations are available rapidly enough to influence ongoing processing when: new information is perceived; old information is retrieved; and representations are held on-line to be evaluated, manipulated, integrated, and used in service of behavioral performance. That is, the hippocampus is critical not only for the ability to form new enduring memories and to recover the past, but also for the creation, maintenance, updating, and use of on-line representations in support of ongoing information processing. These findings raise the possibility of hippocampal involvement in real-time semantic processing.

The hippocampus has also long been associated with explicit and conscious processing. Recent work, however, implicates the hippocampus in the incremental and implicit/unconscious processing of arbitrary relations (for review, see Hannula and Greene, 2012), suggesting that consciousness alone is not a reliable predictor of what neural region or memory system contributes to a given behavioral phenomena. Although implicit semantic processing tasks have often been assumed to be hippocampal independent, these new findings raise the possibility that the hippocampus may contribute to some aspects of unconscious or implicit semantic processing (also see Gaskell et al., 2019). Initial support for such a prediction comes from data pointing to hippocampal contributions to statistical learning, the process by which individuals uncover patterns in their environment by tracking co-occurrence frequencies amongst stimuli. In language, statistical learning is the proposed mechanism by which we learn to segment

words from continuous speech (Saffran et al., 1996), uncover grammatical structure (Gómez, 2002; Saffran and Wilson, 2003), and learn to recognize the phonotactic, orthographic, and morphological regularities (Chambers et al., 2003; Pacton et al., 2005). There is also evidence to suggest that statistical learning mechanisms contribute to semantic knowledge by supporting the mapping of word meanings onto word forms (Graf Estes et al., 2007; Lany and Saffran, 2011; Lany, 2014). Although considered an implicit learning process, recent work (imaging and patient studies) demonstrates a role for the hippocampus in the tracking of statistical regularities in the environment, across stimulus modalities (Schapiro et al., 2012, 2014; Covington et al., 2018).

Taken together with the long-acknowledged role of the hippocampus in relational binding, these new findings have significant implications for understanding the role the hippocampus may play in various stages of acquisition, maintenance, activation, and use of semantic information. By combining broader theoretical and methodological approaches to semantic memory and the functionality of the hippocampus, there is a growing literature demonstrating hippocampal contributions to semantic progressing in the moment. Next, we highlight studies that have documented hippocampus contributions in on-line semantic memory processing.

## Hippocampal Contributions to Semantic Processing in the Moment

A particularly innovative approach to studying hippocampal contributions to on-line semantic memory processing comes from intracranial recordings from depth electrodes in patients with intractable epilepsy. These studies have the advantage of a high degree of both spatial and temporal specificity, allowing for tests of the nature and time course of hippocampal contributions to semantic processing. Two such studies demonstrate hippocampal coding for semantic representations depending on a similar mechanism to hippocampal coding for space/episodes: hippocampal theta power. The role of the hippocampus is well-established in the encoding of relations for representing and navigating physical space (O'Keefe and Nadel, 1978; Nadel, 1991). Solomon et al. (2019) ask if hippocampal theta oscillations represent semantic distances between words (i.e., the similarity or likeness in meaning between words as measured by corpus analysis), similar to how these same oscillations code for relations in physical space. In this study, patients with depth electrodes with contacts on hippocampus completed study and recall of sets of 12-item lists. During recall, patients demonstrated the expected behavioral pattern of clustering list items based on both their temporal relations (e.g., words in close serial proximity during the study were recalled in clusters during recall) and also based on semantic relations (e.g., words closer in semantic space were recalled in clusters during recall). Hippocampal theta power prior to the retrieval event was predictive of the semantic relationship in the two subsequently recalled words, suggesting that hippocampal theta power codes for semantic relatedness in multi-dimensional word space. These data are striking as they suggest a role for the hippocampus in tracking and

representing the relations among words in semantic memory in a manner that is similar to how the hippocampus tracks and represents relations in physical space and events in episodic memory.

Piai et al. (2016) demonstrated relationships between hippocampal theta power and semantic processing during language comprehension. In contrast to the list learning in the Solomon study, patients in the Piai study were not required to learn any new information. In this study, the patients listened to sentences with the final word omitted and were then presented with a picture to name that could complete the sentence. In the experiment, half of the sentences presented to the patients began with a sentence stem that linguistically constrained the possible final word [e.g., “She locked the door with the” (picture: key)] while the other half were linguistically unconstrained [e.g., “She walked in here with the” (picture: key)]. The results demonstrated that constraining sentence stems facilitated the picture naming response, and that hippocampal theta power increased during the sentence stem for the constrained vs. unconstrained sentence stems, prior to the picture onset. Further analysis of these data demonstrated that the increases in theta power were related to increasing semantic associations between words in the sentence. Using latent semantic analysis (LSA), Piai et al. (2016) determined the “context-defining word” for each sentence (i.e., the word with the strongest LSA association to the final picture name). In the constrained condition, all patients demonstrated increased theta power at this keyword compared to the preceding word, a pattern that was not present in the unconstrained condition. These results demonstrated that the hippocampus contributes to tracking and building semantic associations across words, and suggest a role for the hippocampus in predictive language processing (also see Bonhage et al., 2015), consistent with its role in predictive processing in other domains (Buckner, 2010; Covington and Duff, 2016).

In a similar study to Piai et al. (2016), Jafarpour et al. (2017) examined patterns of hippocampal activity, specifically hippocampal high-frequency band (HFB) power, during the 0.5-second pause between the sentence stem and the appearance of the to-be-named picture. Greater HFB power was observed during the pre-picture period during the highly constraining vs. low constraint sentences, suggesting pre-activation of the expected semantic representation. Indeed, patterns of HFB power in the pre-picture and picture intervals were compared using time series analyses, and the degree of similarity between these patterns was higher for highly constrained items. These patterns of hippocampal HFB power were then compared to one another based on semantic similarity (as calculated using LSA). Results indicated that HFB power pre-activation patterns were more similar for pictures that were closer in semantic distance to one another.

Finally, data from intracranial recordings also suggest that the hippocampus contributes to word retrieval during picture naming (Hamamé et al., 2014). During picture naming, left hippocampal HFB power increased during the period between picture presentation and word production, relative to the pre-stimulus baseline. Peak-latency of this hippocampal response

was predictive of participants’ trial-by-trial naming latency. The authors suggest that these results point to a role for the hippocampus in retrieving the arbitrary associations between objects and their names.

The results from these intracranial recording studies suggest that, in addition to the role for the hippocampus in the acquisition of new semantic memory and maintenance of remote semantic memory, the hippocampus also encodes, tracks, and builds semantic relations of previously acquired words during on-line sentence processing to create meaning in the moment and to facilitate communication (see Cross et al., 2018; Gaskell et al., 2019). The role of the hippocampus in semantic memory processing appears remarkably similar to the role the hippocampus plays in its support of episodic memory. Building on this work, interdisciplinary approaches to the study of hippocampal contributions to semantic memory promise to expand and refine the theories and methods across fields and may offer researchers new paradigms that will allow for integrating the study of episodic and semantic memory.

## CONCLUSION

It has been nearly 50 years since Tulving (1972) suggested that memory research may benefit from observing a distinction between episodic and semantic memory. Unquestionably, Tulving’s thought experiment has been a significant catalyst in the empirical and theoretical study of multiple memory systems. The shared neural correlates and the commonalities in processing and representation of semantic and episodic memory suggest to us that these forms of memory have more in common than Tulving’s initial distinction, and the work that followed, suggested (also see Renoult et al., 2019). Indeed, like episodic memory, semantic memory is a highly flexible, (re)constructive, relational and multimodal knowledge system. Furthermore, like episodic memory, semantic memory also depends critically on the hippocampus; patients with dense amnesia following hippocampal damage cannot acquire new semantic memory fully normally, just as they do not have the normal capacity for acquiring new episodic memory. This review highlights the role the hippocampus plays across nearly all stages of semantic memory including acquisition, maintenance, and processing in real-time.

There is growing recognition that the history of studying memory systems in isolation and the search for dissociations has led many to overlook the well-documented interdependence of episodic and semantic memory (Greenberg and Verfaellie, 2010; Ferreira et al., 2019; Renoult et al., 2019). Recent work also highlights the pivotal role semantic memory plays across many, if not all, forms of episodic memory, irrespective of time constraints (Irish and Piguet, 2013). Future work developing methods and materials that fully capture the depth and breadth of semantic memory and processing will be critical in facilitating comparison across forms of memory and in understanding their cognitive and neural (inter)dependencies as well as in testing the psychological and anatomical reality of the distinction in memory between semantic and episodic memory.

Integrating the study of episodic and semantic, understanding their interactions, interdependencies, and shared mechanisms, promises to advance our understanding of how words, concepts, and meaning, as well as episodes and events, are integrated, instantiated and maintained in memory, giving new insights into our two most quintessentially human abilities: memory and language.

## REFERENCES

- Adelman, J. S., Brown, G. D. A., and Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychol. Sci.* 17, 814–823. doi: 10.1111/j.1467-9280.2006.01787.x
- Arnon, I., and Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Top. Cogn. Sci.* 9, 621–636. doi: 10.1111/tops.12271
- Baddeley, A., Vargha-Khadem, F., and Mishkin, M. (2001). Preserved recognition in a case of developmental amnesia: implications for the acquisition of semantic memory? *J. Cogn. Neurosci.* 13, 357–369. doi: 10.1162/08989290151137403
- Baddeley, A. D., and Wilson, B. A. (1994). When implicit learning fails: amnesia and problem of error elimination. *Neuropsychologia* 32, 53–68. doi: 10.1016/0028-3932(94)90068-x
- Bayley, P., and Squire, L. (2002). Medial temporal lobe amnesia: gradual acquisition of factual information by nondeclarative memory. *J. Neurosci.* 22, 5741–5748. doi: 10.1523/JNEUROSCI.22-13-05741.2002
- Bennett, S. D., Burnett, A. N., Siakaluk, P. D., and Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behav. Res. Methods* 43, 1100–1109. doi: 10.3758/s13428-011-0117-5
- Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001
- Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055
- Blumenthal, A., Duke, D., Bowles, A., Gilboa, R., Rosenbaum, R., Kohler, K., et al. (2017). Abnormal semantic knowledge as a case of developmental amnesia. *Neuropsychologia* 102, 237–247. doi: 10.1093/cercor/bhp055
- Bonhage, C., Mueller, L., Friederici, A., and Fiebach, C. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex* 68, 33–47. doi: 10.1016/j.cortex.2015.04.011
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., and Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One* 5:e10773. doi: 10.1371/journal.pone.0010773
- Brodeur, M. B., Guérard, K., and Bouras, M. (2014). Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS One* 9:e106953. doi: 10.1371/journal.pone.0106953
- Buckner, R. (2010). The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol.* 61, 27–48. doi: 10.1146/annurev.psych.60.110707.163508
- Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci.* 11, 49–57. doi: 10.1016/j.tics.2006.11.004
- Buckner, R., and Carroll, D. (2007). Self-projection and the brain. *Trends Cogn. Sci.* 11, 49–57. doi: 10.1016/j.tics.2006.11.004
- Chambers, K. E., Onishi, K. H., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87, B69–B77. doi: 10.1016/s0010-0277(02)00233-0
- Clarke, A., and Tyler, L. (2015). Understanding what we see: how we derive meaning from vision. *Trends Cogn. Sci.* 19, 677–687. doi: 10.1016/j.tics.2015.08.008
- Cohen, N. J. (1984). “Preserved learning capacity in amnesia: evidence for multiple memory systems,” in *Neuropsychology of Memory*, eds L. R. Squire and N. Butters (New York, NY: Guilford Press), 83–103.
- Cohen, N. J., and Eichenbaum, H. (1993). *Memory, Amnesia and the Hippocampal System*. Cambridge, MA: MIT Press.
- Cohen, N. J., Poldrack, R. A., and Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory* 5, 131–178. doi: 10.1080/741941149
- Cohen, N. J., and Squire, L. R. (1980). Preserved learning and retention of a pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* 210, 207–210. doi: 10.1126/science.7414331
- Corkin, S. (2002). What’s new with the amnesic patient H.M.? *Nat. Rev. Neurosci.* 3, 153–160. doi: 10.1038/nrn726
- Covington, N. V., Brown-Schmidt, S., and Duff, M. C. (2018). The necessity of the hippocampus for statistical learning. *J. Cogn. Neurosci.* 30, 680–697. doi: 10.1162/jocn\_a\_01228
- Covington, N., and Duff, M. C. (2016). Expanding the language network: direct contributions from hippocampus. *Trends Cogn. Sci.* 20, 869–870. doi: 10.1016/j.tics.2016.10.006
- Covington, N., Kurczek, J., and Duff, M. C. (2017). “Metaphor generation and comprehension in individuals with hippocampal amnesia,” in *Poster Presentation at the American Speech Language Hearing Association* (Los Angeles, CA, USA).
- Cross, Z. R., Kohler, M. J., Schlesewsky, M., Gaskell, M. G., and Bornkessel-Schlesewsky, I. (2018). Sleep-dependent memory consolidation and incremental sentence comprehension: computational dependencies during language learning as revealed by neuronal oscillations. *Front. Hum. Neurosci.* 12:18. doi: 10.3389/fnhum.2018.00018
- Damasio, A. R., Tranel, D., and Damasio, H. (1989). “Amnesia caused by herpes simplex encephalitis, infarctions in basal forebrain, Alzheimer’s disease and anoxia/ischemia,” in *Handbook of Neuropsychology* (Vol. 3), eds F. Boller and J. Grafman (Amsterdam: Elsevier), 149–166.
- Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* 16, 693–700. doi: 10.1016/j.conb.2006.10.012
- Davachi, L., Mitchell, J., and Wagner, A. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci. U S A* 100, 2157–2162. doi: 10.1073/pnas.0337195100
- Davies, R. R., Graham, K. S., Xuereb, J. H., Williams, G. B., and Hodges, J. R. (2004). The human perirhinal cortex and semantic memory. *Eur. J. Neurosci.* 20, 2441–2446. doi: 10.1111/j.1460-9568.2004.03710.x
- Davis, M., and Gaskell, G. (2009). A complementary systems account of word learning: neural and behavioral evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- Dawood, M., Cole, A., Hilverman, C., and Duff, M. C. (2018). “Hippocampal contributions to remote semantic memory: evidence of impaired naming in amnesia,” in *Poster Presentation at the American Speech Language Hearing Association* (Boston, MA, USA).
- Dewar, M., Garcia, Y. F., Cowan, N., and Della Sala, S. (2009). Delaying interference enhances memory consolidation in amnesic patients. *Neuropsychology* 23, 627–634. doi: 10.1037/a0015568
- Duff, M. C., Hengst, J., Tranel, D., and Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nat. Neurosci.* 9, 140–146. doi: 10.1038/nn1601
- Duff, M. C., Hengst, J., Tranel, D., and Cohen, N. J. (2009). Hippocampal amnesia disrupts verbal play and the creative use of language in social interaction. *Aphasiology* 23, 926–939. doi: 10.1080/02687030802533748
- Duñabeitia, J. A., Avilés, A., and Carreiras, M. (2008). NoA’s ark: influence of the number of associates in visual word recognition. *Psychon. Bull. Rev.* 15, 1072–1077. doi: 10.3758/pbr.15.6.1072
- Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron* 95, 1007–1018. doi: 10.1016/j.neuron.2017.06.036

## AUTHOR CONTRIBUTIONS

MD and NJC planned the scope and content of the review. MD did the majority of the writing for the initial version of the manuscript with assistance from NVC and CH. All authors contributed to the final version of the manuscript, intellectually and in the writing and editing.



- Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*. New York, NY: Oxford University Press.
- Eichenbaum, H. B. (1998). Amnesia, the hippocampus, and episodic memory. *Hippocampus* 8:197. doi: 10.1002/(sici)1098-1063(1998)8:3<205::aid-hipo3>3.0.co;2-i
- Eichenbaum, H., Yonelinas, A. P., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* 30, 123–152. doi: 10.1146/annurev.neuro.30.051606.094328
- Ellis, N. C., and Ogden, D. C. (2017). Thinking about multiword constructions: usage-based approaches to acquisition and processing. *Topics Cogn. Sci.* 9, 604–620. doi: 10.1111/tops.12256
- Elward, R. L., Dzieciol, A. M., and Vargha-Khadem, F. (2019). Little evidence for fast mapping in adults with developmental amnesia. *Cogn. Neurosci.* 10, 215–217. doi: 10.1080/17588928.2019.1593123
- Elward, R. L., and Vargha-Khadem, F. (2018). Semantic memory in developmental amnesia. *Neurosci. Lett.* 680, 23–30. doi: 10.1016/j.neulet.2018.04.040
- Ferreira, C., Charest, I., and Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. *Neuroimage* 201:115996. doi: 10.1016/j.neuroimage.2019.07.009
- Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annu. Rev. Psychol.* 49, 87–115. doi: 10.1146/annurev.psych.49.1.87
- Gabrieli, J. D. E., Cohen, N. J., and Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain Cogn.* 7, 157–177. doi: 10.1016/0278-2626(88)90027-9
- Gardiner, J. M., Brandt, K. R., Baddeley, A. D., Vargha-Khadem, F., and Mishkin, M. (2008). Charting the acquisition of semantic knowledge in a case of developmental amnesia. *Neuropsychologia* 46, 2865–2868. doi: 10.1016/j.neuropsychologia.2008.05.021
- Gaskell, M. G., Cairney, S., and Rodd, J. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition* 182, 109–126. doi: 10.1016/j.cognition.2018.09.007
- Glisky, E. L. (1992). Acquisition and transfer of declarative and procedural knowledge by memory-impaired patients: a computer data-entry task. *Neuropsychologia* 30, 899–910. doi: 10.1016/0028-3932(92)90034-j
- Glisky, E. L., Schacter, D. L., and Tulving, E. (1986). Computer learning by memory-impaired patients: acquisition and retention of complex knowledge. *Neuropsychologia* 24, 313–328. doi: 10.1016/0028-3932(86)90017-5
- Gold, J., and Squire, L. (2005). Quantifying medial temporal lobe damage in memory-impaired patients. *Hippocampus* 15, 79–85. doi: 10.1002/hipo.20032
- Gordon Hayman, C. A., Macdonald, C. A., and Tulving, E. (1993). The role of repetition and associative interference in new semantic learning in amnesia: a case experiment. *J. Cogn. Neurosci.* 5, 375–389. doi: 10.1162/jocn.1993.5.4.375
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–437. doi: 10.1111/1467-9280.00476
- Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260. doi: 10.1111/j.1467-9280.2007.01885.x
- Greenberg, D. L., and Verfaellie, M. (2010). Interdependence of episodic and semantic memory: evidence from neuropsychology. *J. Int. Neuropsychol. Soc.* 16, 748–753. doi: 10.1017/s1355617710000676
- Grilli, M. D., and Verfaellie, M. (2014). Personal semantic memory: insights from neuropsychological research on amnesia. *Neuropsychologia* 61, 56–64. doi: 10.1016/j.neuropsychologia.2014.06.012
- Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19. doi: 10.1016/j.jml.2008.09.001
- Gupta Gordon, R., Duff, M. C., and Cohen, N. J. (2018). “Applications of collaborative memory: patterns of success and failure in individuals with hippocampal amnesia,” in *Collaborative Remembering: How Remembering With Others Influences Memory*, eds M. Meade, A. Barnier, P. Van Bergen, C. Harris and J. Sutton.
- Gupta, P. (2012). “Word learning as the confluence of memory mechanisms: computational and neural evidence,” in *The Handbook of the Neuropsychology of Language*, Vol. 1 (1st Edn.), ed. M. Faust (Chichester, England: Wiley-Blackwell), 146–163.
- Hamann, S. B., and Squire, L. R. (1995). On the acquisition of new declarative knowledge in amnesia. *Behav. Neurosci.* 109, 1027–1044. doi: 10.1037/0735-7044.109.6.1027
- Hamamé, C. M., Alario, F. X., Llorens, A., Liégeois-Chauvel, C., and Trébuchon-Da Fonseca, A. (2014). High frequency  $\gamma$  activity in the left hippocampus predicts visual object naming performance. *Brain Lang.* 135, 104–114. doi: 10.1016/j.bandl.2014.05.007
- Hannula, D., and Duff, M. C. (Eds). (2017). *The Hippocampus From Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition*. Switzerland: Springer International Publishing.
- Hannula, D. E., and Greene, A. (2012). The hippocampus reevaluated in unconscious learning and memory: at a tipping point? *Front. Hum. Neurosci.* 6:80. doi: 10.3389/fnhum.2012.00080
- Hannula, D. E., and Ranganath, C. (2008). Medial temporal lobe activity predicts successful relational memory binding. *J. Neurosci.* 28, 116–124. doi: 10.1523/JNEUROSCI.3086-07.2008
- Hannula, D., Ryan, J., and Warren, D. (2017). “Beyond long-term declarative memory: hippocampal contributions to preception, short term retention and unconscious memory expression,” in *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition*, eds D. Hannula and M. C. Duff (Springer International Publishing: Switzerland), 281–336.
- Hannula, D. E., Tranel, D., and Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26, 8352–8359. doi: 10.1523/JNEUROSCI.5222-05.2006
- Hargreaves, I. S., and Pexman, P. M. (2012). Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. *Front. Hum. Neurosci.* 6:234. doi: 10.3389/fnhum.2012.00234
- Hargreaves, I. S., Pexman, P. M., Johnson, J. S., and Zdravilova, L. (2012). Richer concepts are better remembered: number of features effects in free recall. *Front. Hum. Neurosci.* 6:73. doi: 10.3389/fnhum.2012.00073
- Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nat. Rev. Neurosci.* 11, 523–532. doi: 10.1038/nrn2850
- Heyworth, N., and Squire, L. (2019). The nature of recollection across months and years after medial temporal lobe damage. *Proc. Natl. Acad. Sci. U S A* 116, 4619–4624. doi: 10.1073/pnas.1820765116
- Hilverman, C., Cook, S. W., and Duff, M. C. (2016). Hippocampal declarative memory supports gesture production: evidence from amnesia. *Cortex* 85, 25–36. doi: 10.1016/j.cortex.2016.09.015
- Hilverman, C., Cook, S. W., and Duff, M. C. (2017). Influence of hippocampal declarative memory on word use: patients with amnesia use less imageable words. *Neuropsychologia* 106, 179–186. doi: 10.1016/j.neuropsychologia.2017.09.028
- Irish, M., Eyre, N., Dermody, N., O’Callaghan, C., Hodges, J. R., Hornberger, M., et al. (2016). Neural substrates of semantic prospection—evidence from the dementias. *Front. Behav. Neurosci.* 10:96. doi: 10.3389/fnhbeh.2016.00096
- Irish, M., and Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Front. Behav. Neurosci.* 7:27. doi: 10.3389/fnhbeh.2013.00027
- Jafarpour, A., Piai, V., Lin, J. J., and Knight, R. T. (2017). Human hippocampal pre-activation predicts behavior. *Sci. Rep.* 7:5959. doi: 10.1038/s41598-017-06477-5
- Katz, A. N. (1989). On choosing the vehicles of metaphors: referential concreteness, semantic distances, and individual differences. *J. Mem. Lang.* 28, 486–499. doi: 10.1016/0749-596x(89)90023-5
- Keane, M., Bousquet, K., Wank, A., and Verfaellie, M. (2019). Relational processing in the semantic domain is impaired in medial temporal lobe amnesia. *J. Neuropsychol.* doi: 10.1111/jnp.12196 [Epub ahead of print].
- Kensinger, E. A., Ullman, M. T., and Corkin, S. (2001). Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: evidence from the amnesic patient H.M. *Hippocampus* 11, 347–360. doi: 10.1002/hipo.1049.abs

- Kinsbourne, M., and Wood, F. (1975). "Short term memory processes and the amnesic syndrome," in *Short-Term Memory*, eds D. Deutsch and J. S. Deutsch (San Diego, CA: Academic Press), 258–291.
- Klooster, N., and Duff, M. C. (2015). Remote semantic memory is impoverished in hippocampal amnesia. *Neuropsychologia* 79, 42–52. doi: 10.1016/j.neuropsychologia.2015.10.017
- Klooster, N., Tranel, D., and Duff, M. C. (2019). Hippocampus and semantic memory across time. *Brain Lang.* 201:104711. doi: 10.1016/j.bandl.2019.104711
- Koutstaal, W. (2019). Other 'routes in'? Has the 'fast' in the fast mapping concept led us astray? *Cogn. Neurosci.* 10, 213–214. doi: 10.1080/17588928.2019.1593124
- Kovner, R., Mattis, S., and Goldmeier, E. (1983). A technique for promoting robust free recall in chronic organic amnesia. *J. Clin. Neuropsychol.* 5, 65–71. doi: 10.1080/01688638308401151
- Lakoff, G., and Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cogn. Sci.* 4, 195–208. doi: 10.1207/s15516709cog0402\_4
- Lambon Ralph, M., Lowe, C., and Rogers, T. (2007). Neural basis of category-specific semantic deficits for living things: evidence from semantic dementia, HSVE, and a neural network model. *Brain* 130, 1127–1137. doi: 10.1093/brain/awm025
- Lany, J. (2014). Judging words by their covers and the company they keep: probabilistic cues support word learning. *Child Dev.* 85, 1727–1739. doi: 10.1111/cdev.12199
- Lany, J., and Saffran, J. R. (2011). Interactions between statistical and semantic information in infant language development. *Dev. Sci.* 14, 1207–1219. doi: 10.1111/j.1467-7687.2011.01073.x
- Levine, B., Svoboda, E., Hay, J. F., Winocur, G., and Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol. Aging* 17, 677–689. doi: 10.1037/0882-7974.17.4.677
- Lucas, H. D., Duff, M. C., and Cohen, N. J. (2019). The hippocampus promotes effective saccadic information gathering in humans. *J. Cogn. Neurosci.* 31, 186–201. doi: 10.1162/jocn\_a\_01336
- Lucas, H. D., Hubbard, R. J., and Federmeier, K. D. (2017). Flexible conceptual combination: electrophysiological correlates and consequences for associative memory. *Psychophysiology* 54, 833–847. doi: 10.1111/psyp.12840
- Makkai, A., Boatner, T., and Gates, J. E. (1995). *Dictionary of American Idioms*. Los Angeles, CA: Barrons International.
- Manns, J. R. (2004). J.F.K., L.B.J., and H.M.: the famous memories of a famous amnesic. *Hippocampus* 14, 411–412. doi: 10.1002/hipo.20010
- Manns, J. R., Hopkins, R. O., and Squire, L. R. (2003). Semantic memory and the human hippocampus. *Neuron* 38, 127–133. doi: 10.1016/s0896-6273(03)00146-6
- Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201. doi: 10.1016/s0959-4388(00)00196-3
- 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
- McGlone, M. S., Glucksberg, S., and Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Process.* 17, 167–190. doi: 10.1080/01638539409544865
- Meeter, M., and Murre, J. (2004). Consolidation of long-term memory: evidence and alternatives. *Psychol. Bull.* 130, 843–857. doi: 10.1037/0033-2909.130.6.843
- Mishkin, M., Vargha-Khadem, F., and Gadian, D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus* 8, 212–216. doi: 10.1002/(sici)1098-1063(1998)8:3<212::aid-hipo4>3.0.co;2-1
- Mollica, F., and Piantadosi, S. T. (2019). Humans store about 1.5 megabytes of information during language acquisition. *R. Soc. Open Sci.* 6:181393. doi: 10.1098/rsos.181393
- Moscovitch, M., Cabeza, R., Winocur, G., and Nadel, L. (2016). Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* 67, 105–134. doi: 10.1146/annurev-psych-113011-143733
- Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., and Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr. Opin. Neurobiol.* 16, 179–190. doi: 10.1016/j.conb.2006.03.013
- Nadel, L. (1991). The hippocampus and space revisited. *Hippocampus* 1, 221–229. doi: 10.1002/hipo.450010302
- Nadel, L., and Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol.* 7, 217–227. doi: 10.1016/s0959-4388(97)80010-4
- O'Kane, G., Kensinger, E. A., and Corkin, S. (2004). Evidence for semantic learning in profound amnesia: an investigation with patient H.M. *Hippocampus* 14, 417–425. doi: 10.1002/hipo.20005
- O'Keefe, J., and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.
- Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., and Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *J. Neurosci.* 26, 4596–4601. doi: 10.1523/JNEUROSCI.1923-05.2006
- Pacton, S., Fayol, M., and Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Dev.* 76, 324–339. doi: 10.1111/j.1467-8624.2005.00848.x
- Pawley, A., and Syder, F. H. (1980). "Two puzzles for linguistic theory: natelike selection and natelike fluency," in *Communicative Competence*, eds J. C. Richards and R. Schmidt (London: Longmans), 191–225.
- Pexman, P. M., Holyk, G. G., and Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Mem. Cognit.* 31, 842–855. doi: 10.3758/bf03196439
- Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549. doi: 10.3758/bf03196311
- Piai, V., Anderson, K., Lin, J., Dewar, C., Parvizi, J., Dronkers, N., et al. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proc. Natl. Acad. Sci. U S A* 113, 11366–11371. doi: 10.1073/pnas.1603312113
- Postle, B. R., and Corkin, S. (1998). Impaired word-stem completion priming but intact perceptual identification priming with novel words: evidence from the amnesic patient H.M. *Neuropsychologia* 36, 421–440. doi: 10.1016/s0028-3932(97)00155-3
- Race, E., Keane, M. M., and Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *J. Neurosci.* 31, 10262–10269. doi: 10.1523/JNEUROSCI.1145-11.2011
- Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus* 20, 1263–1290. doi: 10.1002/hipo.20852
- Reber, P., Knowlton, B., and Squire, L. (1996). Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative knowledge. *Behav. Neurosci.* 110, 861–871. doi: 10.1037/0735-7044.110.5.861
- Reilly, J., Peelle, J. E., Garcia, A., and Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: the dynamic multilevel reactivation framework. *Psychon. Bull. Rev.* 23, 1002–1014. doi: 10.3758/s13423-015-0824-5
- Renoult, L., Irish, M., Moscovitch, M., and Rugg, M. D. (2019). From Knowing to Remembering: the Semantic-Episodic distinction. *Trends Cogn. Sci.* 23, 1041–1057. doi: 10.1016/j.tics.2019.09.008
- Ritchey, M., Libby, L., and Ranganath, C. (2015). Cortico-hippocampal systems involved in memory and cognition: the PMAT framework. *Prog. Brain Res.* 219, 45–64. doi: 10.1016/bs.pbr.2015.04.001
- Robin, J., Rivest, J., Rosenbaum, R. S., and Moscovitch, M. (2019). Remote spatial and autobiographical memory in cases of episodic amnesia and topographical disorientation. *Cortex* 119, 237–257. doi: 10.1016/j.cortex.2019.04.013
- 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. *Psychol. Rev.* 111, 205–235. doi: 10.1037/0033-295x.111.1.205
- Rosenbaum, R. S., Gilboa, A., Levine, B., Winocur, G., and Moscovitch, M. (2009). Amnesia as an impairment of detail generation and binding: evidence from personal, fictional, and semantic narratives in K.C. *Neuropsychologia* 47, 2181–2187. doi: 10.1016/j.neuropsychologia.2008.11.028
- Rosenbaum, R. S., Kohler, S., Schacter, D., Moscovitch, M., Westmacott, R., Black, S., et al. (2005). The case of K.C.: contributions of a memory-impaired

- person to memory theory. *Neuropsychologia* 43, 989–1021. doi: 10.1016/j.neuropsychologia.2004.10.007
- Rosenbaum, R. S., Moscovitch, M., Foster, J. K., Schnyer, D. M., Gao, F., Kovacevic, N., et al. (2008). Patterns of autobiographical memory loss in medial-temporal lobe amnesic patients. *J. Cogn. Neurosci.* 20, 1490–1506. doi: 10.1162/jocn.2008.20105
- Rubin, R. D., Schwarb, H., Lucas, H. D., Dulas, M. R., and Cohen, N. J. (2017). Dynamic hippocampal and prefrontal contributions to memory processes and representations blur the boundaries of traditional cognitive domains. *Brain Sci.* 7:82. doi: 10.3390/brainsci7070082
- Rugg, M. D., Johnson, J. D., and Uncapher, M. R. (2015). “Encoding and retrieval in episodic memory: insights from fMRI,” in *Handbook on the Cognitive Neuroscience of Memory*, eds A. Duarte, M. Barense and D. R. Addis (Oxford, UK: Wiley-Blackwell), 84–107.
- Ryan, J. D., Althoff, R. R., Whitlow, S., and Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychol. Sci.* 11, 454–461. doi: 10.1111/1467-9280.00288
- Ryan, L., Cox, C., Hayes, S., and Nadel, L. (2008). Hippocampal activation during episodic and semantic memory retrieval: comparing category production and category cued recall. *Neuropsychologia* 46, 2109–2121. doi: 10.1016/j.neuropsychologia.2008.02.030
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Saffran, J. R., and Wilson, D. P. (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy* 4, 273–284. doi: 10.1207/s15327078in0402\_07
- 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. Lond. B Biol. Sci.* 362, 773–786. doi: 10.1098/rstb.2007.2087
- Schacter, D. L., Addis, D. R., and Buckner, R. (2008). Episodic simulation of future events: concepts, data, and applications. *Ann. N Y Acad. Sci.* 1124, 39–60. doi: 10.1196/annals.1440.001
- Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., and Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* 26, 1736–1747. doi: 10.1162/jocn\_a\_00578
- Schapiro, A. C., Kustner, L. V., and Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* 22, 1622–1627. doi: 10.1016/j.cub.2012.06.056
- Sharon, T., Moscovitch, M., and Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proc. Natl. Acad. Sci. U S A* 108, 1146–1151. doi: 10.1073/pnas.1005238108
- Schlichting, M. L., and Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Curr. Opin. Behav. Sci.* 1, 1–8. doi: 10.1016/j.cobeha.2014.07.005
- Schmolck, H., Kensinger, E., Corkin, S., and Squire, L. (2002). Semantic knowledge in patient H.M. and other patients with bilateral medial and lateral temporal lobe lesions. *Hippocampus* 12, 520–533. doi: 10.1002/hipo.10039
- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–12. doi: 10.1136/jnnp.20.1.11
- Shimamura, A. P., and Squire, L. (1984). Paired-associate learning and priming effects in amnesia: a neuropsychological study. *J. Exp. Psychol. Gen.* 113, 556–570. doi: 10.1037/0096-3445.113.4.556
- Skotko, B., Kensinger, E., Locascio, J., Einstein, G., Rubin, D., Tupler, L., et al. (2004). Puzzling thoughts for H.M.: can new semantic information be anchored to old semantic memories? *Neuropsychology* 18, 756–769. doi: 10.1037/0894-4105.18.4.756
- Solomon, E. A., Lega, B. C., Sperling, M. R., and Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proc. Natl. Acad. Sci. U S A* 116, 24343–24352. doi: 10.1073/pnas.1906729116
- Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys and humans. *Psychol. Rev.* 99, 195–231. doi: 10.1037/0033-295x.99.2.195
- Squire, L. R., and Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proc. Natl. Acad. Sci. U S A* 93, 13515–13522. doi: 10.1073/pnas.93.24.13515
- Squire, L. R., and Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus* 8, 205–211. doi: 10.1002/(sici)1098-1063(1998)8:3<205::aid-hipo3>3.0.co;2-i
- Staresina, B., and Davachi, L. (2009). Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron* 63, 267–276. doi: 10.1016/j.neuron.2009.06.024
- Stark, S. M., Stark, C. E., and Gordon, B. (2005). New semantic learning and generalization in an amnesic patient. *Neuropsychology* 19, 139–151. doi: 10.1037/0894-4105.19.2.139
- St-Laurent, M., Moscovitch, M., Jadd, R., and Mcandrews, M. P. (2014). The perceptual richness of complex memory episodes is compromised by medial temporal lobe damage. *Hippocampus* 24, 560–576. doi: 10.1002/hipo.22249
- Taler, V., Kousaie, S., and López Zunini, R. (2013). ERP measures of semantic richness: the case of multiple senses. *Front. Hum. Neurosci.* 7:5. doi: 10.3389/fnhum.2013.00005
- Thompson-Schill, S. (2003). Neuroimaging studies of semantic memory: inferring “how” from “where”. *Neuropsychologia* 41, 280–292. doi: 10.1016/s0028-3932(02)00161-6
- Tulving, E. (1972). “Episodic and semantic memory,” in *Organization of Memory*, eds E. Tulving and W. Donaldson (New York, NY: Academic Press), 381–403.
- Tulving, E., and Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus* 8, 198–204. doi: 10.1002/(sici)1098-1063(1998)8:3<198::aid-hipo2>3.0.co;2-g
- Tulving, E., Hayman, C. A., and Macdonald, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: a case experiment. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 595–617. doi: 10.1037/0278-7393.17.4.595
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connely, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science* 277, 376–380. doi: 10.1126/science.277.5324.376
- Verfaellie, M., Bousquet, K., and Keane, M. M. (2014). Medial temporal and neocortical contributions to remote memory for semantic narratives: evidence from amnesia. *Neuropsychologia* 61, 105–112. doi: 10.1016/j.neuropsychologia.2014.06.018
- Verfaellie, M., Koseff, P., and Alexander, M. P. (2000). Acquisition of novel semantic information in amnesia: effects of lesion location. *Neuropsychologia* 38, 484–492. doi: 10.1016/s0028-3932(99)00089-5
- Wang, J., Cohn, N. J., and Voss, J. (2015). Covert rapid action-memory simulation (CRAMS): a hypothesis of hippocampal-prefrontal interactions for adaptive behavior. *Neurobiol. Learn. Mem.* 117, 22–33. doi: 10.1016/j.nlm.2014.04.003
- Warren, D. E., and Duff, M. C. (2014). Not so fast: hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus* 24, 920–933. doi: 10.1002/hipo.22279
- Warren, D. E., and Duff, M. C. (2019). Fast mappers, slow learners: word learning without hippocampus is slow and sparse irrespective of methodology. *Cogn. Neurosci.* 10, 210–212. doi: 10.1080/17588928.2019.1593120
- Wilson, B., and Moffat, N. (1983). “Rehabilitation of memory for everyday life,” in *Everyday Memory: Actions and Absent-Mindedness*, eds J. E. Harris and P. Morris (London: Academic Press), 207–233.
- Wojcik, E. H. (2018). The development of lexical-semantic networks in infants and toddlers. *Child Dev. Perspect.* 12, 34–38. doi: 10.1111/cdep.12252
- Wojcik, E. H., and Saffran, J. R. (2013). The ontogeny of lexical networks: toddlers encode the relationships amongst referents when learning novel words. *Psychol. Sci.* 24, 1898–1905. doi: 10.1177/0956797613478198
- Wojcik, E. H., and Saffran, J. R. (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition* 138, 10–20. doi: 10.1016/j.cognition.2015.01.015
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00072
- Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behav. Brain Res.* 254, 34–44. doi: 10.1016/j.bbr.2013.05.030

- Yonelinas, A. P., Ranganath, C., Ekstrom, A., and Wiltgen, B. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nat. Rev. Neurosci.* 20, 364–375. doi: 10.1038/s41583-019-0150-4
- Zacks, J. M., Speer, N. K., and Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *J. Exp. Psychol. Gen.* 138, 307–327. doi: 10.1037/a0015305
- Zacks, J. M., and Swallow, K. M. (2007). Event segmentation. *Curr. Dir. Psychol. Sci.* 16, 80–84. doi: 10.1111/j.1467-8721.2007.00480.x
- Zettersten, M., Wojcik, E., Benitez, V., and Saffran, J. (2018). The company objects keep: linking referents together during cross-situational word learning. *J. Mem. Lang.* 99, 62–67. doi: 10.1016/j.jml.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 © 2020 Duff, Covington, Hilverman and Cohen. 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.





# Verbal Working Memory as Emergent from Language Comprehension and Production

Steven C. Schwering and Maryellen C. MacDonald\*

Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States

## OPEN ACCESS

### Edited by:

Arthur M. Jacobs,  
Freie Universität Berlin, Germany

### Reviewed by:

Christos Sallis,  
Newcastle University,  
United Kingdom  
Melissa Duff,  
Vanderbilt University Medical Center,  
United States

### \*Correspondence:

Maryellen C. MacDonald  
mcmacdonald@wisc.edu

### Specialty section:

This article was submitted to Speech and Language, a section of the journal *Frontiers in Human Neuroscience*

**Received:** 31 July 2019

**Accepted:** 13 February 2020

**Published:** 12 March 2020

### Citation:

Schwering SC and MacDonald MC (2020) Verbal Working Memory as Emergent from Language Comprehension and Production. *Front. Hum. Neurosci.* 14:68. doi: 10.3389/fnhum.2020.00068

This article reviews current models of verbal working memory and considers the role of language comprehension and long-term memory in the ability to maintain and order verbal information for short periods of time. While all models of verbal working memory posit some interaction with long-term memory, few have considered the character of these long-term representations or how they might affect performance on verbal working memory tasks. Similarly, few models have considered how comprehension processes and production processes might affect performance in verbal working memory tasks. Modern theories of comprehension emphasize that people learn a vast web of correlated information about the language and the world and must activate that information from long-term memory to cope with the demands of language input. To date, there has been little consideration in theories of verbal working memory for how this rich input from comprehension would affect the nature of temporary memory. There has also been relatively little attention to the degree to which language production processes naturally manage serial order of verbal information. The authors argue for an emergent model of verbal working memory supported by a rich, distributed long-term memory for language. On this view, comprehension processes provide encoding in verbal working memory tasks, and production processes maintenance, serial ordering, and recall. Moreover, the computational capacity to maintain and order information varies with language experience. Implications for theories of working memory, comprehension, and production are considered.

**Keywords:** working memory, language comprehension, language production, serial order, long-term memory, lexical representations

## INTRODUCTION

When Ebbinghaus (1885) published his extensive verbal memory experiments and observations, he established a new theoretical approach to cognitive psychology through the formal study of memory. In his quest to isolate the properties of memory, Ebbinghaus observed that immediate recall of verbal material was utterly contaminated by long-term knowledge of the language. He found it impossible to isolate immediate memory when he probed recall of meaningful verbal memoranda such as lines of poetry or narratives, and he established critical methodological practices aimed at stripping away confounding factors. In his attempt to isolate immediate memory, Ebbinghaus developed a collection of nonwords, thousands of consonant-vowel-consonant

syllables that could be used to construct lists for immediate recall. The contamination of long-term experience persisted, as certain nonwords exhibited “very important and almost incomprehensible variations as to the ease or difficulty with which they are learned” (p. 23). Moreover, Ebbinghaus noted that even these novel materials could not completely isolate immediate memory from other cognitive processes; visual, acoustic, and articulatory components of verbal perception and action necessarily affected task performance.

Over 130 years of research now contributes to answering the questions posed by Ebbinghaus, and it is useful to ask how his catalyzing observations continue to influence theoretical and methodological approaches to memory research. In this article, we critically analyze Ebbinghaus’s goal of isolating immediate memory as well as his warning that such isolation may be impossible. Following some establishment of terms and definitions and a brief sketch of some current models of immediate memory, we consider several intersecting points, all of which stem from a language-based perspective on the ability to temporarily maintain verbal information. First, we consider the dependence of immediate memory on long-term language knowledge, as Ebbinghaus first observed, and consider the impact of these relationships on modern theories of working memory. These modern accounts recognize some role for long-term memory, but we argue that they have been slow to embrace more modern approaches to the nature of long-term word representations and processing. Instead, we argue that language comprehension and production processes underpin encoding, maintenance, and production of old and new verbal memoranda without the need for separable buffers that are common in some current memory models. A key development in some models of immediate memory is the assumption that memory for words is separate from memory for their orders. In contrast, we consider the many ways in which various word and order representations are intertwined in language comprehension and production research and propose a new emergent account that incorporates these representations in VWM. In closing, we consider the implications of our perspective on theories of language use and on related research areas.

## WORKING MEMORY MODELS AND TERMINOLOGY

There exists a fundamental disagreement about the definition of working memory (e.g., Cowan, 2008; Aben et al., 2012), as evidenced by a wide array of both qualitative descriptions of immediate memory and competing memory models (see Cowan, 2017). We will focus on two general classes of models for how humans can encode verbal material, maintain it for a brief period of time, and produce the memoranda by speaking or writing. Proponents of the two types of models that we discuss, the multi-component models (e.g., Baddeley et al., 1975; Baddeley, 2000) and emergent models (e.g., Cowan, 1993; Postle, 2006), do not always use terms in the same way, and so we begin with some definitions.

Verbal working memory (VWM) is commonly viewed as the temporary maintenance of verbal information (i.e., some

aspects of language). Some researchers distinguish VWM as an immediate memory for processing of information (converting speech to meaning, say) from short-term memory (STM), a passive temporary store. However, as Buchsbaum and D’Esposito (2019) have noted, information is always being transformed in some way in the service of goal-directed behavior, and so we will use the term VWM to refer to both storage and processing, except where we specifically refer to theories invoking an STM component. Finally, VWM researchers have increasingly investigated the ability to recall verbal material in the same order it was presented. Thus, we discuss abilities to recall a word or nonword (termed *item memory*) and recall in the correct order in a list (*order memory*).

Multicomponent models, which get their name from the distinct components posited in the working memory system (Baddeley, 1992), draw a sharp distinction between passive storage of information in “buffers” and processing mechanisms such as speech perception and production processes. In this respect, multicomponent models are aligned with classical theories of working memory advanced by Ebbinghaus. In this view, the sole function of STM is to act as a site of storage. Specifically, multicomponent models posit a short-term buffer that maintains a rapidly degrading representation of memoranda (Baddeley et al., 1984). Critically, in this perspective, long-term memory is separate from STM (e.g., Shallice and Warrington, 1970, 1974), but *via* a process called *redintegration* (e.g., Hulme et al., 1997), LTM can provide cues to rebuild STM as it degrades (Lewandowsky and Farrell, 2000). LTM can interact with STM in other ways. With respect to language processing, some researchers claim that verbal STM is a buffer that stores partially processed linguistic representations (e.g., Martin and Romani, 1994; Martin and He, 2004), or is a specific subcomponent of language processing mechanisms dedicated to storage (Shallice and Papagno, 2019). Certain theories propose that the buffer holds copies of or pointers to representations derived from LTM that may require further processing in the future (Norris, 2017). Thus, whereas Ebbinghaus (1885) tried to isolate STM processes within an interacting system, the multi-component models have converted that research goal into an architectural claim: STM is a distinct system with only the most limited, indirect contact with LTM and language processing mechanisms.

Although multicomponent accounts are the dominant perspective in VWM research, there is a long history of caution about this approach. More than 25 years ago, Crowder (1993) predicted a wholesale reassessment of multi-component models of VWM in favor of alternative approaches. He described the notion of a separate, dedicated short-term store (the multicomponent model) as “archaic and, to some of us, even downright quaint” and suggested that “Increasingly, the field is turning instead to a procedural attitude toward memory” (p. 143). Crowder’s predictions were wildly inaccurate in their timeline, as multi-component models of memory remain important and useful theories of VWM now many decades after Crowder predicted their demise. Nevertheless, Crowder correctly predicted the rise of alternative, emergent models of VWM that did away with separate buffers.

Emergent approaches do not generally distinguish between storage and processing mechanisms. Some earlier variants were called procedural models, defining VWM as a secondary product of procedures in support of other cognitive processes (Craik and Lockhart, 1972; Kolers and Roediger, 1984; Crowder, 1993; Jones et al., 2004). Early theorizing by Saffran and Martin (1997) explored relationships between aphasic patients' VWM in the context of their language production abilities, informed by Dell (1986) spreading activation model of language production (Martin et al., 1996; Saffran and Martin, 1997). We advocate this "rich emergent" approach here, where VWM is the activated portion of linguistic LTM (Cowan, 1993; Postle, 2006; Acheson and MacDonald, 2009a,b; Hasson et al., 2015; MacDonald, 2016; Buchsbaum and D'Esposito, 2019). This approach emphasizes VWM as a complex of skills, honed by past language comprehension and production experience. In this view, knowledge of word meanings and other forms of linguistic knowledge shape performance in VWM tasks. Performance on VWM tasks co-opts language LTM, by which we mean any parts of LTM involved in language tasks, including knowledge of events, word meanings, word order, phonological form, and other information (MacDonald and Christiansen, 2002; Acheson and MacDonald, 2009b; MacDonald, 2016). LTM itself is characterized as a set of processing mechanisms employed to achieve goal-directed behavior rather than store a static set of memoranda chunked or compressed from prior experience (Postle, 2006; Buchsbaum and D'Esposito, 2019). In the case of WM for linguistic memoranda, we have proposed that the language production architecture is co-opted to maintain and order the memoranda, obviating the need for a separate memory buffer (Acheson and MacDonald, 2009b; MacDonald, 2016). Whereas, in the multicomponent model, effects of prior language knowledge in LTM have been attributed to secondary mechanisms (e.g., Hulme et al., 1997; Lewandowsky and Farrell, 2000), we see these LTM effects arising naturally from language production and comprehension processes. For example, language production is well known to favor serial orders that have been used frequently or recently (Bock, 1986a) and to group related words together in an utterance (Solomon and Pearlmutter, 2004). These biases in production may underlie the effects of semantic grouping and similarity to natural language that has been observed in recall tasks (Miller and Selfridge, 1950; Jones and Farrell, 2018). Thus, we view temporary maintenance and ordering as the job of action systems, which must construct an action plan and maintain it before it can be executed, so that the action plan is the "memory of what is to come" (Rosenbaum et al., 2007, p. 528). For language, the action planning system is language production, and the utterance plan is the memory of both what is to be produced and the order in which it will be produced at several levels, including words, phonemes, and articulatory gestures (Martin et al., 1996; Acheson and MacDonald, 2009b; MacDonald, 2016). In this view, VWM is simply the skill of maintaining and ordering linguistic material, and that skill, as with all subcomponents of language production and comprehension, emerges from actions of the language systems and varies with experience (MacDonald and Christiansen, 2002; MacDonald, 2016).

In contrast to the "rich emergent" account described above, some "limited emergent" accounts posit a more restricted interaction with language processes, with different systems working in parallel to support memory for items and their orders (Majerus, 2013, 2019). On this view, item memory engages ventral language pathways that process semantics, with dorsal pathways supporting order within the item (i.e., phonemes). In contrast, order memory for sequences of words themselves engages frontal-parietal networks and networks closely associated with attentional mechanisms. The item/order memory distinction has been supported by findings that word characteristics, like frequency of use (Poirier and Saint-Aubin, 1996; Saint-Aubin and Poirier, 1999) and semantics (Majerus and D'Argembeau, 2011), largely affect memory for items but not memory for order. Furthermore, memory for items and order appear to engage distinct neural populations, as indicated by neuroimaging results (Majerus et al., 2006, 2008; Guidali et al., 2019) and aphasic patient data (e.g., Majerus et al., 2007, 2015).

The separate item/order memory of more limited emergent accounts is consistent with a multicomponent approach, namely that LTM is able to support STM only in cases where the items and order conform to prior experience. Multicomponent models are particularly emphatic about this point, arguing that this is a critical reason an STM buffer must exist distinct from LTM (e.g., Norris, 2017). Some emergent accounts also recognize that there are limitations to LTM. For example, Majerus (2013) suggests that "the representations of the language system are able to support familiar item and order information, but not unfamiliar order information" (p. 4). This distinction between familiar and unfamiliar orders is problematic because it presumes a dichotomy between the novel and familiar when similarity to prior experience is actually continuous. We consider this point further in the section entitled "Problems with Limited Emergence."

In the next sections, we contrast our rich emergent account against a variety of alternative multi-component and more limited emergent memory models. Specifically, we describe current research on the nature of LTM language representations and the language comprehension and production processes that interact with LTM. Because all accounts of VWM must refer in some way to LTM, we argue that this characterization of language knowledge informs all theories of encoding, maintaining, and ordering verbal information.

## WORD REPRESENTATIONS IN VWM AND LANGUAGE RESEARCH: NO WORD IS AN ISLAND

Since the time of Ebbinghaus, most VWM models have assumed discrete representations or "items" in memory. Often, verbal memory is conceptualized by the unit of the word or word-like collections of phonemes (nonwords). For example, there are a multitude of studies investigating immediate or delayed word recall that document word accuracy across list position (e.g., Murdock, 1962; Watkins and Watkins, 1977), word omissions

(e.g., Roodenrys et al., 2002), word intrusions (e.g., Coltheart, 1993), and so on. Furthermore, measurement of VWM capacity is often indexed by list span, or the average number of words recalled from lists (e.g., Daneman and Carpenter, 1980; Hulme and Tordoff, 1989). In part, such descriptions are a convenient shorthand for bits of information (Miller, 1956), but they also reflect certain assumptions about the isolability of memory representations. One common assumption is that word memory is supported by fully separable phonological and semantic codes (Martin, 1987; Martin et al., 1999; Howard and Nickels, 2005). Another is that order memory is separable from the memory for the word, itself; this view is further compounded by viewing the words in lists as separate from each other, especially in the case of novel word orders (Majerus, 2013, 2019).

Considering that all major memory models posit some kinds of ties with language representations, it bears asking how a compartmentalized view of item and order representations, and a compartmentalized view of item components (e.g., phonology, semantics, grammatical role), accords with language research. In this section, we describe developments in both comprehension and production research that is completely antithetical to the isolated representations prevalent in much memory research. This work shows that different levels of language representation used in production and comprehension, what we refer to as language LTM, influence each other and are integrated. We suggest that this integration, and the statistical regularities between classically defined and supposedly dissociable representations that are critical for language research, have significant consequences for how verbal information is maintained. In other words, we argue that the nature of linguistic LTM representations, as revealed in research on language comprehension and production, is highly relevant to theories of VWM.

## Integrated Representations in Language Processing

Researchers' views about the nature of word representations and their use in comprehension and production have undergone enormous change in the last several decades. Initially, researchers believed that comprehension processes were modular, such that dedicated components worked independently to interpret language input (e.g., Seidenberg et al., 1982; Frazier, 1987; see also Almeida and Gleitman, 2018 for more historical context and current views of modularity). Similarly, models of production were highly staged, with minimal interaction between different language representations (e.g., Levelt et al., 1999). Theories of word representation pointed to a lexicon with distinct levels (phonological, syntactic, semantic, e.g., Allport and Funnell, 1981). Importantly, these models assumed that, regardless of the nature of LTM, language processes could selectively extract and operate over subcomponents of linguistic knowledge, such as processing phonology or syntax without meaning, with some later integration stage (Forster, 1985; Frazier, 1987). While this work did not often invoke VWM, the notions of separable language components and isolated processing systems are compatible with the orientation of multi-component models.

More recent theories of language comprehension are far less aligned with these compartmentalized approaches. Instead, they have emphasized extensive interaction between different kinds of language representations. This is most clearly demonstrated behaviorally in instances where certain information cannot be "turned off," even when it is beneficial to do so (e.g., Stroop, 1935). For example, Seidenberg and Tanenhaus (1979) demonstrated that the orthographic form of a word interfered with judgments of phonological form, meaning that one form of information in LTM (orthographic information) interfered with another form of information in LTM (phonological form). While early neuropsychological studies suggested that the subcomponents of language knowledge were represented with discrete neural codes (Dapretto and Bookheimer, 1999), more recent analyses support integrated representations. For example, Siegelman et al. (2019) argue against previous evidence for divisions between syntactic and semantic representations during sentence comprehension. Similarly, Dikker et al. (2010) found that phonological/orthographic information contributes to syntactic analyses within 100 ms, even before a word has been recognized because the phonological form is correlated with, and therefore provides information about, the likely grammatical category (noun, verb, etc.) of the to-be-recognized word. Together, this work and others (e.g., Pereira et al., 2018) suggest that word comprehension and LTM representations are much more interconnected than was previously recognized.

This article is not the place for a full specification of how representations are integrated, nor for the natural ongoing debates concerning how to characterize linguistic knowledge, but it is worth noting why a number of researchers now assume extensive interaction and integration among what has been traditionally described as distinct levels of linguistic information. In more integrated accounts, multiple sources of information interact in perception and comprehension because interactions are beneficial, essential really, to comprehend and produce language in real-time. Language contains strong correlations between different levels of representation, between language and the world, and between information earlier and later in a linguistic signal to be interpreted. People are voracious statistical learners, and they leverage their LTM of the statistical regularities between different kinds of information to comprehend and produce language efficiently and accurately (Seidenberg and MacDonald, 2018). Indeed, the combination of several partially informative information sources (phonology and semantics, for example) is now seen as central to accounting for the speed with which comprehenders interpret incoming language input despite the massive ambiguity known to pervade language; an individual source of information only weakly constrains interpretation alone but is highly effective in combination with other constraints (Seidenberg, 1997; MacDonald and Seidenberg, 2006; Graves et al., 2010; Joanisse and McClelland, 2015). Each language comprehension experience is a source of learning (Chang et al., 2000), and a consequence of learning all this combinatorial information is that any single source of information, including words, cannot be atomic or isolated (Willits et al., 2015). Instead, words and other classically defined levels of representation are highly intertwined, because



learning (and therefore LTM) must capture a complex web of statistical structure to maximize performance during language comprehension and production. Word representations can be modeled as attractors in networks comprising various types of information (phonological, semantic, etc., Hinton and Shallice, 1991), and some linguists and psycholinguists now consider discrete notions such as *word* and *phoneme* to be convenient fictions, highly useful for researchers' discussions but having more to do with people's conscious intuitions than with the way that language is actually represented and processed in the brain (Bybee and McClelland, 2005; Baayen et al., 2016; Ramscar and Port, 2016).

## Separated Representations in Memory-Models

These highly interactive approaches have not yet penetrated much of the theorizing in most current multi-component and emergent models of VWM, which continue to emphasize individual "items" of memory. Multicomponent models posit specialized, separate buffers, such as the phonological loop (Baddeley and Hitch, 1974), which encode a single type of information. Initially, patient lesion data seemed to provide further support to modular memory and language approaches, as in patients who exhibited impaired memory abilities with spared language abilities (often called "STM patients," e.g., Warrington and Shallice, 1969) and in cases reporting double dissociations of phonological and semantic information in memory and language tasks, leading to a separation between phonology and semantics in multicomponent models (Martin and Romani, 1994; Martin et al., 1994). This dissociation between representations extends into memory for items and their order. Certain aphasic patients demonstrate apparently isolable item or order memory impairments (Attout et al., 2012; Majerus et al., 2015), and this behavioral pattern is accompanied by neuroimaging evidence suggesting item and order memory are supported by distinct neural populations (Kalm and Norris, 2014; Attout et al., 2019).

A strict notion of "item" in memory becomes more complicated when considering the qualities of statistical information in linguistic LTM. For example, phonotactic long-term knowledge influences recall of novel words. Non-words consistent with the transitional probabilities of phonemes (or acoustic properties or articulatory gestures) in natural language are recalled better than non-words inconsistent with these patterns (Gathercole et al., 1999; Thorn and Frankish, 2005). Researchers have likewise extended these findings to suggest that both lexical and sublexical properties affect recall of non-words (Roodenrys et al., 2002; Majerus et al., 2004). Tanida et al. (2019) further demonstrated an effect of forward and backward bimora transition probabilities on ordered recall. Together, these results suggest that memory of one phoneme or acoustic pattern influences memory of others *via* LTM of the phonological statistical structure of language. These "neighborhoods" of patterns in LTM can be quite subtle, as evidenced by the improved recall for nonwords with regular pitch accent compared to irregular pitch accent, an effect moderated by phonotactic frequency (Tanida et al., 2015; see

also Yuzawa and Saito, 2006). Not only do these studies suggest that LTM is relevant for VWM, but they suggest multiple grain sizes of phonological information interact to inform performance in memory tasks.

Beyond phonological information, language users also track and leverage complex statistical regularities between different types of linguistic representations, such as between phonology and semantics. Our claim is not that phonology and semantics are completely merged (they are clearly not), but rather that they are intertwined to a degree that affects language use and VWM performance. Such regularities are not always obvious. Indeed, with some exceptions (Farmer et al., 2006; Schmidtke et al., 2014; Christiansen and Monaghan, 2016), the mapping between phonology and semantics seems largely arbitrary. If phonology and semantics were completely distinct, then each representation could be stored in a separable buffer, consistent with multicomponent accounts. However, claims for a strict semantic-phonological divide break down when considering morphologically complex words, such as *painter*, *ideas*, *friendship*, and *working*. These words contain morphemes (*-er*, *-s*, *-ship*, *-ing*) for which the mapping from phonology to semantics is not arbitrary. The same mapping occurs repeatedly through the language (e.g., *worker*, *baker*, *seeker*, etc.), and words sharing these affixes form semantic-phonological neighborhoods that shape language LTM and behavior (Rueckl et al., 1997; Seidenberg and Gonnerman, 2000). These relationships also encode grammatical form (e.g., *-er* is associated with nouns, *-ing* with verbs). It might be tempting to consider morphologically complex words as marginal and not part of more "typical" language, but morphologically complex words are common in English and their phonological-semantic-grammatical regularities have been shown to affect word learning in infants (Willits et al., 2014). In adults, regularities between phonological, orthographic, semantic, and grammatical knowledge drive very early stages of language comprehension, even before conscious word recognition (Dikker et al., 2010). Even so, recent reviews suggest there is a "notorious lack of consensus" (p. 37) in the imaging literature about the brain representations of phonological, semantic, and morphological relationships among more complex words (Leminen et al., 2019). As such, it is clear that many representations simultaneously impact language comprehension and production, and it is unclear how any single representation could be extricated from this web of processing.

Given these regularities in language use, it is not surprising that morphophonological regularities also impact VWM. For example, the use of morphophonological cues has been well-studied in children's nonword repetition. Nonwords with morphophonological cues are recalled better than nonwords without such cues, and children with language impairments may be less sensitive to this effect (Archibald and Gathercole, 2006; Casalini et al., 2007; Estes et al., 2007). Thus, experience with language, specifically the regular co-occurrences between phonology and semantics in morphologically complex words, affects VWM for nonwords (though see Szwedczyk et al., 2018). These results have largely been examined with children completing single word repetition tasks. It would be worthwhile

to extend this work to other tasks and populations. Incorporating regularities between phonology and semantics in stimuli (e.g., *via* the use of affixes) could alter the apparent separability of phonology and semantics, as has been suggested by many memory and language studies (e.g., Martin et al., 1994).

The “primary systems” approach to memory and language use begins to incorporate some current insights about language representations and argues for phonology and semantics as separable yet interacting representations (Ueno et al., 2014; Savill et al., 2019). Broadly, this approach supports emergent memory accounts, suggesting that the effects of semantics and phonology on word and non-word recall reflect a balance of processing. For example, when phonological support is weak, semantic support affects recall to a larger degree compared to when phonological support is strong (Savill et al., 2019). In such accounts, the interactions between phonology and semantics emerge from processing in a quasi-regular domain, resulting in integrated representations. Ueno et al. (2014) demonstrated that words with low imageability are recalled worse than words with high imageability (i.e., the effect of semantics), and this effect is exacerbated in words with an atypical pitch accent (i.e., effect of phonotactics). In line with the primary systems account, this suggests that the effect of phonotactics on recall depends in part on semantics. Interestingly, the researchers developed a neurobiologically constrained connectionist model of word comprehension, repetition, and production, demonstrating that phonological (ventral) and semantic (dorsal) language pathways are differentially engaged when processing typical and atypical phonotactic patterns. As a result, the semantic pathway was more engaged in processing atypical phonotactic patterns. Such research suggests that subtle phonological information may infiltrate a putative semantic pathway (see also Jefferies et al., 2005).

The tracking of complex statistical patterns in support of language comprehension, production, and memory is not limited to within-word representations like phonology and semantics; statistical regularities also support the representation of word order. This point gets to the heart of the item vs. order distinction in VWM theorizing. Memory researchers readily agree that sentences are recalled better than scrambled lists of words (Brenner, 1940), and this effect scales with list approximation to natural language sequence statistics (Miller and Selfridge, 1950). These effects are typically attributed to semantic coherence or episodic pattern recognition (Baddeley et al., 2009; Allen et al., 2018). However, episodic memory is not sufficient to explain the full range of results. Memory is similarly facilitated for lists of non-words that approximate natural language syntax (Epstein, 1961, 1962). Thus, meaning does not seem to be necessary for the effect. Jones and Farrell (2018) further demonstrated that people are more likely to recall sentence-like lists in an order consistent with syntactic knowledge and that errors are more likely to conform to prior syntactic knowledge than expected by chance (for corpus analyses tying language experience to memory performance, see Perham et al., 2009; Jones et al., 2020). In each case, inter-item information affected memory for order *via* long-term knowledge of language syntax, suggesting that memory for items and their order interact to support each

other. For example, experience using English builds an LTM of the word *pull*. The LTM of *pull* not only encodes meaning and sound but also co-occurrence tendencies; *pull* is often flanked by words denoting animate entities and objects involved in a pulling event (as in *The girl pulled the cart*). We are emphatically not claiming that linguistic knowledge is limited to co-occurrence, merely that such knowledge includes linear relationships and that what might be viewed as multi-word frequency knowledge shapes both language use (Seidenberg and MacDonald, 2018) and memory (Arnon and Snider, 2010). While strict chaining accounts of ordering have generally fallen out of favor in memory research (e.g., Hurlstone et al., 2014), these studies suggest that inter-item associations are not only encoded and leveraged for performance in memory tasks (for discussion, see also Fischer-Baum and McCloskey, 2015) but reinforced by LTM. Such effects are likely amplified by the presence of multi-morphemic words (such as *pulled*), because, as noted above, morphemes such as *-ed* also contain grammatical information and provide cues to inter-word relationships (see Epstein, 1961, 1962). Thus, it is unclear to what extent item knowledge can be separated from order knowledge if the source of the order benefit is derived from the information associated with the individual words.

## The Role of Language Processes in Performing VWM Tasks

If performing a VWM task is dependent on language processes, such as comprehension for encoding (MacDonald and Christiansen, 2002), lexical production for item memory (Page et al., 2007), or sentence production skills for item ordering (Acheson and MacDonald, 2009b; MacDonald, 2016), then theories of VWM must consider how theories of language comprehension and production constrain memory performance. Here, we describe some current models of language comprehension and production with a specific eye toward describing statistical regularities in language and the integrated representations in LTM that capture those regularities. Of course, these models were not explicitly designed to model performance in VWM tasks. There is an essential tension between the complexity of LTM representations and modeling: the more complex and intertwined the representations are thought to be, the more difficult it is to capture this complexity in a computational model. Few explicit emergent models of VWM exist, as some researchers have noted (Norris, 2017), though many models adopt principles consistent with the emergent approach (e.g., Botvinick and Plaut, 2006). However, from the language emergent perspective, theories of language comprehension and production should serve as a useful analog, continuing the role models of language use have played in shaping memory research (e.g., Martin et al., 1994).

In this view, language comprehension and production processes underlie the encoding and retrieval mechanisms posited in memory accounts, respectively. Language comprehension processes extract meaning from input by mapping an input signal to a semantic representation of the entities and events being referred to (MacDonald and Hsiao, 2018). Often, comprehension processes involve partial

predictions of upcoming input (Federmeier, 2007; Altmann and Mirković, 2009; Kuperberg and Jaeger, 2016), which means that comprehension processes routinely involve not only semantic integration of words that have been encountered but also generation of serial order expectations among representations of words that are likely upcoming in the input. Similarly, the interpretation of some language input can depend on the material that comes later (Connine and Clifton, 1987; MacDonald, 1994). There are many language comprehension models that depend on integrated representations, variously capturing word segmentation (Christiansen et al., 1998), utterance interpretation without a separate word segmentation stage (Baayen et al., 2016), the learning of phonological forms (Plaut and Kello, 1999), word reading and its relationship to phonology (Seidenberg and McClelland, 1989; Plaut et al., 1996), the learning of grammatical knowledge (Allen and Seidenberg, 1999), behavior in the visual world paradigm (Mayberry et al., 2009), disorders of comprehension in individuals with developmental language disorder (also called specific language impairment, Joanisse and Seidenberg, 2003), and more. In turn, language production models attempt to generate a well-formed utterance from a message representation, either externally motivated in the case of a repetition task or internally generated in the case of self-generated production. Several interactive models exist, capturing lexical selection (i.e., retrieving words from LTM, Dell et al., 1997) and phrase (Dell et al., 1997) or sentence production (Chang et al., 2006; Dell and Chang, 2014). The Lichtheim-2 model implements an account of single-word comprehension and repetition as well as the degradation of those processes in aphasia (Ueno et al., 2011). All of these models share several core features that tie them to the emergent account. In each, learning algorithms, such as backpropagation, encode statistical knowledge in the connection weights updated through experience, forming the model's LTM. Each of these models also develops a VWM through learning; for example the TRACE model of speech perception (McClelland and Elman, 1986) got its name from the claim that the STM trace of the model emerged from the interacting layers of the network. No separable STM buffers divorced from LTM are employed in any of the above models.

Critically, integrated representations are a core part of these language models, most commonly instantiated as distributed representations in a network. Distributed representations as their name implies, spread a representation over the entire network *via* connection weights between layers. Integrated representations exhibit at least two key ties to distributed representations in connectionist language models. First, integrated representations emerge in processing *via* bidirectional spreading activation between layers, a feature evident in models of human comprehension and production (e.g., Dell, 1986; Seidenberg and McClelland, 1989). Second, the integrated representations blend processed information across the network such that phonological, semantic, lexical, and grammatical information cannot be strictly separated from other types of information (e.g., McClelland et al., 2010). Of course, we are not claiming that language models do not develop certain specializations for phonological, semantic, lexical, grammatical, and other types of

information. Instead, specialization is a matter of degree, where complete modularity and complete overlap are less likely than an intermediate state (McClelland et al., 2010). For example, in some models, impairments of a discrete representation (e.g., phonology) disrupt the use of other representations (e.g., semantics), *via* layers that allow interaction between those representations (e.g., Monaghan and Woollams, 2017). Such models are most consistent with primary systems accounts (e.g., Ueno et al., 2014; Savill et al., 2019). In other models, the integrated representations are not as explicit. For example, simple recurrent networks of comprehension and production, allow information to be processed through time. Such networks cross item and order information *via* recurrent connections (Elman, 1991; Joanisse and Seidenberg, 2003; Botvinick and Plaut, 2006), and there is no clear way in which item and order information can be separated.

Distributed representations as they are captured in connectionist models are not the only way to characterize integrated representations. We have focused on variations in distributed connectionist approaches as examples that most clearly embrace the interconnected representations that should affect theorizing about VWM, but other computational approaches could also incorporate integrated representations in processing (e.g., Frank and Goodman, 2014). Furthermore, localist representations, like the one implemented in Dell et al. (1997), also have interaction among different types of information and have proven incredibly useful in describing mechanisms by which LTM engages with VWM.

## Potential Research Directions and Predictions for a Language-Emergent VWM

There are several predictions for VWM research that stem from the language emergent view, the first of which emphasizes the role of language production processes in the serial ordering of the items in a memory list. Previous research has argued that production processes are engaged in maintenance and recall of verbal material, specifically that the utterance plan that maintains the to-be-uttered words in order also serves the maintenance and ordering functions during VWM tasks (Acheson and MacDonald, 2009b; MacDonald, 2016). As MacDonald (2016) discussed, this claim is much more controversial for some kinds of VWM tasks and performance than others. For example, Page et al. (2007) posited a limited role for language production processes in ordering at the item level. They argued that parallels between word production processes and word recall in VWM tasks pointed to individual, word-level utterance plans playing a role in phonological maintenance in VWM, but ordering the words themselves (order memory) must be the purview of a dedicated short-term store. Lombardi and Potter (1992) and Potter and Lombardi (1998) hypothesized a different role for language processing: in VWM tasks involving whole sentence repetition, the comprehension system interprets the meaning of the sentence and the production system regenerates it from that meaning. The model we advocate incorporates the language



system for remembering individual words, whole sentences, as well as all cases in-between, including ordering of word sequences that are less than full, coherent sentences. As there are very few tests of these ideas in the existing literature, our discussion addresses the kinds of word-ordering phenomena in language production that may be relevant to performance in VWM tasks.

An essential task in language production is the creation of serial order over many levels, including messages, words, sub-lexical forms such as phonemes, and articulatory gestures that enable overt language (Dell et al., 1997). Acheson and MacDonald (2009b) extensively reviewed how the interactivity of phonological information with other information predicted serial order phenomena through the lens of language production research. They concluded that "...one key insight about the serial ordering of verbal information in language production is that serial ordering results from interactions across multiple levels of representation over time, that is to say, as a result of recurrent connectivity" (p. 54). For example, word ordering in language production is more likely to go awry when words share features, including both grammatical features (e.g., noun) and phonological features (Dell and Reich, 1981), meaning that phonological and lexico-grammatical information are together affecting serial ordering processes. Given Acheson and MacDonald's review, we do not focus on phonological interactions with word order here, but it is worth noting a few more recent phenomena relevant to their claims. A number of studies have investigated semantic-phonological interactions termed semantic binding, the finding that lexico-semantic knowledge affects the nature of phonological representations in VWM and other tasks (e.g., Patterson et al., 1994; Hoffman et al., 2009; Savill et al., 2017). Relatedly, Acheson and colleagues conducted several studies suggesting that phonological and semantic information jointly affect serial order in VWM tasks in a way that would be expected from how information interacts in comprehension and production (Acheson et al., 2010, 2011b; see also Poirier et al., 2014). Similarly, Macken et al. (2014) investigated the memory implications for prosody, the intonation patterns that span whole phrases and sentences in everyday language use, in VWM tasks. Like syntactic and discourse relations, prosody is another multi-word phenomenon that does not fit neatly into the item/order distinctions in memory tasks. Macken et al. (2014) found that prosodic phrasing does affect recall, which argues against individual word units in memory.

Far less research concerns the nature of sentence-level language planning and serial ordering in VWM tasks. We mention three findings from language production research that seem particularly relevant to claims about the role of language production in VWM. All three point to the essential non-independence of words and word orders in utterance planning. First, a central tenet across essentially all approaches to language production is that lexico-semantic characteristics of individual words strongly affect their order in a sentence (Bock, 1987; Levelt, 1993). An example is that animate entities like *woman* tend to appear earlier in utterances than inanimate words like *book*. This effect is thought to reflect a more

general phenomenon linked to LTM retrieval, in which early-retrieved words enter the utterance plan first and end up in earlier positions in the utterance (Bock, 1987). Semantic features such as animacy affect retrieval and, consequently, serial position in the sentence (Bock, 1987; MacDonald, 2013). Second, the word orders that people produce tend to be ones that have been recently produced (Weiner and Labov, 1983; Bock, 1986b), but the strength of this effect is modulated by the particular words in the sentence: repeated words lead to more repeated word orders (for review, see Pickering and Ferreira, 2008). Again, words and their orders are interdependent. Third, word orders and the presence/absence of optional words in sentences vary with semantic relationships between words, where semantic similarity between two words yields more word omissions and different word orders than in the absence of semantic similarity across words (Gennari et al., 2012; Hsiao et al., 2014; Montag et al., 2017). Thus, whereas the first two examples illustrated interactions between properties of a particular word and word order of an entire utterance, this example shows that semantic relationships between two words also affect word order. All of these examples of word and word order interdependence are broadly compatible with models of language production that represent production as activation of learned weights in a connectionist architecture; these representations arguably cross item and order memory (Chang et al., 2006; Dell and Chang, 2014; McCauley and Christiansen, 2014). In this view, language production models could serve as highly informative models of serial recall, especially when the models engage in sentence repetition (see Ueno et al., 2011 for word repetition and Fischer-Baum, 2018 for other potential commonalities in serial order representations). We see this approach as inconsistent with the currently dominant views of VWM, that memory for items (the words) and memory for their serial order are unrelated, accomplished by independent mechanisms (Henson et al., 2003; Majerus, 2009; Guidali et al., 2019).

These results and approaches offer several avenues for investigations of the relationship between serial ordering of words in language production and VWM tasks. For example, it is worth further consideration of the item-order distinction in some theories of VWM, particularly those that posit a role for LTM and language production for item memory but a special-purpose system for ordering the items (Page et al., 2007; Majerus, 2009). From the point of language production, serial order is crucial both across items (i.e., word order) and within items (syllable, phoneme, articulatory gesture orders). It is curious that within-word serial order demands are considered "item memory" rather than another example of ordering memory. For current purposes, a key difference between the two types of serial order would seem to be their regularity, in that phonological order is much more rigid than syntactic order. For example, the phonemes and articulatory gestures must be in a particular order to produce a given word, and the semantic identity of the word "binds" the sub-lexical representations and their order together—the semantic binding hypothesis (Patterson et al., 1994). Dell and Chang (2014) posit a similar kind of binding from message-level semantics to the serial orders of words,



but this binding is weaker and more variable than in the word-phoneme case; there are statistical regularities between types of messages and sentence forms, but messages can usually also be conveyed with alternative word orders (MacDonald, 2016). In other words, the item-order distinction is really one of two different kinds of serial ordering demands and LTM, and the one called “item memory” (which includes the ordering of phonological codes) is much stronger and more regular than the one called “order memory.” On that view, it should be possible to manipulate these contingencies in simulations or experimentally, perhaps with artificial languages in which “word” order and “phoneme” order vary in their rigidity. If, after learning the artificial language, participants had to perform a memory task, we predict that performance at both levels should respond to the regularities of past experience and thus strength of LTM constraints, in contrast to accounts positing a rigid item/order distinction (see also Acheson and MacDonald, 2009a for discussion of “item” vs. phoneme errors and Botvinick and Bylsma, 2005 for recall in artificial languages).

Another interesting domain is performance in Hebb repetition tasks, in which participants repeatedly encounter certain serial orders across lists (Page et al., 2013; Guerrette et al., 2018). Performance in these tasks should at least initially be moderated by statistical regularities in the broader language (that is, in LTM, *via* prior experience with language), where certain words occur in certain serial orders more frequently than others. For example, we might expect that words referring to animate entities (*child*, *teacher*) would yield different serial order behavior than inanimate words (*book*, *table*) in ordered recall, because people’s broader experience ordering different types of words in their history of language production would affect how rapidly repeated patterns are learned. More generally, we expect serial ordering behavior to reflect both long-term language use and also rapid adaptation to more recent ordering contexts, a phenomenon that is robust in both language comprehension (Fine et al., 2013) and production (Bock, 1986a). Whereas Hebb repetition effects have been described in terms of repetition of specific tokens, syntactic priming effects in language processing carry across multi-word grammatical and semantic relations. If there are interactive representations between word and grammatical roles, then classic Hebb repetition effects should carry across these abstract relational categories and be moderated by fit with the category. Indeed, some studies have begun to examine these effects in sentence repetition (Allen et al., 2018; Jones and Farrell, 2018) and in recall of lists with grammatical dependencies (Perham et al., 2009) by considering how lists consistent with grammatical knowledge are recalled better than lists inconsistent with these patterns. The emergent account described here would further predict that the effect of grammatical knowledge would be moderated by semantic information of words, such as animacy, and morphophonological cues, reflecting interrelationships in LTM. For example, recall of animate nouns should be greater than recall of inanimate nouns in the context of word lists that encourage a noun to be interpreted as an agent, because

animate nouns are commonly agents of actions and inanimate nouns are not. Furthermore, this account would suggest rapid adaptation to novel orders would affect memory in a manner consistent with models of language production that learn over experience.

## Challenges for the Multicomponent Approach

Rather than viewing memory representations as graded, integrated, and distributed, as described above, multicomponent models separate various representations into discrete components. For example, the phonological loop stores phonological representations in a buffer separate from other representations (Baddeley and Hitch, 1974). Likewise, other researchers posit separate phonological and semantic buffers stemming from language mechanisms (Martin and Romani, 1994). These models are reminiscent of older, modular models of language comprehension and production that employ discrete stores and restricted interaction of information (Forster, 1985; Frazier, 1987). To fully capture the rich and interactive tapestry of language representations that are invoked in more current language research, multicomponent models would seem to require a combinatorial explosion of additional buffers for each form of interaction. In terms of parsimony and plausibility, this seems unlikely to be a tenable solution. Martin and Freedman (2001) offered a possible solution in which various language representations may interact in a multi-component memory model by passing the activity through layers with phonological and semantic buffers. This approach may allow more interaction but is also inconsistent with much language research, as it specifically implies that certain language representations are processed independently and in sequence (MacDonald and Seidenberg, 2006). As far as we are aware, no research has explicitly considered how different forms of interactive representations could be modeled in VWM in a manner consistent with language comprehension and production research. Even so, it is unclear how integrated representations and interactive processing could be implemented in a multicomponent account.

An important route for LTM effects on VWM performance in multicomponent models is redintegration, a process that rebuilds decaying memory traces from LTM (Roodenrys and Hinton, 2002; Roodenrys et al., 2002; Allen and Hulme, 2006; Clarkson et al., 2017). The redintegration mechanism not only rebuilds the phonological loop with phonological information from LTM (Clarkson et al., 2017), it also is the mechanism invoked to account for other LTM effects that go beyond phonological structure, including influences of word frequency and long-term knowledge of semantics and word co-occurrences on VWM (Hulme et al., 1997; Walker and Hulme, 1999; Roodenrys et al., 2002; Stuart and Hulme, 2009). In this view, the redintegration process must use LTM outside the phonological domain to shore up decaying phonological buffers. It is not clear how that process would work if LTM representations are highly integrated. Such a process would imply that phonological representations

are first stripped from their richly integrated encoding in LTM, maintained in a separate phonological buffer, and then recombined with their integrated representations at the time of recall.

Currently, empirical evidence in favor of emergent (Postle, 2006; Acheson et al., 2011a; Buchsbaum and D'Esposito, 2019) and multicomponent accounts (for review, see Shallice and Papagno, 2019; Yue et al., 2019) has established little consensus. We recognize that many of the claims above are logical arguments, and further empirical evidence could prove some of our assumptions faulty. Proponents of emergent models should see language comprehension and production mechanisms as consistent with VWM systems that stem from a richly structured and integrated LTM (Acheson and MacDonald, 2009b; Jones and Macken, 2015; Hughes et al., 2016). Proponents of multicomponent models, however, may see these discussions of a rich language LTM and the processes that operate with it as simply more evidence for the sorts of information that could be encoded *via* language processes or that reintegration could use to reconstruct memory traces. Regardless, defining LTM representations is important for the advancement of memory models, and language models should provide insight into these LTM representations.

## Challenges for Limited Emergence

Perhaps one of the most persistent complaints against emergent accounts is their inability to handle aphasic patient data (Shallice and Papagno, 2019). Classically, patterns of behavior by patients with aphasia have been seen as evidence for the notion that STM and LTM are supported by distinct neural populations. Lesions to the medial temporal lobe have appeared to yield deficits of LTM with spared STM, typically assessed using lexical decision tasks and digit span tasks, respectively (Scoville and Milner, 1957; Penfield and Milner, 1958; Baddeley and Warrington, 1970; Warrington et al., 1971; Cave and Squire, 1992). In contrast, damage to left parietal regions have been interpreted to cause impairments in verbal recognition tasks and digit span tasks with spans greater than 1 or 2 while sparing other cognitive functions and LTM (e.g., Warrington and Shallice, 1969, 1972; Shallice and Warrington, 1970, 1974; Vallar and Baddeley, 1984). Thus, these studies of patients appeared to show a double dissociation of STM and LTM.

Some patient data may also support a dissociation between language processing and STM. For example, the patient K.F. reported in Warrington and Shallice (1969) exhibited strong repetition deficits with spared word knowledge, which would typically classify the patient as having conduction aphasia. However, given that the patient exhibited recognition deficits even when no verbal output was required by the task (i.e., pointing), Warrington and Shallice concluded that the patient's impairment was not limited to language repetition. Later work reinforced this notion in patients with impaired phonological discrimination with spared word recognition and short sentence comprehension (Basso et al., 1982; Vallar and Baddeley, 1984; Silveri and Cappa, 2003) as well as in patients with dissociable speech and STM deficits (Martin and Breedin, 1992). In a similar way, more recent research has attempted to

unconfound item and order memory (Attout et al., 2012; Majerus et al., 2015).

However, the putative pure deficits of STM are frequently tainted by subtle language impairments (Martin and Saffran, 1992). For example, Warrington et al. (1971) described a selective impairment of STM in a group of patients, yet those same patients exhibited difficulty in the repetition of abstract words, reading, and fluent speech. Vallar and Baddeley (1984) claimed to have found a pure deficit of STM in one patient, yet that same patient exhibited impaired comprehension of longer sentences compared to other participants. Even the patients identified with fluent speech also exhibited abnormalities. For example, the patient described by Shallice and Butterworth (1977) exhibited paraphasic errors in speaking names and had difficulty comprehending spoken discourse and written text. Furthermore, comprehension difficulty was exacerbated for complex sentences. Jacquemot et al. (2006) claimed to have found patients with a specific STM impairment, yet those same patients also exhibited difficulty in language comprehension tasks and sentence repetition tasks, resulting in phonological paraphasias. A truly pure deficit has proven quite elusive (though see Martin and Breedin, 1992). Rather than see these language deficits as stemming from a specific STM impairment, we see both as being driven by deficits in LTM. A complementary pattern is seen in other lines of research. For example, Hannula et al. (2006) found that hippocampal deficits cause impairments in relational processing at both short and long durations, upsetting prominent research suggesting that hippocampal activity is associated only with LTM. A strongly emergent perspective accords neatly with this data.

A recurrent theme in this review has been that the relationship between VWM and LTM depends on the nature of language LTM. Patient data is no exception. Reference to models of language production and comprehension reveals how apparent STM deficits could be captured by damage to LTM. Martin and Saffran (1992) presented the case of a patient with deep dysphasia who exhibited apparent errors of STM: difficulty producing nonwords and semantic errors in repetition. This patient exhibited fluent speech with semantic and phonological paraphasias. The researchers evaluated this patient's performance through the lens of the Dell (1986) interactive spreading activation model of lexical retrieval. This model employs discrete representations of phonology, lexical entries, and semantics that interact in a bidirectional network. The model was able to produce human-like lexical selection behaviors. Critically, the model was able to capture putatively pure STM patient data solely through perturbation of the model parameters and without the inclusion of a distinct memory buffer. In this specific case, an increased decay rate reduced the ability of lexical representations to support lexical selection. The predictions afforded by this model were later confirmed in additional analyses of patient data by Martin et al. (1994; see also Dell et al., 1997), and patient recovery was also able to be modeled using the same framework (Martin et al., 1996). These results suggest that a specification of the LTM representations relevant to

language comprehension and production may help test claims about the representational basis of VWM and its relationship to LTM.

Findings such as these point to the need for contact between theories of VWM and perspectives on long-term representations of serial order in language. That is, the extent to which the above or similar results affect VWM models depends on the hypothesized nature of LTM, particularly the extent to which LTM could contribute to representations of novel memoranda and their order. Language LTM captures relations between words and levels of linguistic representation and therefore allows generalization to new cases. Indeed, any linguistic input is novel in many ways, such as a new word-order, new speaker, new acoustic environment, and so on. By definition, the goal of language comprehension processes is to cope with novel input, and language production processes constantly generate novel utterances. The VWM literature offers a different perspective, with some claiming that buffers are needed explicitly to represent novel material (Norris, 2017). One challenge for memory research is the need to characterize a clear divide between “old” and “new,” especially given that novelty means very different things in different memory models. Distributed language models provide a key demonstration of the emergent perspective. In such models, novel stimuli are processed with respect to their similarity to prior experience, without any need for separate systems dedicated to handling the particularities of novel items or orders. In parallel, emergent models of VWM are capable of producing novel sequences just using LTM, without dedicated short-term buffers (e.g., Botvinick and Bylsma, 2005; Botvinick and Plaut, 2006, 2009). Perhaps greater adoption of graded representations of novelty could bridge the divide between language emergent and pure memory accounts. Important behavioral data linking graded phonotactic LTM to VWM (e.g., Tanida et al., 2015) and graded grammatical LTM to VWM (e.g., Jones and Farrell, 2018) already speaks to the usefulness of this approach.

## IMPLICATIONS FOR LANGUAGE AND VWM RESEARCH

We have cited a broad range of work in both VWM and in language comprehension and production, and one of the striking features of that work is how very little the fields have to say about each other. For example, it is completely uncontroversial that language comprehension and production processes are constrained by what is commonly called “verbal working memory capacity” in those fields, and yet the specific mechanisms posited in classic VWM models are, with only a few exceptions, absent from theorizing about how limited capacities shape language processes (for review and a different perspective, see Caplan and Waters, 2013). Similarly, while VWM accounts assume that VWM abilities must be used in everyday activities, the connection to actual theories of language use is equally scant. Here we discuss several fronts with more potential for interaction among the fields.

## Implications for Relating WM Assessments to Other Measures

The approach that we have advocated, in which performance on VWM tasks is heavily supported by language processes, which are themselves dependent on long-term knowledge, naturally leads to questions about what VWM tasks actually measure. This question is not only central to theories of working memory but also has enormous practical significance because there is wide usage of tasks that are described as VWM assessments in clinical and educational contexts—in typical and atypical child development, young adults, older adults, and patients with brain injury. Whereas some researchers have considered poor VWM performance as a cause of poor language skills, potentially ameliorated by working memory training (e.g., Ingvalson et al., 2015), our language-emergent VWM view suggests that poor VWM performance is a symptom associated with poor language skill. In other words, the abilities to encode, maintain, and order verbal information are skills that emerge from language use, and individuals who have higher language skills have richer LTM representations and more practiced comprehension and production processes (see also Jones et al., 2020). Thus, we can view tasks that are described as VWM tasks not as assessments of a separate VWM capacity but rather as measures of a person’s skill in encoding and maintaining verbal information. Consistent with this approach, there are now a number of reassessments of tasks that have previously been called “working memory tasks,” with arguments that they are better viewed as assessments of language skill, including but not limited to encoding, maintenance, and ordering. Tasks that have been reinterpreted in this way include reading span (MacDonald and Christiansen, 2002), digit span (Jones and Macken, 2015), nonword repetition (Edwards et al., 2004; Estes et al., 2007), sentence repetition (Klem et al., 2015), and immediate serial recall of word lists (Perham et al., 2009). In each of these examples, the argument has the same character. The apparent “verbal working memory task” does not measure a separate memory capacity but instead measures the quantity and quality of language skill and experience relevant to the specific demands of the task (see also Jones and Macken, 2018). Thus, nonword repetition performance can be traced to the knowledge of phonological patterns and vocabulary (Edwards et al., 2004; Gupta and Tisdale, 2009), digit span performance can be linked to prior experience with and statistical learning of digit sequences (Jones and Macken, 2015), and so on. The overarching conclusion from this work is that computational capacity to perform some task is not independent of long-term language knowledge and experience (MacDonald and Christiansen, 2002). That is an essential claim of an emergent perspective.

The emergent perspective also helps to elucidate so-called “brain training” research. If VWM is emergent from language LTM, then training VWM should only be beneficial (e.g., Soveri et al., 2017) if training improves relevant language skills. In contrast, VWM training should not be effective if it merely attempts to manipulate some independent notion of capacity. VWM training has been applied to therapeutic contexts, such as with aphasic patients, but the effectiveness of such interventions



is unclear, driven in part by methodological limitations of single case studies (Zakariás et al., 2019). VWM treatments almost always employ linguistic stimuli of some sort, meaning they inherently provide some language practice. Therefore, VWM is rarely divorced from linguistic LTM in the training. VWM training research could benefit from a consideration of the emergent perspective defined hereby further developing language skills as opposed to separate memory capacity.

## Implications for Attention, Task Subcomponents, and Domain Generality

All theories of VWM have some mix of domain-specific and domain-general components. For example, the multicomponent model has the domain-specific phonological loop but also the general Central Executive, which guides behavior beyond the maintenance of phonological forms. Similarly, emergent views have domain-general attention and other cognitive control processes, but LTM can be domain-specific, in that linguistic knowledge need not have the same properties as a memory for smell or spatial relations. The specific emergent approach advocated here, in which language LTM and language comprehension and production processes underlie VWM functions, might initially seem strongly domain-specific in character, given the modular perspective that has pervaded language research. However, “emergent from language processes” need not be “domain-specific.” Indeed, there has been new interest in investigating how language use is supported by domain-general processes of attention and episodic memory (Nozari et al., 2016; Van de Cavey and Hartsuiker, 2016; Hepner and Nozari, 2019), and interest in how distinct brain networks must coordinate to accomplish language comprehension and other complex cognitive processes (Fedorenko et al., 2011; Fedorenko, 2014). Close ties with attention have long been a component of emergent models (e.g., Cowan, 1993), and researchers are now considering the interrelationships between language and attention mechanisms with respect to VWM (Majerus, 2019). More generally, there is real interest in considering the extent to which language production processes are related to or are themselves emergent from more general action planning processes or domain-general sequencing systems (Van de Cavey and Hartsuiker, 2016; Anderson and Dell, 2018; Guidali et al., 2019). Long-term ordering knowledge across domains (e.g., Kaiser, 2012; Van de Cavey and Hartsuiker, 2016) may inform sequence ordering, further tying together domain-general perspectives, emergent models, and language research. If language research continues to embrace more domain-general processes, this development could have substantial consequences for debates about the relationship between language processes and VWM, including distinctions between multicomponent and emergent accounts. That is, if VWM and language researchers both incorporate the same domain-general processes, then the distinction between multicomponent models and emergent models becomes less theoretically important.

Perhaps one of the most compelling examples of how domain-general processes affect language use and temporary maintenance may be seen in conversational turn-taking, which draws on episodic memory (Duff and Brown-Schmidt, 2012;

Rubin et al., 2014) and cognitive control. Using data from recordings of conversations in 10 languages, Stivers et al. (2009) found that speakers typically begin speaking less than 500 ms after the previous speaker has ended their conversational turn. A number of researchers have argued that this closely time-locked behavior requires extensive attention, maintenance, and cognitive control because the next speaker simultaneously juggles a number of disparate tasks, some of which bear a close similarity to demands of VWM tasks. The conversational demands on the person who will soon speak include: comprehending the person currently speaking; planning a response and maintaining that utterance plan until time to speak; predicting the timing of the current speaker’s endpoint, which often involves predicting the actual words that the current speaker is likely to end on; and triggering an anticipatory in-breath and then exhalation to allow the speech to begin (de Ruiter et al., 2006; Torreira et al., 2015; Levinson, 2016). Not surprisingly, turn-taking and planning before speaking have high processing loads, as measured in a variety of methods (Kemper et al., 2011; Boiteau et al., 2014; Barthel and Sauppe, 2019). Thus, while a participant’s overall goals in a conversation and a VWM task are very different, it should be clear that the task demands of both activities overlap, including simultaneously encoding input while developing and maintaining plans to generate a response. Researchers are actively investigating the attention and cognitive control demands of language planning in advance of speaking, including serial ordering and monitoring of utterance plans (for review, see Nozari and Novick, 2017 and Fischer-Baum, 2018 for potential implications for VWM tasks). Some methods manipulating selective attention to individual words in a list could prove to be useful for new studies of both VWM tasks and more typical language production (e.g., Nozari and Dell, 2012; Nozari and Thompson-Schill, 2013). We see this research as complicating the domain-specific/general debates but also as an important arena for collaboration between VWM and language researchers.

## Implications for Language Production Research

The view that language production underlies maintenance of verbal information has significant implications for language production research. If every VWM study can be seen as a particular form of language production, the radically emergent perspective we describe has the potential to inform theories of language production. Interaction between the fields has long been evident at phonological levels. There has been keen interest in phonological level speech errors as important data for theories of serial ordering in language production (Dell, 1984; Dell et al., 1997), and there are extensive discussions of relationships between speech errors and recall errors in VWM tasks (Ellis, 1980; Hartley and Houghton, 1996; Page et al., 2007; Acheson and MacDonald, 2009a). In addition, VWM research has increasingly investigated the Hebb Repetition effect, the improved recall of repeated lists (Hebb, 1961; Oberauer et al., 2015). In parallel, production researchers have investigated the effects of learning on serial ordering and speech errors in production (Dell et al., 2000; Anderson



et al., 2019). These investigations may be mutually informative, especially when placed in the context of computational models of ordering in VWM and models of language production which produce ordered sequences. As we have noted, some of these models have already suggested some parallels in ordering mechanisms between the two domains (Page and Norris, 2009; Hartley et al., 2016).

There are also potential parallels beyond the phonological level, relevant to questions concerning the relationship between words and their production in ordered sequences. MacDonald (2016) argued that of the three most obvious task demand differences between immediate serial recall and everyday language production (item list vs. coherent message, recall signal vs. spontaneous production, and producing exact list order vs. flexible language production), the latter was particularly important for understanding relationships between language production and VWM. Whereas serial recall, by definition, must be in the presented order, a hallmark of language production at the phrase or sentence level is serial order flexibility—that almost any message can be conveyed *via* several different words and word orders. This difference is informative when considering how interference among similar words can affect performance in language production and VWM tasks. Interference among list items leads to item omissions and re-ordering of list items in the recall; these are naturally treated as ordering errors, given the task demands in immediate serial recall (Baddeley, 1966; Page et al., 2007; though see Saint-Aubin and Poirier, 1999). Language production is also subject to interference among words, which leads to omissions and alternative word orders, compared to production conditions without interference (Gennari et al., 2012; Hsiao et al., 2014). These shifts and omissions are not considered errors but in some sense evidence of production skill, that is, evidence for how the speaker uses alternative ordering to maintain fluency in the face of interference. What is missing in this literature is a better understanding of interference during production planning and maintenance, and how alternative word orders emerge in the face of this interference. These questions seem ripe for insight from and collaboration with VWM research.

## Implications for Language Comprehension

Theories of language comprehension aim to explain how language percepts are recognized and interpreted. Important data in this endeavor have been measures of comprehension difficulty, or, more specifically, the relative difficulty of some kind of language compared to another. In the case of sentence-level comprehension research, the focus has been on why some kinds of sentences are harder than others, and VWM capacity has been a common explanatory factor in this field (MacDonald and Hsiao, 2018). Many researchers have invoked decay in VWM to explain comprehension difficulty of certain kinds of sentences, as the difficult sentences require integration over distant information that has degraded in working memory (Just and Carpenter, 1992; Gibson, 1998; Babyonyshev and Gibson, 1999; Grodner and Gibson, 2005). An alternative approach suggests that VWM and comprehension difficulty are constrained by interference rather than decay or capacity

limitations (Lewis et al., 2006; Van Dyke and Johns, 2012; Glaser et al., 2013). This work emphasizes that both encoding and retrieval of information becomes more difficult with the increased semantic similarity between words, meaning sentences with more interfering elements are more difficult to comprehend (for review, see Van Dyke and Johns, 2012). This area is, therefore, another in which VWM research could inform comprehension, particularly the influence of decay and/or interference (Oberauer et al., 2016). More generally, though, while language comprehension researchers have often invoked VWM limitations in accounts of comprehension difficulty, they have not necessarily aligned themselves with particular VWM models of encoding, maintenance, and retrieval processes (for some exceptions, see Just and Carpenter, 1992; Martin and Romani, 1994; Lewis et al., 2006; Caplan and Waters, 2013).

At least initially, very few accounts of language comprehension ascribed a major role for experience in language comprehension difficulty. These accounts were, at least in principle, aligned with a multi-component perspective. A separate, temporary store, separate from long-term language knowledge, provided a bottleneck in encoding and maintenance that could explain comprehension difficulty. More recently, a number of researchers have suggested that both VWM capacity and language experience are important components in processing difficulty (Demberg and Keller, 2008; Staub, 2010). In a more fully emergent approach of VWM, the capacity to encode and maintain information (whether for everyday language use or a working memory task) is not independent of long-term memory, and thus not independent of experience with language (McClelland and Elman, 1986; MacDonald and Christiansen, 2002; Botvinick and Plaut, 2006; Acheson and MacDonald, 2009a; Jones and Macken, 2015). We see this emphasis on experience-based capacity as a basis for investigating parallels between comprehension processes and VWM. Moreover, the emphasis on experience also casts language use and memory as intertwined, learned skills, as noted in the discussion of revised interpretations of VWM tasks above. For example, memory researchers have noted relationships between novel word learning and the Hebb repetition effect (Szmalec et al., 2009). If word representations are highly intertwined, as our emergent perspective claims, then sensitivity to the Hebb repetition effect and novel word learning should exhibit exploitation of statistical regularities between different sources of information (e.g., Cassidy and Kelly, 1991; Nygaard et al., 2009) rather than mere memory capacity of the learner.

## CONCLUSIONS

In this article, we have aimed to describe the rich nature of linguistic LTM and its consequences for VWM. While Ebbinghaus (1885) had inklings that LTM could not be fully set aside in studying VWM, we have suggested that the linkage between language LTM and VWM is far stronger than he imagined, in part because LTM has a different quality than he and many others had hypothesized. A more thorough understanding of the nature of language processing, attention, and LTM, we claim, will accelerate the advancement of both VWM and

language research. We have argued that words are not unrelated islands in LTM representations, and therefore they should not be treated as isolated items in VWM research. We have further argued that the processes of language comprehension and production underlie a person's ability to encode, maintain, and order verbal information. These skills are essential for everyday language use, change with experience and the richness of LTM, and are brought to bear on VWM tasks. On this view, VWM and language research should be mutually informative.

## REFERENCES

- Aben, B., Stapert, S., and Blokland, A. (2012). About the distinction between working memory and short-term memory. *Front. Psychol.* 3:301. doi: 10.3389/fpsyg.2012.00301
- Acheson, D. J., Hamidi, M., Binder, J. R., and Postle, B. R. (2011a). A common neural substrate for language production and verbal working memory. *J. Cogn. Neurosci.* 23, 1358–1367. doi: 10.1162/jocn.2010.21519
- Acheson, D. J., MacDonald, M. C., and Postle, B. R. (2011b). The effect of concurrent semantic categorization on delayed serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 44–59. doi: 10.1037/a0021205
- Acheson, D. J., and MacDonald, M. C. (2009a). Twisting tongues and memories: explorations of the relationship between language production and verbal working memory. *J. Mem. Lang.* 60, 329–350. doi: 10.1016/j.jml.2008.12.002
- Acheson, D. J., and MacDonald, M. C. (2009b). Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychol. Bull.* 135, 50–68. doi: 10.1037/a0014411
- Acheson, D. J., Postle, B. R., and MacDonald, M. C. (2010). The interaction of concreteness and phonological similarity in verbal working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 17–36. doi: 10.1037/a0017679
- Allen, R. J., Hitch, G. J., and Baddeley, A. D. (2018). Exploring the sentence advantage in working memory: insights from serial recall and recognition. *Q. J. Exp. Psychol.* 71, 2571–2585. doi: 10.1177/1747021817746929
- Allen, R., and Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *J. Mem. Lang.* 55, 64–88. doi: 10.1016/j.jml.2006.02.002
- Allen, J., and Seidenberg, M. S. (1999). “The emergence of grammaticality in connectionist networks,” in *The Emergence of Language*, ed. B. MacWhinney (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 115–151.
- Allport, D. A., and Funnell, E. (1981). Components of the mental lexicon. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 295, 397–410. doi: 10.1098/rstb.1981.0148
- Almeida, R. G. D., and Gleitman, L. R. (2018). *On Concepts, Modules, and Language: Cognitive Science at Its Core*. New York, NY: Oxford University Press.
- Altmann, G. T. M., and Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cogn. Sci.* 33, 583–609. doi: 10.1111/j.1551-6709.2009.01022.x
- Anderson, N. D., and Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proc. Natl. Acad. Sci. U S A* 115, 3617–3622. doi: 10.1073/pnas.1721107115
- Anderson, N. D., Holmes, E. W., Dell, G. S., and Middleton, E. L. (2019). Reversal shift in phonotactic learning during language production: evidence for incremental learning. *J. Mem. Lang.* 106, 135–149. doi: 10.1016/j.jml.2019.03.002
- Archibald, L. M. D., and Gathercole, S. E. (2006). Nonword repetition: a comparison of tests. *J. Speech Lang. Hear. Res.* 49, 970–983. doi: 10.1044/1092-4388(2006/070)
- Arnott, I., and Snider, N. (2010). More than words: frequency effects for multi-word phrases. *J. Mem. Lang.* 62, 67–82. doi: 10.1016/j.jml.2009.09.005
- Attout, L., Ordonez Magro, L., Szmalec, A., and Majerus, S. (2019). The developmental neural substrates of item and serial order components of verbal working memory. *Hum. Brain Mapp.* 40, 1541–1553. doi: 10.1002/hbm.24466
- Attout, L., Van der Kaa, M.-A., George, M., and Majerus, S. (2012). Dissociating short-term memory and language impairment: the importance of item and

## AUTHOR CONTRIBUTIONS

SS and MM contributed to all phases of writing this review.

## FUNDING

Preparation of this work was supported by NSF Grant number 1849236.

- serial order information. *Aphasiology* 26, 355–382. doi: 10.1080/02687038.2011.604303
- Baayen, R. H., Shaoul, C., Willits, J., and Ramscar, M. (2016). Comprehension without segmentation: a proof of concept with naive discriminative learning. *Lang. Cogn. Neurosci.* 31, 106–128. doi: 10.1080/23273798.2015.1065336
- Babyonyshev, M., and Gibson, E. (1999). The complexity of nested structures in Japanese. *Language* 75, 423–450. doi: 10.2307/417056
- Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/s1364-6613(00)01538-2
- Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Q. J. Exp. Psychol.* 18, 302–309. doi: 10.1080/14640746608400047
- Baddeley, A. D., and Hitch, G. (1974). “Working memory,” in *Psychology of Learning and Motivation*, (Vol. 8) ed. G. H. Bower (New York, NY: Academic Press), 47–89.
- Baddeley, A. D., Hitch, G. J., and Allen, R. J. (2009). Working memory and binding in sentence recall. *J. Mem. Lang.* 61, 438–456. doi: 10.1016/j.jml.2009.05.004
- Baddeley, A., Lewis, V., and Vallar, G. (1984). Exploring the articulatory Loop. *Q. J. Exp. Psychol. A* 36, 233–252. doi: 10.1080/14640748408402157
- Baddeley, A. D., Thomson, N., and Buchanan, M. (1975). Word length and the structure of short-term memory. *J. Verbal Learn. Verbal Behav.* 14, 575–589. doi: 10.1016/s0022-5371(75)80045-4
- Baddeley, A. D., and Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *J. Verbal Learn. Verbal Behav.* 9, 176–189. doi: 10.1016/s0022-5371(70)80048-2
- Barthel, M., and Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science* 43:e12768. doi: 10.1111/cogs.12768
- Basso, A., Spinnler, H., Vallar, G., and Zanolio, M. E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory. A case study. *Neuropsychologia* 20, 263–274. doi: 10.1016/0028-3932(82)90101-4
- Bock, J. K. (1986a). Meaning, sound, and syntax: lexical priming in sentence production. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 575–586. doi: 10.1037/0278-7393.12.4.575
- Bock, J. K. (1986b). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi: 10.1016/0010-0285(86)90004-6
- Bock, K. (1987). An effect of the accessibility of word forms on sentence structures. *J. Mem. Lang.* 26, 119–137. doi: 10.1016/0749-596x(87)90120-3
- Boiteau, T. W., Malone, P. S., Peters, S. A., and Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *J. Exp. Psychol. Gen.* 143, 295–311. doi: 10.1037/a0031858
- Botvinick, M., and Bylsma, L. M. (2005). Regularization in short-term memory for serial order. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 351–358. doi: 10.1037/0278-7393.31.2.351
- Botvinick, M. M., and Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychol. Rev.* 113, 201–233. doi: 10.1037/0033-295X.113.2.201
- Botvinick, M. M., and Plaut, D. C. (2009). Empirical and computational support for context-dependent representations of serial order: reply to Bowers, Damian, and Davis (2009). *Psychol. Rev.* 116, 998–1001. doi: 10.1037/a0017113
- Brener, R. (1940). An experimental investigation of memory span. *J. Exp. Psychol.* 26, 467–482. doi: 10.1037/h0061096

- Buchsbaum, B. R., and D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex* 112, 134–148. doi: 10.1016/j.cortex.2018.11.010
- Bybee, J., and McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguist. Rev.* 22, 381–410. doi: 10.1515/tlir.2005.22.2-4.381
- Caplan, D., and Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychon. Bull. Rev.* 20, 243–268. doi: 10.3758/s13423-012-0369-9
- Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., et al. (2007). Non-word repetition in children with specific language impairment: a deficit in phonological working memory or in long-term verbal knowledge? *Cortex* 43, 769–776. doi: 10.1016/s0010-9452(08)70505-7
- Cassidy, K. W., and Kelly, M. H. (1991). Phonological information for grammatical category assignments. *J. Mem. Lang.* 30, 348–369. doi: 10.1016/0749-596x(91)90041-h
- Cave, C. B., and Squire, L. R. (1992). Intact verbal and nonverbal short-term memory following damage to the human hippocampus. *Hippocampus* 2, 151–163. doi: 10.1002/hipo.450020207
- Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Rev.* 113, 234–272. doi: 10.1037/0033-295x.113.2.234
- Chang, F., Dell, G. S., Bock, K., and Griffin, Z. M. (2000). Structural priming as implicit learning: a comparison of models of sentence production. *J. Psycholinguist. Res.* 29, 217–230. doi: 10.1023/a:1005101313330
- Christiansen, M. H., Allen, J., and Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: a connectionist model. *Lang. Cogn. Process.* 13, 221–268. doi: 10.1080/016909698386528
- Christiansen, M. H., and Monaghan, P. (2016). Division of labor in vocabulary structure: insights from corpus analyses. *Top. Cogn. Sci.* 8, 610–624. doi: 10.1111/tops.12164
- Clarkson, L., Roodenrys, S., Miller, L. M., and Hulme, C. (2017). The phonological neighbourhood effect on short-term memory for order. *Memory* 25, 391–402. doi: 10.1080/09658211.2016.1179330
- Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Mem. Cognit.* 21, 539–545. doi: 10.3758/bf03197185
- Connine, C. M., and Clifton, C. Jr. (1987). Interactive use of lexical information in speech perception. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 291–299. doi: 10.1037/0096-1523.13.2.291
- Cowan, N. (1993). Activation, attention, and short-term memory. *Mem. Cognit.* 21, 162–167. doi: 10.3758/bf03202728
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Essence of Memory*, eds W. S. Sossin, J.-C. Lacaille, V. F. Castellucci and S. Belleville, (Amsterdam, The Netherlands: Elsevier), 323–338.
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychon. Bull. Rev.* 24, 1158–1170. doi: 10.3758/s13423-016-1191-6
- 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
- Crowder, R. G. (1993). Short-term memory: where do we stand? *Mem. Cognit.* 21, 142–145. doi: 10.3758/bf03202725
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Dapretto, M., and Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427–432. doi: 10.1016/s0896-6273(00)80855-7
- de Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130
- Dell, G. S. (1984). Representation of serial order in speech: evidence from the repeated phoneme effect in speech errors. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 222–233. doi: 10.1037/0278-7393.10.2.222
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321. doi: 10.1037/0033-295x.93.3.283
- Dell, G. S., Burger, L. K., and Svec, W. R. (1997). Language production and serial order: a functional analysis and a model. *Psychol. Rev.* 104, 123–147. doi: 10.1037/0033-295x.104.1.123
- Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20120394. doi: 10.1098/rstb.2012.0394
- Dell, G. S., Reed, K. D., Adams, D. R., and Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1355–1367. doi: 10.1037/0278-7393.26.6.1355
- Dell, G. S., and Reich, P. A. (1981). Stages in sentence production: an analysis of speech error data. *J. Verbal Learn. Verbal Behav.* 20, 611–629. doi: 10.1016/s0022-5371(81)90202-4
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., and Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* 104, 801–838. doi: 10.1037/0033-295x.104.4.801
- Demberg, V., and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109, 193–210. doi: 10.1016/j.cognition.2008.07.008
- Dikler, S., Rabagliati, H., Farmer, T. A., and Pykkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychol. Sci.* 21, 629–634. doi: 10.1177/0956797610367751
- Duff, M. C., and Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Front. Hum. Neurosci.* 6:69. doi: 10.3389/fnhum.2012.00069
- Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. New York, NY: Dover.
- Edwards, J., Beckman, M. E., and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *J. Speech Lang. Hear. Res.* 47, 421–436. doi: 10.1044/1092-4388(2004)034
- Ellis, A. W. (1980). Errors in speech and short-term memory: the effects of phonemic similarity and syllable position. *J. Verbal Learn. Verbal Behav.* 19, 624–634. doi: 10.1016/s0022-5371(80)90672-6
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Mach. Learn.* 7, 195–225. doi: 10.1007/bf00114844
- Epstein, W. (1961). The influence of syntactical structure on learning. *Am. J. Psychol.* 74, 80–85. doi: 10.2307/1419827
- Epstein, W. (1962). A further study of the influence of syntactical structure on learning. *Am. J. Psychol.* 75, 121–126. doi: 10.2307/1419550
- Estes, K. G., Evans, J. L., and Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 50, 177–195. doi: 10.1044/1092-4388(2007)015
- Farmer, T. A., Christiansen, M. H., and Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proc. Natl. Acad. Sci. U S A* 103, 12203–12208. doi: 10.1073/pnas.0602173103
- Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x
- Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Lang. Sci.* 5:335. doi: 10.3389/fpsyg.2014.00335
- Fedorenko, E., Behr, M. K., and Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U S A* 108, 16428–16433. doi: 10.1073/pnas.1112937108
- Fine, A. B., Jaeger, T. F., Farmer, T. A., and Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One* 8:e77661. doi: 10.1371/journal.pone.0077661
- Fischer-Baum, S. (2018). “A common representation of serial position in language and memory,” in *Psychology of Learning and Motivation*, eds K. D. Federmeier and D. G. Watson, (London, UK), 31–54.
- Fischer-Baum, S., and McCloskey, M. (2015). Representation of item position in immediate serial recall: evidence from intrusion errors. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1426–1446. doi: 10.1037/xlm0000102
- Forster, K. I. (1985). Lexical acquisition and the modular lexicon. *Lang. Cogn. Process.* 1, 87–108. doi: 10.1080/01690968508402073
- Frank, M. C., and Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cogn. Psychol.* 75, 80–96. doi: 10.1016/j.cogpsych.2014.08.002
- Frazier, L. (1987). “Sentence processing: a tutorial review,” in *Attention and Performance XII*, ed. M. Coltheart (Hillsdale, NJ: Erlbaum), 559–586.



- Gathercole, S. E., Frankish, C. R., Pickering, S. J., and Peaker, S. (1999). Phonotactic influences on short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 84–95. doi: 10.1037/0278-7393.25.1.84
- Gennari, S. P., Mirković, J., and MacDonald, M. C. (2012). Animacy and competition in relative clause production: a cross-linguistic investigation. *Cogn. Psychol.* 65, 141–176. doi: 10.1016/j.cogpsych.2012.03.002
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/s0010-0277(98)00034-1
- Glaser, Y. G., Martin, R. C., Van Dyke, J. A., Hamilton, A. C., and Tan, Y. (2013). Neural basis of semantic and syntactic interference in sentence comprehension. *Brain Lang.* 126, 314–326. doi: 10.1016/j.bandl.2013.06.006
- Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., and Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *NeuroImage* 53, 638–646. doi: 10.1016/j.neuroimage.2010.06.055
- Grodner, D., and Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cogn. Sci.* 29, 261–290. doi: 10.1207/s15516709cog0000\_7
- Guerrette, M.-C., Saint-Aubin, J., Richard, M., and Guérard, K. (2018). Overt language production plays a key role in the Hebb repetition effect. *Mem. Cognit.* 46, 1389–1397. doi: 10.3758/s13421-018-0844-2
- Guidali, G., Pisoni, A., Bolognini, N., and Papagno, C. (2019). Keeping order in the brain: the supramarginal gyrus and serial order in short-term memory. *Cortex* 119, 89–99. doi: 10.1016/j.cortex.2019.04.009
- Gupta, P., and Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *J. Mem. Lang.* 61, 481–502. doi: 10.1016/j.jml.2009.08.001
- Hannula, D. E., Tranel, D., and Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *J. Neurosci.* 26, 8352–8359. doi: 10.1523/JNEUROSCI.5222-05.2006
- Hartley, T., and Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *J. Mem. Lang.* 35, 1–31. doi: 10.1006/jmla.1996.0001
- Hartley, T., Hurlstone, M. J., and Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: a model and tests of its stimulus-driven mechanism. *Cogn. Psychol.* 87, 135–178. doi: 10.1016/j.cogpsych.2016.05.001
- Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci.* 19, 304–313. doi: 10.1016/j.tics.2015.04.006
- Hebb, D. O. (1961). “Distinctive features of learning in the higher animal,” in *Brain Mechanisms and Learning*, ed. J. F. Delafresnaye (Oxford: Blackwell), 37–46.
- Henson, R., Hartley, T., Burgess, N., Hitch, G., and Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: a new paradigm and tests of a timing-signal hypothesis. *Q. J. Exp. Psychol.* A 56, 1307–1334. doi: 10.1080/02724980244000747
- Hepner, C. R., and Nozari, N. (2019). Resource allocation in phonological working memory: same or different principles from vision? *J. Mem. Lang.* 106, 172–188. doi: 10.1016/j.jml.2019.03.003
- Hinton, G. E., and Shallice, T. (1991). Lesioning an attractor network: investigations of acquired dyslexia. *Psychol. Rev.* 98, 74–95. doi: 10.1037/0033-295x.98.1.74
- Hoffman, P., Jefferies, E., Ehsan, S., Jones, R. W., and Lambon Ralph, M. A. (2009). Semantic memory is key to binding phonology: converging evidence from immediate serial recall in semantic dementia and healthy participants. *Neuropsychologia* 47, 747–760. doi: 10.1016/j.neuropsychologia.2008.12.001
- Howard, D., and Nickels, L. (2005). Separating input and output phonology: semantic, phonological, and orthographic effects in short-term memory impairment. *Cogn. Neuropsychol.* 22, 42–77. doi: 10.1080/02643290342000582
- Hsiao, Y., Gao, Y., and MacDonald, M. C. (2014). Agent-patient similarity affects sentence structure in language production: evidence from subject omissions in Mandarin. *Front. Psychol.* 5:1015. doi: 10.3389/fpsyg.2014.01015
- Hughes, R. W., Chamberland, C., Tremblay, S., and Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *J. Mem. Lang.* 90, 126–146. doi: 10.1016/j.jml.2016.04.006
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., and Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a reintegration process in immediate serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1217–1232. doi: 10.1037/0278-7393.23.5.1217
- Hulme, C., and Tordoff, V. (1989). Working memory development: the effects of speech rate, word length, and acoustic similarity on serial recall. *J. Exp. Child Psychol.* 47, 72–87. doi: 10.1016/0022-0965(89)90063-5
- Hurlstone, M. J., Hitch, G. J., and Baddeley, A. D. (2014). Memory for serial order across domains: an overview of the literature and directions for future research. *Psychol. Bull.* 140, 339–373. doi: 10.1037/a0034221
- Ingvalson, E. M., Dhar, S., Wong, P. C. M., and Liu, H. (2015). Working memory training to improve speech perception in noise across languages. *J. Acoust. Soc. Am.* 137, 3477–3486. doi: 10.1121/1.4921601
- Jacquemot, C., Dupoux, E., Decouche, O., and Bachoud-Lévi, A.-C. (2006). Misperception in sentences but not in words: speech perception and the phonological buffer. *Cogn. Neuropsychol.* 23, 949–971. doi: 10.1080/02643290600625749
- Jefferies, E., Jones, R. W., Bateman, D., and Ralph, M. A. L. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cogn. Neuropsychol.* 22, 183–212. doi: 10.1080/02643290442000068
- Joanisse, M. F., and McClelland, J. L. (2015). Connectionist perspectives on language learning, representation and processing. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 235–247. doi: 10.1002/wcs.1340
- Joanisse, M. F., and Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: evidence from a connectionist model. *Brain Lang.* 86, 40–56. doi: 10.1016/s0093-934x(02)00533-3
- Jones, T., and Farrell, S. (2018). Does syntax bias serial order reconstruction of verbal short-term memory? *J. Mem. Lang.* 100, 98–122. doi: 10.1016/j.jml.2018.02.001
- Jones, G., Justice, L. V., Cabiddu, F., Lee, B. J., Iao, L. S., Harrison, N., et al. (2020). Does short-term memory develop? *Cognition* 198:104200. doi: 10.1016/j.cognition.2020.104200
- Jones, G., and Macken, B. (2015). Questioning short-term memory and its measurement: why digit span measures long-term associative learning. *Cognition* 144, 1–13. doi: 10.1016/j.cognition.2015.07.009
- Jones, D., and Macken, W. (2018). In the beginning was the deed: verbal short-term memory as object-oriented action. *Curr. Dir. Psychol. Sci.* 27, 351–356. doi: 10.1177/0963721418765796
- Jones, D. M., Macken, W. J., and Nicholls, A. P. (2004). The phonological store of working memory: is it phonological and is it a store? *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 656–674. doi: 10.1037/0278-7393.30.3.656
- Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149. doi: 10.1037/0033-295x.99.1.122
- Kaiser, E. (2012). Taking action: a cross-modal investigation of discourse-level representations. *Front. Psychol.* 3:156. doi: 10.3389/fpsyg.2012.00156
- Kalm, K., and Norris, D. (2014). The representation of order information in auditory-verbal short-term memory. *J. Neurosci.* 34, 6879–6886. doi: 10.1523/JNEUROSCI.4104-13.2014
- Kemper, S., Hoffman, L., Schmalzried, R., Herman, R., and Kieweg, D. (2011). Tracking talking: dual task costs of planning and producing speech for young versus older adults. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 18, 257–279. doi: 10.1080/13825585.2010.527317
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J.-E., and Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Dev. Sci.* 18, 146–154. doi: 10.1111/desc.12202
- Kolers, P. A., and Roediger, H. L. III. (1984). Procedures of mind. *J. Verbal Learn. Verbal Behav.* 23, 425–449. doi: 10.1016/S0022-5371(84)90282-2
- Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299
- Leminen, A., Smolka, E., Duñabeitia, J. A., and Pliatsikas, C. (2019). Morphological processing in the brain: the good (inflection), the bad (derivation) and the ugly (compounding). *Cortex* 116, 4–44. doi: 10.1016/j.cortex.2018.08.016
- Levelt, W. J. (1993). *Speaking: From Intention to Articulation*, Cambridge, MA: MIT Press.
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/s0140525x99001776



- Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010
- Lewandowsky, S., and Farrell, S. (2000). A reintegration account of the effects of speech rate, lexicality and word frequency in immediate serial recall. *Psychol. Res.* 63, 163–173. doi: 10.1007/pl00008175
- Lewis, R. L., Vasisht, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007
- Lombardi, L., and Potter, M. C. (1992). The regeneration of syntax in short term memory. *J. Mem. Lang.* 31, 713–733. doi: 10.1016/0749-596x(92)90036-w
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Lang. Cogn. Process.* 9, 157–201. doi: 10.1080/01690969408402115
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Front. Psychol.* 4:226. doi: 10.3389/fpsyg.2013.00226
- MacDonald, M. C. (2016). Speak, act, remember: the language-production basis of serial order and maintenance in verbal memory. *Curr. Direct. Psychol. Sci.* 25, 47–53. doi: 10.1177/0963721415620776
- MacDonald, M. C., and Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychol. Rev.* 109, 35–54; discussion 55–74. doi: 10.1037/0033-295x.109.1.35
- MacDonald, M. C., and Hsiao, Y. (2018). “Sentence comprehension,” in *Oxford Handbook of Psycholinguistics*, eds S.-A. Rueschmeyer and G. Gaskell (Oxford, UK: Oxford University Press), 171–196.
- MacDonald, M. C., and Seidenberg, M. S. (2006). “Constraint satisfaction accounts of lexical and sentence comprehension,” in *Handbook of Psycholinguistics*, eds M. Traxler and M. A. Gernsbacher (London, UK: Elsevier), 581–611.
- Macken, W. J., Taylor, J. C., and Jones, D. M. (2014). Language and short-term memory: the role of perceptual-motor affordance. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 1257–1270. doi: 10.1037/a0036845
- Majerus, S. (2009). “Verbal short-term memory and temporary activation of language representations: the importance of distinguishing item and order information,” in *Interactions Between Short-Term and Long-Term Memory in the Verbal Domain*, eds A. Thorn and M. Page (New York, NY: Psychology Press), 244–276.
- Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Front. Hum. Neurosci.* 7:357. doi: 10.3389/fnhum.2013.00357
- Majerus, S. (2019). Verbal working memory and the phonological buffer: the question of serial order. *Cortex* 112, 122–133. doi: 10.1016/j.cortex.2018.04.016
- Majerus, S., Attout, L., Artielle, M.-A., and Van der Kaa, M.-A. (2015). The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia* 77, 165–176. doi: 10.1016/j.neuropsychologia.2015.08.010
- Majerus, S., Belayachi, S., De Smedt, B., Leclercq, A. L., Martinez, T., Schmidt, C., et al. (2008). Neural networks for short-term memory for order differentiate high and low proficiency bilinguals. *NeuroImage* 42, 1698–1713. doi: 10.1016/j.neuroimage.2008.06.003
- Majerus, S., and D’Argembeau, A. (2011). Verbal short-term memory reflects the organization of long-term memory: further evidence from short-term memory for emotional words. *J. Mem. Lang.* 64, 181–197. doi: 10.1016/j.jml.2010.10.003
- Majerus, S., Norris, D., and Patterson, K. (2007). What does a patient with semantic dementia remember in verbal short-term memory? Order and sound but not words. *Cogn. Neuropsychol.* 24, 131–151. doi: 10.1080/02643290600989376
- Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., et al. (2006). The left intraparietal sulcus and verbal short-term memory: focus of attention or serial order? *Magn. Reson. Med.* 32, 880–891. doi: 10.1016/j.neuroimage.2006.03.048
- Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., and Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: evidence from an incidental phonotactic learning paradigm. *J. Mem. Lang.* 51, 297–306. doi: 10.1016/j.jml.2004.05.002
- Martin, N., Dell, G. S., Saffran, E. M., and Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain Lang.* 47, 609–660. doi: 10.1006/brln.1994.1061
- Martin, N., and Saffran, E. M. (1992). A computational account of deep dysphasia: evidence from a single case study. *Brain Lang.* 43, 240–274. doi: 10.1016/0093-934x(92)90130-7
- Martin, N., Saffran, E. M., and Dell, G. S. (1996). Recovery in deep dysphasia: evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain Lang.* 52, 83–113. doi: 10.1006/brln.1996.0005
- Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. *Brain Lang.* 32, 159–192. doi: 10.1016/0093-934x(87)90122-2
- Martin, R. C., and Breedin, S. D. (1992). Dissociations between speech perception and phonological short-term memory deficits. *Cogn. Neuropsychol.* 9, 509–534. doi: 10.1080/02643299208252070
- Martin, R. C., and Freedman, M. L. (2001). Short-term retention of lexical-semantic representations: implications for speech production. *Memory* 9, 261–280. doi: 10.1080/09658210143000173
- Martin, R. C., and He, T. (2004). Semantic short-term memory and its role in sentence processing: a replication. *Brain Lang.* 89, 76–82. doi: 10.1016/s0093-934x(03)00300-6
- Martin, R. C., Lesch, M. F., and Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *J. Mem. Lang.* 41, 3–29. doi: 10.1006/jmla.1999.2637
- Martin, R. C., and Romani, C. (1994). Verbal working memory and sentence comprehension: a multiple-components view. *Neuropsychology* 8, 506–523. doi: 10.1037/0894-4105.8.4.506
- Martin, R. C., Shelton, J. R., and Yaffee, L. S. (1994). Language processing and working memory: neuropsychological evidence for separate phonological and semantic capacities. *J. Mem. Lang.* 33, 83–111. doi: 10.1006/jmla.1994.1005
- Mayberry, M. R., Crocker, M. W., and Knoeferle, P. (2009). Learning to attend: a connectionist model of situated language comprehension. *Cogn. Sci.* 33, 449–496. doi: 10.1111/j.1551-6709.2009.01019.x
- McCauley, S. M., and Christiansen, M. H. (2014). Acquiring formulaic language: a computational model. *Mental Lexicon* 9, 419–436. doi: 10.1075/ml.9.3.03mcc
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356. doi: 10.1016/j.tics.2010.06.002
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158
- Miller, G. A., and Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *Am. J. Psychol.* 63, 176–185. doi: 10.2307/1418920
- Monaghan, P., and Woollams, A. M. (2017). “Implementing the ‘Simple’ model of reading deficits: a connectionist investigation of interactivity,” in *Neurocomputational Models of Cognitive Development and Processing: Proceedings of the 14th Neural Computation and Psychology Workshop*, (Lancaster, UK: Lancaster University), 69–81.
- Montag, J. L., Matsuki, K., Kim, J. Y., and MacDonald, M. C. (2017). Language specific and language general motivations of production choices: a multi-clause and multi-language investigation. *Collabra Psychol.* 3, 1–22. doi: 10.1525/collabra.94
- Murdock, B. B. Jr. (1962). The serial position effect of free recall. *J. Exp. Psychol.* 64, 482–488. doi: 10.1037/h0045106
- Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychol. Bull.* 143, 992–1009. doi: 10.1037/bul0000108
- Nozari, N., and Dell, G. S. (2012). Feature migration in time: reflection of selective attention on speech errors. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1084–1090. doi: 10.1037/a0026933
- Nozari, N., and Novick, J. (2017). Monitoring and control in language production. *Curr. Direct. Psychol. Sci.* 26, 403–410. doi: 10.1177/0963721417702419
- Nozari, N., and Thompson-Schill, S. L. (2013). More attention when speaking: does it help or does it hurt? *Neuropsychologia* 51, 2770–2780. doi: 10.1016/j.neuropsychologia.2013.08.019
- Nozari, N., Trueswell, J. C., and Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychon. Bull. Rev.* 23, 1942–1953. doi: 10.3758/s13423-016-1068-8
- Nygaard, L. C., Cook, A. E., and Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition* 112, 181–186. doi: 10.1016/j.cognition.2009.04.001

- Oberauer, K., Farrell, S., Jarrold, C., and Lewandowsky, S. (2016). What limits working memory capacity? *Psychol. Bull.* 142, 758–799. doi: 10.1037/bul0000046
- Oberauer, K., Jones, T., and Lewandowsky, S. (2015). The Hebb repetition effect in simple and complex memory span. *Mem. Cognit.* 43, 852–865. doi: 10.3758/s13421-015-0512-8
- Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., and Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *J. Mem. Lang.* 69, 506–526. doi: 10.1016/j.jml.2013.07.001
- Page, M. P. A., Madge, A., Cumming, N., and Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: evidence suggesting a common locus. *J. Mem. Lang.* 56, 49–64. doi: 10.1016/j.jml.2006.09.002
- Page, M. P. A., and Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3737–3753. doi: 10.1098/rstb.2009.0173
- Patterson, K., Graham, N., and Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *J. Cogn. Neurosci.* 6, 57–69. doi: 10.1162/jocn.1994.6.1.57
- Penfield, W., and Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *AMA. Arch. Neurol. Psychiatry* 79, 475–497. doi: 10.1001/archneurpsyc.1958.02340050003001
- Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., et al. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* 9:963. doi: 10.1038/s41467-018-03068-4
- Perham, N., Marsh, J. E., and Jones, D. M. (2009). Syntax and serial recall: how language supports short-term memory for order. *Q. J. Exp. Psychol.* 62, 1285–1293. doi: 10.1080/17470210802635599
- Pickering, M. J., and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychol. Bull.* 134, 427–459. doi: 10.1037/0033-2909.134.3.427
- Plaut, D. C., and Kello, C. T. (1999). “The emergence of phonology from the interplay of speech comprehension and production: a distributed connectionist approach,” in *The Emergence of Language*, ed. B. MacWhinney (Mahwah, NJ: Psychology Press), 381–415.
- 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
- Poirier, M., and Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Can. J. Exp. Psychol.* 50, 408–412. doi: 10.1037/1196-1961.50.4.408
- Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., and Tolan, A. (2014). Order recall in verbal short-term memory: the role of semantic networks. *Mem. Cognit.* 43, 489–499. doi: 10.3758/s13421-014-0470-6
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38. doi: 10.1016/j.neuroscience.2005.06.005
- Potter, M. C., and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *J. Mem. Lang.* 38, 265–282. doi: 10.1006/jmla.1997.2546
- Ramscar, M., and Port, R. F. (2016). How spoken languages work in the absence of an inventory of discrete units. *Lang. Sci.* 53, 58–74. doi: 10.1016/j.langsci.2015.08.002
- Roodenrys, S., and Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Neuropsychologia* 28, 29–33. doi: 10.1037/0278-7393.28.1.29
- Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., and Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 1019–1034. doi: 10.1037/0278-7393.28.6.1019
- Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., and van der Wel, R. (2007). The problem of serial order in behavior: lashley’s legacy. *Hum. Mov. Sci.* 26, 525–554. doi: 10.1016/j.humov.2007.04.001
- Rubin, R. D., Watson, P. D., Duff, M. C., and Cohen, N. J. (2014). The role of the hippocampus in flexible cognition and social behavior. *Front. Hum. Neurosci.* 8:742. doi: 10.3389/fnhum.2014.00742
- Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., and Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *J. Mem. Lang.* 36, 382–405. doi: 10.1006/jmla.1996.2489
- Saffran, E. M., and Martin, N. (1997). Language and auditory-verbal short-term memory impairments: evidence for common underlying processes. *Cogn. Neuropsychol.* 14, 641–682. doi: 10.1080/026432997381402
- Saint-Aubin, J., and Poirier, M. (1999). The influence of long-term memory factors on immediate serial recall: an item and order analysis. *Int. J. Psychol.* 34, 347–352. doi: 10.1080/002075999399675
- Savill, N. J., Cornelissen, P., Pahor, A., and Jefferies, E. (2019). RTMS evidence for a dissociation in short-term memory for spoken words and nonwords. *Cortex* 112, 5–22. doi: 10.1016/j.cortex.2018.07.021
- Savill, N., Cornelissen, P., Whiteley, J., Woollams, A., and Jefferies, E. (2019). Individual differences in verbal short-term memory and reading aloud: semantic compensation for weak phonological processing across tasks. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 1815–1831. doi: 10.1037/xlm0000675
- Savill, N., Ellis, R., Brooke, E., Koa, T., Ferguson, S., Rojas-Rodriguez, E., et al. (2017). Keeping it together: semantic coherence stabilizes phonological sequences in short-term memory. *Mem. Cognit.* 46, 426–437. doi: 10.3758/s13421-017-0775-3
- Schmidtke, D., Conrad, M., and Jacobs, A. M. (2014). Phonological iconicity. *Front. Psychol.* 5:80. doi: 10.3389/fpsyg.2014.00080
- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21. doi: 10.1136/jnnp.20.1.11
- Seidenberg, M. S. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science* 275, 1599–1603. doi: 10.1126/science.275.5306.1599
- Seidenberg, M. S., and Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends Cogn. Sci.* 4, 353–361. doi: 10.1016/s1364-6613(00)01515-1
- Seidenberg, M. S., and MacDonald, M. C. (2018). The impact of language experience on language and reading: a statistical learning approach. *Top. Lang. Disord.* 38, 66–83. doi: 10.1097/tld.0000000000000144
- Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568. doi: 10.1037/0033-295x.96.4.523
- Seidenberg, M. S., and Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *J. Exp. Psychol. Hum. Learn. Mem.* 5, 546–554. doi: 10.1037/0278-7393.5.6.546
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., and Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: some limitations of knowledge-based processing. *Cogn. Psychol.* 14, 489–537. doi: 10.1016/0010-0285(82)90017-2
- Shallice, T., and Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia* 15, 729–735. doi: 10.1016/0028-3932(77)90002-1
- Shallice, T., and Papagno, C. (2019). Impairments of auditory-verbal short-term memory: do selective deficits of the input phonological buffer exist? *Cortex* 112, 107–121. doi: 10.1016/j.cortex.2018.10.004
- Shallice, T., and Warrington, E. K. (1970). Independent functioning of verbal memory stores: a neuropsychological study. *Q. J. Exp. Psychol.* 22, 261–273. doi: 10.1080/00335557043000203
- Shallice, T., and Warrington, E. K. (1974). The dissociation between short term retention of meaningful sounds and verbal material. *Neuropsychologia* 12, 553–555. doi: 10.1016/0028-3932(74)90087-6
- Siegelman, M., Blank, I. A., Mineroff, Z., and Fedorenko, E. (2019). An attempt to conceptually replicate the dissociation between syntax and semantics during sentence comprehension. *Neuroscience* 413, 219–229. doi: 10.1016/j.neuroscience.2019.06.003
- Silveri, M. C., and Cappa, A. (2003). Segregation of the neural correlates of language and phonological short-term memory. *Cortex* 39, 913–925. doi: 10.1016/s0010-9452(08)70870-0
- Solomon, E. S., and Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cogn. Psychol.* 49, 1–46. doi: 10.1016/j.cogpsych.2003.10.001
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., and Laine, M. (2017). Working memory training revisited: a multi-level meta-analysis of n-back training studies. *Psychon. Bull. Rev.* 24, 1077–1096. doi: 10.3758/s13423-016-1217-0

- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition* 116, 71–86. doi: 10.1016/j.cognition.2010.04.002
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U S A* 106, 10587–10592. doi: 10.1073/pnas.0903616106
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Stuart, G. P., and Hulme, C. (2009). “Lexical and semantic influences on immediate serial recall: a role for reintegration,” in *Interactions Between Short-Term and Long-Term Memory in the Verbal Domain*, eds A. Thorne and M. Page (New York, NY: Psychology Press), 157–176.
- Szewczyk, J. M., Marecka, M., Chiat, S., and Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: evidence from a multi-factorial analysis. *Cognition* 179, 23–36. doi: 10.1016/j.cognition.2018.06.002
- Szmalc, A., Duyck, W., Vandierendonck, A., Mata, A. B., and Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Q. J. Exp. Psychol.* 62, 435–443. doi: 10.1080/17470210802386375
- Tanida, Y., Nakayama, M., and Saito, S. (2019). The interaction between temporal grouping and phonotactic chunking in short-term serial order memory for novel verbal sequences. *Memory* 27, 507–518. doi: 10.1080/09658211.2018.1532008
- Tanida, Y., Ueno, T., Lambon Ralph, M. A., and Saito, S. (2015). The roles of long-term phonotactic and lexical prosodic knowledge in phonological short-term memory. *Mem. Cognit.* 43, 500–519. doi: 10.3758/s13421-014-0482-2
- Thorn, A. S. C., and Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 729–735. doi: 10.1037/0278-7393.31.4.729
- Torreira, F., Bögel, S., and Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Front. Psychol.* 6:284. doi: 10.3389/fpsyg.2015.00284
- Ueno, T., Saito, S., Rogers, T. T., and Lambon Ralph, M. A. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron* 72, 385–396. doi: 10.1016/j.neuron.2011.09.013
- Ueno, T., Saito, S., Saito, A., Tanida, Y., Patterson, K., and Lambon Ralph, M. A. (2014). Not lost in translation: generalization of the primary systems hypothesis to Japanese-specific language processes. *J. Cogn. Neurosci.* 26, 433–446. doi: 10.1162/jocn\_a\_00467
- Vallar, G., and Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: a neuropsychological case study. *Cogn. Neuropsychol.* 1, 121–141. doi: 10.1080/02643298408252018
- Van de Cavey, J., and Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions and language. *Cognition* 146, 172–184. doi: 10.1016/j.cognition.2015.09.013
- Van Dyke, J. A., and Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Lang. Linguist. Compass* 6, 193–211. doi: 10.1002/lnc3.330
- Walker, I., and Hulme, C. (1999). Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 1256–1271. doi: 10.1037/0278-7393.25.5.1256
- Warrington, E. K., Logue, V., and Pratt, R. T. C. (1971). The anatomical localisation of selective impairment of auditory verbal short-term memory. *Neuropsychologia* 9, 377–387. doi: 10.1016/0028-3932(71)90002-9
- Warrington, E. K., and Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain* 92, 885–896. doi: 10.1093/brain/92.4.885
- Warrington, E. K., and Shallice, T. (1972). Neuropsychological evidence of visual storage in short-term memory tasks. *Q. J. Exp. Psychol.* 24, 30–40. doi: 10.1080/14640747208400265
- Watkins, O. C., and Watkins, M. J. (1977). Serial recall and the modality effect: effects of word frequency. *J. Exp. Psychol. Hum. Learn. Mem.* 3, 712–718. doi: 10.1037/0278-7393.3.6.712
- Weiner, E. J., and Labov, W. (1983). Constraints on the agentless passive. *J. Linguist.* 19, 29–58. doi: 10.1017/s0022226700007441
- Willits, J. A., Amato, M. S., and MacDonald, M. C. (2015). Language knowledge and event knowledge in language use. *Cogn. Psychol.* 78, 1–27. doi: 10.1016/j.cogpsych.2015.02.002
- Willits, J. A., Seidenberg, M. S., and Saffran, J. R. (2014). Distributional structure in language: contributions to noun-verb difficulty differences in infant word recognition. *Cognition* 132, 429–436. doi: 10.1016/j.cognition.2014.05.004
- Yue, Q., Martin, R. C., Hamilton, A. C., and Rose, N. S. (2019). Non-perceptual regions in the left inferior parietal lobe support phonological short-term memory: evidence for a buffer account? *Cereb. Cortex* 29, 1398–1413. doi: 10.1093/cercor/bhy037
- Yuzawa, M., and Saito, S. (2006). The role of prosody and long-term phonological knowledge in Japanese children’s nonword repetition performance. *Cogn. Dev.* 21, 146–157. doi: 10.1016/j.cogdev.2006.01.003
- Zakariás, L., Kelly, H., Salis, C., and Code, C. (2019). The methodological quality of short-term/working memory treatments in poststroke aphasia: A systematic review. *J. Speech Lang. Hear. Res.* 62, 1979–2001. doi: 10.1044/2018\_JSLHR-L-18-0057

**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 Schwering and MacDonald. 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.



# Post-training Load-Related Changes of Auditory Working Memory – An EEG Study

Helene Gudi-Mindermann<sup>1\*</sup>, Johanna M. Rimmele<sup>2,3</sup>, Patrick Bruns<sup>1</sup>,  
Niels A. Kloosterman<sup>4</sup>, Tobias H. Donner<sup>2</sup>, Andreas K. Engel<sup>2</sup> and Brigitte Röder<sup>1</sup>

<sup>1</sup> Department of Biological Psychology and Neuropsychology, University of Hamburg, Hamburg, Germany, <sup>2</sup> Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, <sup>3</sup> Department of Neuroscience, Max-Planck-Institute for Empirical Aesthetics, Frankfurt am Main, Germany, <sup>4</sup> Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Max-Planck-Institute for Human Development, Berlin, Germany

## OPEN ACCESS

### Edited by:

Vitória Plai,  
Radboud University Nijmegen,  
Netherlands

### Reviewed by:

Caroline Beese,  
Max Planck Institute for Human  
Cognitive and Brain Sciences,  
Germany  
Malte Wöstmann,  
University of Lübeck, Germany

### \*Correspondence:

Helene Gudi-Mindermann  
helene.gudi-mindermann@posteo.de

### Specialty section:

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 20 November 2019

**Accepted:** 19 February 2020

**Published:** 17 March 2020

### Citation:

Gudi-Mindermann H,  
Rimmele JM, Bruns P,  
Kloosterman NA, Donner TH,  
Engel AK and Röder B (2020)  
Post-training Load-Related Changes  
of Auditory Working Memory – An  
EEG Study.  
Front. Hum. Neurosci. 14:72.  
doi: 10.3389/fnhum.2020.00072

Working memory (WM) refers to the temporary retention and manipulation of information, and its capacity is highly susceptible to training. Yet, the neural mechanisms that allow for increased performance under demanding conditions are not fully understood. We expected that post-training efficiency in WM performance modulates neural processing during high load tasks. We tested this hypothesis, using electroencephalography (EEG) ( $N = 39$ ), by comparing source space spectral power of healthy adults performing low and high load auditory WM tasks. Prior to the assessment, participants either underwent a modality-specific *auditory* WM training, or a modality-irrelevant *tactile* WM training, or were not trained (active control). After a modality-specific training participants showed higher behavioral performance, compared to the control. EEG data analysis revealed general effects of WM load, across all training groups, in the theta-, alpha-, and beta-frequency bands. With increased load theta-band power increased over frontal, and decreased over parietal areas. Centro-parietal alpha-band power and central beta-band power decreased with load. Interestingly, in the high load condition a tendency toward reduced beta-band power in the right medial temporal lobe was observed in the modality-specific WM training group compared to the modality-irrelevant and active control groups. Our finding that WM processing during the high load condition changed after modality-specific WM training, showing reduced beta-band activity in voice-selective regions, possibly indicates a more efficient maintenance of task-relevant stimuli. The general load effects suggest that WM performance at high load demands involves complementary mechanisms, combining a strengthening of task-relevant and a suppression of task-irrelevant processing.

**Keywords:** auditory working memory, working memory load, post-training plasticity, EEG, source space

## INTRODUCTION

Working memory (WM) has been defined as the ability to temporarily maintain and manipulate stored information (Baddeley, 2003; D'Esposito, 2007; Jonides et al., 2008). Language processing highly relies on WM processes, as information needs to be maintained and integrated over time, for example during phrasal or sentence level processing (Montgomery, 2000; Emmorey et al., 2017).



Particularly verbal WM is crucial for speech comprehension (Buchsbaum and D'Esposito, 2019), but speech comprehension additionally requires the processing of extralinguistic cues, such as voice features and prosody (Larrouy-Maestri et al., 2013; Hellbernd and Sammler, 2016). Language learning can benefit from prosodic cues, suggesting interactions of verbal and extralinguistic memory (Schon et al., 2008; de Diego-Balaguer et al., 2015). Here, in a voice recognition task, we focus on auditory WM of extralinguistic cues. WM capacity varies among individuals (Luck and Vogel, 2013), but can be improved by training (Morrison and Chein, 2011), such that tasks of higher difficulty can be managed successfully following training. The present study investigated the neural mechanisms that allow enhanced auditory WM performance at high difficulty levels following WM training.

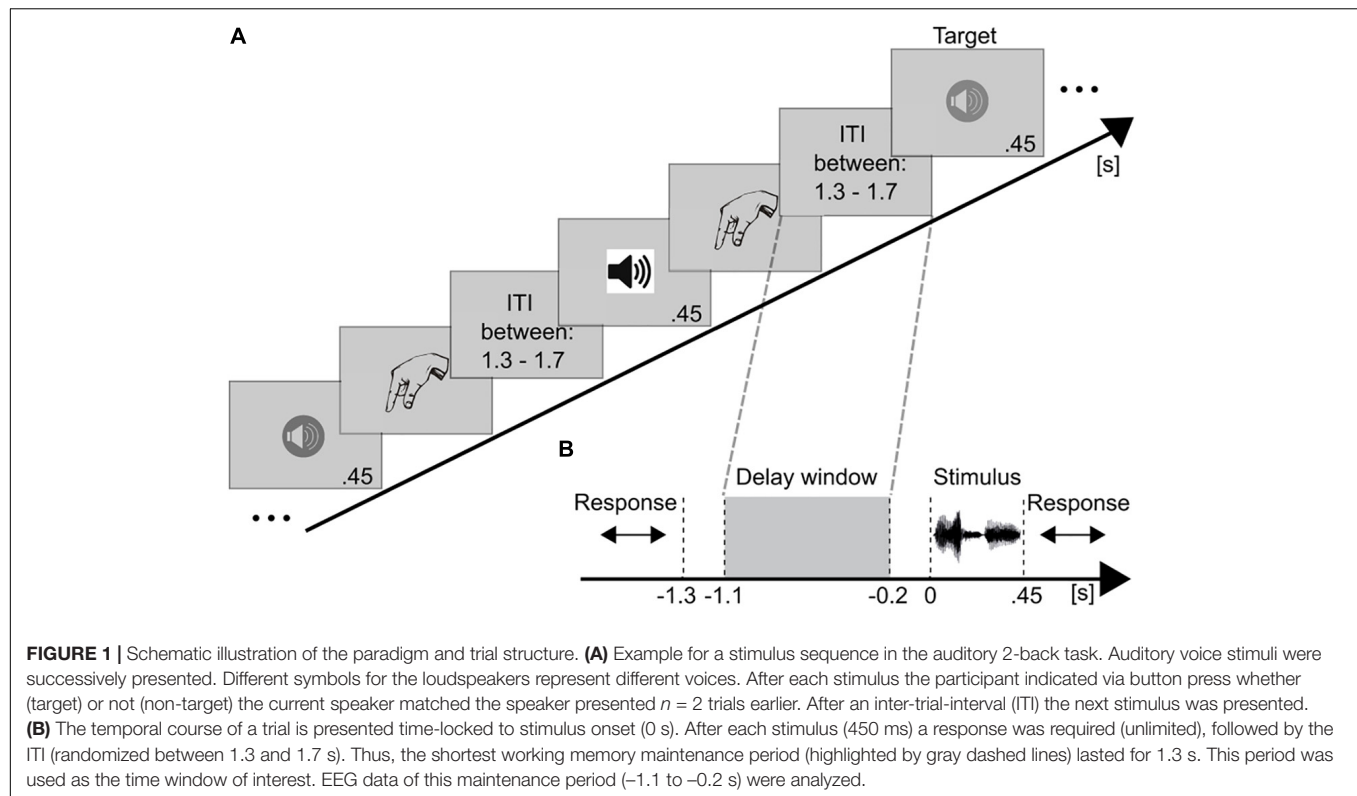
A classical paradigm to assess WM processing at several difficulty levels is the *n*-back task. In *n*-back tasks participants receive a stimulus sequence and have to decide whether or not the current stimulus matches the stimulus presented *n* trials before (Figure 1A). The *n*, thus, represents the adjustable load factor; the higher the *n*, the higher the WM demands. Electroencephalography (EEG) and magnetoencephalography (MEG) studies have reported a parametric relationship between increasing WM load and oscillatory activity (typically neuronal power increases), mainly in the theta- (Krause et al., 2000; Jensen and Tesche, 2002; Hellbernd and Sammler, 2016) and gamma-bands (Kaiser et al., 2003; Palva et al., 2011; Roux et al., 2012). Furthermore, this WM-related theta- and gamma-band activity has been predominantly associated with frontal (Jensen and Tesche, 2002; Barnett et al., 2008; Dupoux et al., 2008; Kaiser et al., 2009) and parietal areas (Sauseng et al., 2009; Scharinger et al., 2017; Kapeller et al., 2018), which are commonly considered to represent the core WM network (for reviews see Wager and Smith, 2003; D'Esposito and Postle, 2015; Eriksson et al., 2015). In WM tasks, the fronto-parietal theta- and gamma-band oscillatory activity have been suggested to reflect retention of relevant information (for a review see Roux and Uhlhaas, 2014).

In contrast, the functional roles of alpha- and beta-band activity in WM have not been yet clearly defined (for a review see Roux and Uhlhaas, 2014). For example, alpha-band modulations have been consistently reported to vary with WM load. Nevertheless, the direction of the relationship remains controversial, reporting both load-induced alpha-band increases (Scheeringa et al., 2009; van Dijk et al., 2010) and load-induced alpha-band decreases (Haegens et al., 2014; Chen and Huang, 2015). A positive relation between WM load and synchronized alpha power (Jokisch and Jensen, 2007; Medendorp et al., 2007) has been interpreted in the light of the inhibition-timing hypothesis (Klimesch et al., 2007), which has linked increases in alpha-band amplitude with an inhibition of task-irrelevant brain regions. Conversely, a decrease in alpha-band amplitude has been associated with a release from inhibition (Klimesch et al., 2007; Klimesch, 2012) and an overall enhanced cortical engagement at higher load demands (Gevins et al., 1997; Palomaki et al., 2012; Chen and Huang, 2015). Similarly, the contribution of beta-band power to WM is not well understood. Although several studies have reported beta-band activity to be modulated by WM

load (Deiber et al., 2007; Chen and Huang, 2015; Palva et al., 2011; Scharinger et al., 2017), beta-band power has not been particularly associated with a specific functional role in WM tasks (for reviews see Roux and Uhlhaas, 2014; D'Esposito and Postle, 2015; Eriksson et al., 2015). Some studies, however, propose a role for (low) beta-band activity in short-term memory (Kopell et al., 2011), cognitive WM processes (Scharinger et al., 2017) or more general integrative functions (Donner and Siegel, 2011) such as the maintenance of the current (motor or cognitive) state, as it is required during WM delay periods (Engel and Fries, 2010). Furthermore, functional connectivity particularly in the beta-band has been found to be enhanced between WM-relevant frontal and parietal regions in WM tasks (Palva et al., 2010, 2011; Salazar et al., 2012).

WM training has been shown to alter oscillatory activity in WM-relevant regions. For example, behavioral training gains in WM tasks were accompanied by training-induced increases in frontal (Gevins et al., 1997; Langer et al., 2013) and fronto-parietal (Jausovec and Jausovec, 2012) theta-band power, suggesting that training strengthened the WM network, thereby, facilitating WM performance. Furthermore, functional connectivity has been found to be altered by WM training (e.g., Langer et al., 2013; Astle et al., 2015; Gudi-Mindermann et al., 2018; Rimmele et al., 2019). While power increases are thought to reflect local processing (Donner and Siegel, 2011; Buzsaki and Wang, 2012), functional connectivity is assumed to reflect the degree of the temporal alignment of brain activity in distributed networks (Engel and Singer, 2001; Varela et al., 2001; Womelsdorf et al., 2007). A training study in children reported that 20–25 sessions of a computerized verbal and spatial WM training relative to a control group enhanced the coupling of resting state MEG activity between a fronto-parietal network and lateralized occipital and inferior temporal regions (Astle et al., 2015). In a consecutive investigation of the same children, the authors reported training-induced increases in cross-frequency phase amplitude coupling (Barnes et al., 2016): Following training, gamma-band power (~90 Hz) in inferior-parietal and temporal areas was phase-locked to a slower beta rhythm (16 Hz) at fronto-parietal areas. These findings demonstrate that neural mechanisms involved in WM processing change as a function of training, as indicated by training-induced changes in both oscillatory power and functional connectivity. Typically, such changes are assessed by contrasting EEG activity prior to and after training while the same task is performed. Therefore, it remains unclear how WM processing at different load demands is affected by the training-induced neuronal changes. While trained and untrained individuals might perform similar at low load demands, the question arises how the neurophysiological training-induced changes facilitate WM processing at high load demands, i.e., how the underlying mechanisms are altered in trained relative to non-trained individuals.

The present study investigated how post-training performance proficiency affects the neural mechanisms involved in successful WM processing at high load demands. Healthy adults performed a low load (2-back) and a high load (adaptive *n*-back) auditory WM task with voices. In the high load condition, the *n*-back level was continuously adjusted to the participants performance, such



that participants were continuously performing at their capacity limits. This ensured that participants, despite interindividual differences in WM capacity, always performed at high load demands. Prior to the test session, participants were adaptively trained (adaptive  $n$ -back) either with the same auditory voice stimuli as in the test session (auditory training group), or with task-irrelevant tactile stimuli (tactile training group), or were not adaptively trained, i.e., the active control group performed a 1-back task throughout all “training” sessions. The EEG power during the maintenance phase of the auditory WM task (**Figure 1B**) was compared between the low load (2-back task) and the high load condition (adaptive  $n$ -back task). We particularly focused on whether load-related changes in neuronal power differ between training groups, since we were interested in whether the neural correlates of WM processing at high load demands differ as a function of post-training performance efficiency. The increase in WM proficiency of trained participants was expected to result in WM processing changes. Importantly, if increased proficiency would result in mere activation differences in the networks activated during low load demands, no group differences would be expected during high load processing, as load levels were adjusted to the individual performance limit across groups. Instead, proficiency-related changes in WM processing should be present despite adjusted load levels across groups. Changes in WM networks were expected to be characterized by more efficient maintenance mechanisms. As suggested by previous studies, such increases in efficiency may be indicated for instance by a shift from attentional control processes to task-specific functions, involving perceptual

processing, thus, by a shift from anterior to posterior activity (Buschkuhl et al., 2012).

## MATERIALS AND METHODS

### Participants

Forty-one healthy adults participated in the study and were pseudo-randomly assigned to three groups (cf. section “Experimental Procedure”). The data of two participants had to be discarded from the analyses, due to a decreased post-training performance relative to their pre-training performance ( $\geq 2SD$  from the mean pre-post-training difference). Thus, the final data included data sets of the remaining 39 participants (**Table 1**). Participants in the three groups did not differ regarding their sex ( $\chi^2(2) = 0.351$ ,  $p = 0.839$ ), age ( $F(2,36) = 0.484$ ,  $p = 0.620$ ), and education (more vs. less than ten years of schooling ( $\chi^2(2) = 2.60$ ,  $p = 0.273$ ). All participants were right-handed, had normal or corrected-to-normal vision, and normal

**TABLE 1 |** Demographic data of 39 participants<sup>a</sup>.

	<i>n</i>	Gender (females)	Mean age in years (range)	Education (>10 years of schooling)
AG	14	6	34 (23–55)	12
TG	11	6	30 (21–48)	9
CG	14	7	33 (21–55)	14

<sup>a</sup>AG, auditory training group; TG, tactile training group; CG, active control group.

hearing (self-report). None of the participants had a history of neurological or psychiatric disorders (self-report). Informed consent was obtained from all individual participants included in the study. Participants received monetary compensation for participation. The study was approved by the German Psychological Association.

Variables such as the perceived current stress level, wellbeing, and intelligence have been shown to affect WM performance (e.g., Ashby et al., 1999; Luethi et al., 2008; Luck and Vogel, 2013). To control for such confounding effects, all participants performed the German version of the Perceived Stress Questionnaire (PSQ-20; Levenstein et al., 1993; German modified version: Fliege et al., 2001) and a wellbeing scale [German: Habituelle Subjektive Wohlbefindens Skala (HSWBS); Dalbert, 1992]. An estimation of the verbal intelligence score was obtained through the MWT-B (German Mehrfachwahl-Wortschatz-Test; Lehrl, 2005). The three groups did not differ in any of the assessed psychological variables (PSQ-R20:  $F(36) = 0.16$ ,  $p = 0.854$ ; HSWBS:  $F(35) = 0.03$ ,  $p = 0.970$ ; MWT-B:  $F(35) = 0.91$ ,  $p = 0.412$ ).

## Experimental Procedures

The reported data were part of a larger WM training study (Table 2), comprising pre-training EEG and MEG recordings, 4 sessions of behavioral WM training, post-training EEG and MEG recordings, and a final magnetic resonance imaging (MRI) session. Here, only behavioral and EEG data from the posttraining assessment and behavioral data from the training sessions will be reported. The other data have been published elsewhere. Previous publications investigated differences of the neural networks changed by WM training between sighted and congenital blind adults, analyzing the WM processing during a 2-back task prior to and after WM training (Gudi-Mindermann et al., 2018; Rimmele et al., 2019). In contrast, the present study focused on neurophysiological differences in WM processing at individual capacity limits, which have been enhanced (as shown on behavioral level; cf. section “Behavioral Performance”) by a previous training treatment, analyzing post-training WM processing of a demanding adaptive  $n$ -back task relative to a low load 2-back task.

**TABLE 2 |** Full training procedure and study design<sup>a</sup>.

Pre-test	Training			Post-test	
EEG/MEG	Behavioral WM Training			EEG/MEG	MR/DTI
<b>Auditory:</b> 2-back task	AG	TG	CG	<b>Auditory:</b> 2-back task adaptive $n$ -back task	
	Auditory:	Tactile:	Auditory:		
			1-back task		
	adaptive $n$ -back task				
<b>Tactile:</b> 2-back task			Tactile:	<b>Tactile:</b> 2-back task adaptive $n$ -back task	
			1-back task		
1 Session	4 Sessions	4 Sessions	4 Sessions	1 Session	1 Session

<sup>a</sup>AG, auditory training group; TG, tactile training group; CG, active control group. Bold black box highlights the conditions that have been analyzed in the present paper. See text for details.

In the pre-training sessions, participants performed an auditory and a tactile 2-back task. After each stimulus the participants had to indicate via button press with one of two fingers (index finger and middle finger of one hand; responding hand and finger were cross-balanced across participants) whether the stimulus was identical or not (target – 1/3: non-target – 2/3 distribution) to the stimulus presented 2 trials ago. Following the participants' response, the next stimulus was presented after an inter-stimulus-interval (ISI) of a varying length, randomized between 1.3 and 1.7 s (Figure 1A).

Auditory stimuli consisted of the pseudo-word “befa” spoken by ten individuals (i.e., 10 different stimuli, 5 females and 5 males, stimulus duration: 450 ms, digitized at 44, 100 Hz and peak normalized at 65–75 dB). Participants had to match the speakers' identity. The tactile stimuli were applied via Braille stimulators (QuaeroSys Medical Devices, Schotten, Germany). Five stimulators were attached to the fingers of one hand. The Braille stimulation generated a tactile motion percept by sequentially activating pairs of pins that were horizontally organized in four-by-two rows ( $4 \times 112.5$  ms, total stimulus duration 450 ms). The apparent motion either started at the fingertips (downward motion) or moved toward the fingertips (upward motion), resulting in ten different stimuli (i.e., 2 possible motion percepts at 5 fingers of one hand). Participants had to match the finger and the motion direction. Thus, the characteristics of the task, i.e., the length of each stimulus and the number of different stimuli, were held constant across the auditory and the tactile tasks.

Participants were pseudo-randomly assigned to three groups: auditory training group (AG), tactile training group (TG), and active control group (CG). In the AG and the TG an adaptive  $n$ -back task either with voices or with tactile stimuli was used during training. WM load was adaptively changed by adjusting the  $n$ -back level: The first block of the first training session started with a 2-back task. After every block the  $n$  was individually adjusted. The  $n$  increased by one if the performance exceeded both a hit rate of 70% and a correct rejection rate of 75%. The  $n$  decreased by one if the performance dropped below 60% in either hits or correct rejections. Otherwise, the  $n$  did not change. Spoken feedback after each block informed participants about their performance and announced the  $n$ -level of the upcoming block. Participants were instructed to strive toward the highest possible  $n$ -level and to prioritize accuracy over speed. Every consecutive training session started with the highest  $n$  that was reached at least three times during the previous training session. Each block consisted of  $30 + n$  trials. For all  $n$ -back levels, sequences of targets and non-targets were constructed pseudo-randomly: The position of the targets ( $10 \times n$ -back targets) varied randomly, while a fixed number of interfering distractors were incorporated ( $3 \times “n-minus-1”$ -back lures and  $3 \times “n-plus-1”$ -back lures per block). A training session comprised 30 blocks and lasted typically for about 2 h, resulting in approximately 8 h of training per participant. All sessions took place on consecutive days or with no more than three days in between.

The AG was trained with auditory stimuli while the TG was trained with tactile stimuli. The CG performed a constant 1-back task throughout the four sessions, while the modality (auditory,

tactile) of the task altered every session. An active control group was included to control for training-unrelated effects, such as familiarity with the stimulus material and procedure, as well as for non-specific effects resulting from being engaged in the training paradigm over several days. After training, the EEG was recorded during the auditory 2-back task (low load condition) and the auditory adaptive  $n$ -back task (high load condition). Note, with respect to the auditory task, the AG received a training in the same modality in which the post-training task was carried out, while the TG was trained in a modality that differed from the modality of the post-training task. In the following, to highlight this difference, the terms modality-specific training (AG) and modality-irrelevant training (TG) are used. The CG did not receive adaptive training.

In the high load condition, we aimed at testing participants' performance at the individual's capacity limit, irrespective of the group affiliation. Therefore, the  $n$ -level was set as follows: In case of the AG the adaptive  $n$ -back version started with the highest  $n$  that was reached at least three times during the last training session. In the case of TG the adaptive  $n$ -back version started 3 levels below the highest  $n$  that was reached at least three times during the last training session, to account for the modality switch between the tactile training and the auditory post-training task. For all CG participants the adaptive  $n$ -back version started with a 3-back task. The adaptive nature of the  $n$ -back task was kept for all three groups throughout the high load condition, to continuously adjust load demands to individual performance, while accounting for instantaneous learning and/or fatigue effects. The post-training EEG session comprised twelve blocks of the auditory 2-back task and fourteen blocks of the auditory adaptive  $n$ -back task. Participants were blindfolded during all EEG and training sessions, since this study was part of a bigger project, which additionally included blind individuals.

## EEG Data Acquisition and Preprocessing

The EEG was continuously recorded with 94 Ag/AgCl scalp electrodes (EasyCap GmbH, Herrsching, Germany), mounted in a cap according to the 10-5 system (Oostenveld and Praamstra, 2001). The electrooculogram (EOG) was recorded to monitor horizontal eye movements (potential difference between F9 and F10) and blinks (potential difference between Fp1 and an electrode placed below the left eye). The EEG signal was amplified with BrainAmp DC amplifiers (Brain Products GmbH, Gilching, Germany) and digitized using the BrainVision Recorder software (Brain Products GmbH). The analog EEG signal was sampled at 5,000 Hz, filtered on-line with a band pass of 0.1–1,000 Hz and then down-sampled on-line to 500 Hz. Impedances of all electrodes were kept below 10 k $\Omega$ .

The EEG data were preprocessed and analyzed using MATLAB 2016a (MathWorks) and the open source MATLAB toolbox Fieldtrip (Oostenveld et al., 2011)<sup>1</sup>. In all  $n$ -back tasks, the first  $n$  trials of a block were discarded from the analysis, since only the  $n + 1$ st stimulus can be compared to the stimulus presented  $n$  trials ago. Only trials with correct responses were considered. For preprocessing the continuous data were

segmented  $\pm 1.9$  s around stimulus onset. Data epochs for correct targets and non-targets were pooled together. Data were high pass filtered at 1 Hz. A standard automatic routine was applied to exclude data epochs contaminated by eye movements and muscle artifacts. The frequency ranges of the signal time courses that typically contain eye artifacts (1–15 Hz in 2 bipolar EOG channels) and muscle artifacts (110–140 Hz, in 94 data channels) were band-pass filtered and  $z$ -normalized per time point and electrode. The  $z$ -scores were averaged over electrodes, in order to accumulate evidence for artifacts, which typically occur in more than one electrode. Trials exceeding a predefined  $z$ -value (eye artifacts:  $z = 4$ ; muscle artifacts:  $z = 15$ ) were considered as artifacts and excluded. Line-noise was removed by subtracting 50-, 100-, and 150 Hz-Fourier components from the signal time course. Electrodes characterized by high variance across trials (visual inspection) were interpolated (spline interpolation; Perrin et al., 1989; mean number of removed electrodes: 2; range: 0–5). Last, all data channels were re-referenced to a common average reference.

Additionally, trials were excluded according to the following criteria: First, all blocks in which the  $n$ -back level dropped to  $n = 2$  (low load condition) were excluded from the high load condition to avoid identical  $n$ -back levels between load conditions. Second, the adaptively changing  $n$ -back levels allowed to instantaneously (block-wise) adjust the load demands ( $n$ -back level) to individual capacity limits throughout the high load condition. To avoid  $n$ -back levels outside individual capacity limits, which might have occurred during the adjusting process, blocks exceeding individual capacity limits (accuracy rates < 60%) and falling below individual capacity limits (accuracy rates > 90%) were excluded. The average number of included trials per participant was 182 trials (47%) [SD = 42 (11%)] in the 2-back low load condition and 103 trials (25%) [SD = 45 (11%)] in the  $n$ -back high load condition.

## Data Analyses

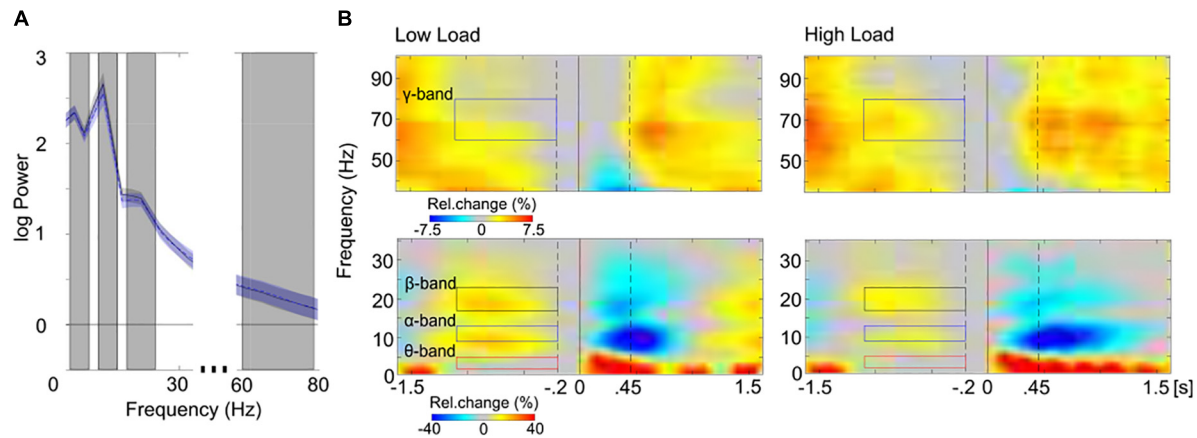
### Estimation of Spectral Power and Source Reconstruction

Discrete Fourier Transforms of EEG maintenance activity ( $-1.1$  to  $-0.2$  s; **Figure 1B**) were calculated at 2.5–100 Hz (segment length: 0.4 s; segment shift: 0.08 s; frequency resolution: 2.5 Hz). The cross-spectrum of the Fourier transformed time segments was retrieved per participant for each electrode and frequency bin in both conditions (low and high load) (Genovese et al., 2002; Nolte et al., 2004). Cross-spectra were averaged across time segments and across trials.

The standard Montreal Neurological Institute (MNI) average brain was used for source reconstruction. Based on the cortical surface of the standard head model a grid was created as a set of 3,000 as equally as possible distributed source points. Every source point represented an equivalent current dipole (cf. Cromer et al., 2010). A standardized three-dimensional map of electrode locations was generated. First, the locations of the 94 employed electrodes were measured three times on a template plastic head, using the ultrasonic Elpos system (zebris Medical GmbH, Isny im Allgäu, Germany). Next, to minimize measurement errors, the

<sup>1</sup><http://www.ru.nl/fcdonders/fieldtrip/>





**FIGURE 2 |** Spectral **(A)** and spectro-temporal **(B)** distribution of maintenance activity. **(A)** The mean overall power spectrum (log power; SEM, shaded area) is displayed separately for the low load (black line) and the high load (blue dashed line) conditions. Power values of the averaged maintenance activity (−1.1 to −0.2 s) were averaged across all voxels and across participants of all training groups. The gray boxes highlight the pre-selected frequency-bands that were used for further analyses (from left to right: theta, 2.5–5 Hz; alpha, 10–12.5 Hz; beta 17.5–22.5 Hz; gamma 60–80 Hz; the spectral resolution of 2.5 Hz is considered). Note, oscillatory peaks are present in all pre-selected frequency bands in the spectral profile of the averaged maintenance activity. **(B)** The sensor-space time-frequency representations of low (left; <36 Hz) and high (right; >36 Hz) frequencies are depicted separately for the low load (top) and the high load (bottom) conditions, averaged across all electrodes and across participants of all training groups. Stimulus-locked (0 s, solid black line) spectral power from −1.6 to 1.6 s (stimulus duration 0.45 s, right dashed black line) is expressed in percent change (Rel. change in%) relative to baseline activity (−0.2, left dashed black line; to 0 s, solid black line). Colored boxes highlight the pre-selected frequency-bands that were used for further analyses (left: red, theta; blue, alpha; black, beta; right: blue, gamma). Note the sustained increase of band-specific activity during the maintenance period (−1.1 to −0.2 s, colored boxes). In further analyses only this maintenance period was analyzed, since the rest of the trial may be overlaid by reaction times (cf. **Figure 1B**).

positions of the three independent measurements were averaged and centered along the midline. These standard electrode locations and the standard grid model were used to compute a standard leadfield matrix. If needed, the standard leadfield was individually adjusted by excluding noisy channels, characterized by high variance across trials, which were interpolated during the individual preprocessing procedure (cf. section “EEG Data Acquisition and Preprocessing”). Exact Low-Resolution Brain Electromagnetic Tomography (eLoreta), a discrete, linear, three-dimensional distributed, weighted minimum norm inverse solution, was used to calculate a spatial filter based on the individually adjusted leadfields. The eLoreta spatial filter localizes the power distribution of the EEG signal with exact maxima for single dipoles (Shafi et al., 2007). The real parts of the cross-spectrum were projected into source space by multiplication with the spatial filter for every source point. Source space power was defined as the maximal eigenvalue of the cross-spectrum, over the three dipole directions. The resulting value per source point represented the power value at that source point (cf. Polomac et al., 2015).

Finally, source power estimates were log-transformed and averaged over frequency ranges of interest, resulting in four frequency bands that were used for further analyses: theta (2.5–5 Hz), alpha (10–12.5 Hz), beta (17.5–22.5 Hz), and gamma (60–80 Hz). These frequency ranges were selected to best represent the core of the theta-, alpha-, beta-, and gamma-bands (cf. Kubicki et al., 1979), given the frequency resolution of 2.5 Hz. The inspection of the overall power spectrum (averaged across voxels) showed that the pre-selected frequency bands in fact captured oscillatory peaks in the

spectral profile of the averaged maintenance activity (**Figure 2A**). Furthermore, for additional confirmation that oscillatory activity in the pre-selected frequency bands represented sustained WM-relevant activity, time-frequency representations (TFRs) of maintenance activity were calculated (overall power across sensors was analyzed in sensor space). The spectral parameters were estimated using multitapers for a trial length of −1.6 to 1.6 s around stimulus onset (**Figure 1B**) with a spectral smoothing of 4.5 Hz for frequencies below 36 Hz and with a spectral smoothing of 8 Hz for frequencies above 36 Hz. For TFRs a sliding time window of 0.4 s was used that was stepped by 50 ms through the trials. TFRs were normalized by 200 ms pre-stimulus activity (baseline). **Figure 2B** shows TFRs of the WM maintenance as percent change relative to baseline. Note, the pre-stimulus baseline, which itself is part of the maintenance period, most probably has reduced WM-relevant activity during the maintenance period. Despite this constraint the inspection of the temporal domain of maintenance activity confirmed that the pre-selected frequency bands represented frequency ranges of the sustained, narrow-band maintenance activity (**Figure 2B**).

### Statistical Analyses

Analyses of behavioral data were conducted using R (R Core Team, 2016) implemented in RStudio (v0.99.486; R Studio Team, 2015)<sup>2</sup>. Behavioral training effects were tested by analyzing the mean WM capacity, represented by the individual *n*-back levels. A one-way analysis of variance (ANOVA) was run on

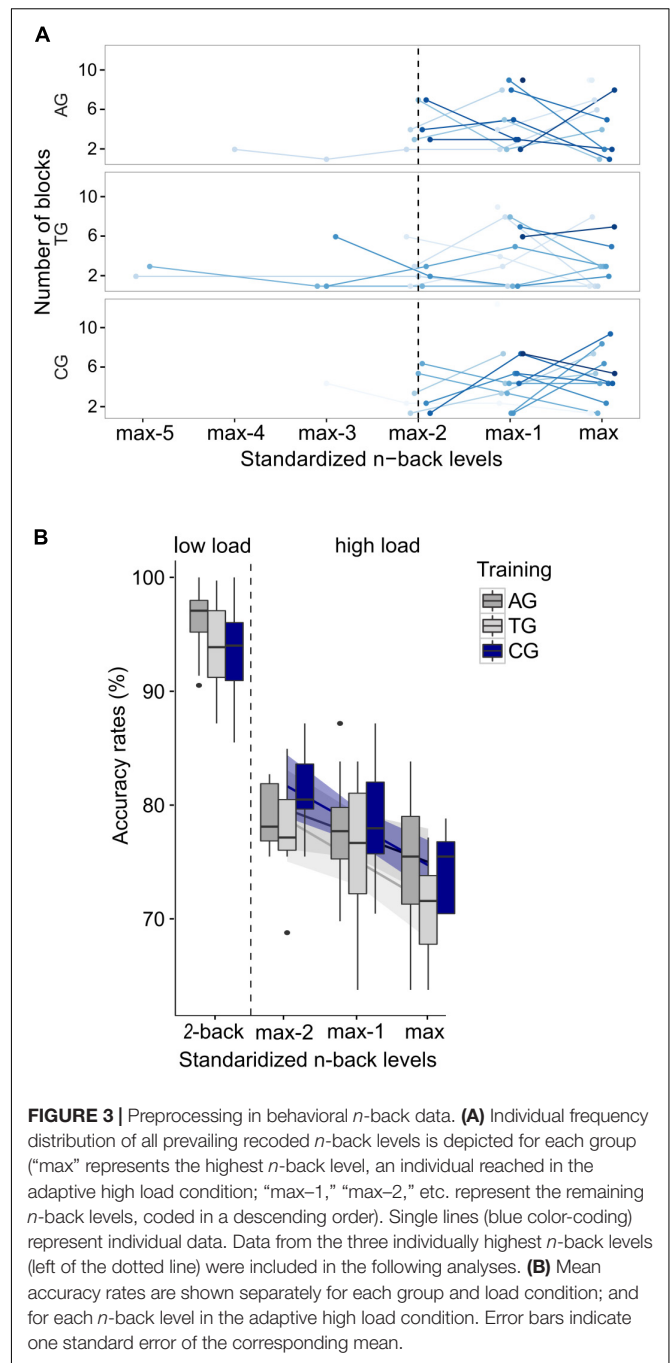
<sup>2</sup><http://www.rstudio.com/>

the mean  $n$ -back levels that were reached during the adaptive high load condition.

Prior to comparing accuracy rates between load conditions, the load demands in the *adaptive* high load condition had to be adjusted to represent individual capacity limits, despite differing  $n$ -back levels between participants (cf. section “Behavioral Results”), differing training experience between participants, and differing WM capacity between participants. To this end, three preprocessing steps were applied. The first two steps were identical to those used for preprocessing EEG data. First, all blocks in which the  $n$ -back level dropped to  $n = 2$  (low load condition) were excluded from the high load condition to avoid identical  $n$ -back levels between load conditions. Second, the adaptively changing  $n$ -back levels allowed to instantaneously (block-wise) adjust the load demands ( $n$ -back level) to individual capacity limits throughout the high load condition. To avoid  $n$ -back levels outside individual capacity limits, which might have occurred during the adjusting process, blocks exceeding individual capacity limits (accuracy rates  $<60\%$ ) and falling below individual capacity limits (accuracy rates  $>90\%$ ) were excluded. The average number of included blocks per participant was 11 blocks (330 trials) out of 14 blocks (420 trials) in total. Finally, performance is by definition negatively correlated with the load factor  $n$ . Thus, in a third preprocessing step, the variance in accuracy, introduced by various  $n$ -back levels within the adaptive high load condition, had to be accounted for. To this end, the individual  $n$ -back levels were recoded for standardization: Each individual's highest  $n$ -back level during the adaptive high load condition represented her/his individual limit in WM capacity. Starting from this personal maximum, the remaining  $n$ -back levels were coded in a descending order (max, max-1, max-2, etc.). The amount of  $n$ -back levels processed during the high load condition varied between 1 and 6 levels (Figure 3A). However, the highest number of different  $n$ -back levels processed by more than one participant in all three groups was 3, i.e.: max, max-1, and max-2 (Figure 3A). All blocks outside these  $n$ -back levels (i.e., max-3, max-4, max-5) were excluded from further analyses<sup>3</sup>.

Having adjusted the load demands within the high load condition a mixed logistic regression model (generalized linear mixed model, GLMM with a logit link function) was run on high load accuracy rates to ensure that the adjusting procedure worked and, thus, similar requirements prevailed within the high load condition across groups and participants despite differing  $n$ -back levels. To this end, the covariate *recoded n-back-levels* (max, max-1, max-2) and the predicting variable *group* (AG, TG, CG) were included as fixed effects. The covariate *recoded n-back-levels* was, furthermore, included into the random effects structure, allowing for random intercepts and slopes for each subject. Significance of fixed effects was tested with Wald  $\chi^2$  tests. The GLMM confirmed that following the adjusting procedure and after having accounted for the covariate ( $\chi^2(1) = 16.01$ ,

<sup>3</sup>For example, during the adaptive  $n$ -back session one particular participant performed several blocks of 4-backs, 5-backs, 6-backs, and 7-backs. This participant's 7-backs were coded as her/his individual maximal level (max), 6-backs as max-1, 5-backs as max-2, and 4-backs as max-3. Blocks with 4-back levels (max-3) were excluded from further analyses.



**FIGURE 3 |** Preprocessing in behavioral  $n$ -back data. **(A)** Individual frequency distribution of all prevailing recoded  $n$ -back levels is depicted for each group (“max” represents the highest  $n$ -back level, an individual reached in the adaptive high load condition; “max-1,” “max-2,” etc. represent the remaining  $n$ -back levels, coded in a descending order). Single lines (blue color-coding) represent individual data. Data from the three individually highest  $n$ -back levels (left of the dotted line) were included in the following analyses. **(B)** Mean accuracy rates are shown separately for each group and load condition; and for each  $n$ -back level in the adaptive high load condition. Error bars indicate one standard error of the corresponding mean.

$p < 0.001$ ) performance within the high load condition no longer differed between groups ( $\chi^2(2) = 4.45$ ,  $p = 0.108$ ; Figure 3B), indicating comparable load demands between groups irrespective of differing  $n$ -back levels. Note that prior to the adjusting procedure performance between groups did differ significantly within the high load condition ( $\chi^2(2) = 6.16$ ,  $p = 0.046$ ; not shown) despite accounting for the covariate ( $\chi^2(1) = 10.16$ ,  $p = 0.001$ ). This indicates the natural differences in load demands due to differing  $n$ -back levels reached in different groups during the adaptive procedure of the task.

The key interest in the present study was to investigate whether WM training alters neural mechanisms when load demands approach individual capacity limits. Particularly, we expected neural mechanisms under high load demands to be altered in response to different types of WM training. Thus, source power of WM maintenance activity was contrasted between the low and the high load condition:  $Pow_{diff} = Pow_{high-load} - Pow_{low-load}$ , resulting in load-related power changes. Importantly, as detailed above, in the adaptive high load condition, load demands had been adjusted prior to the analysis in order to represent individual capacity limits, eliminating the confound of interindividual differences in WM capacity across trained and non-trained participants.

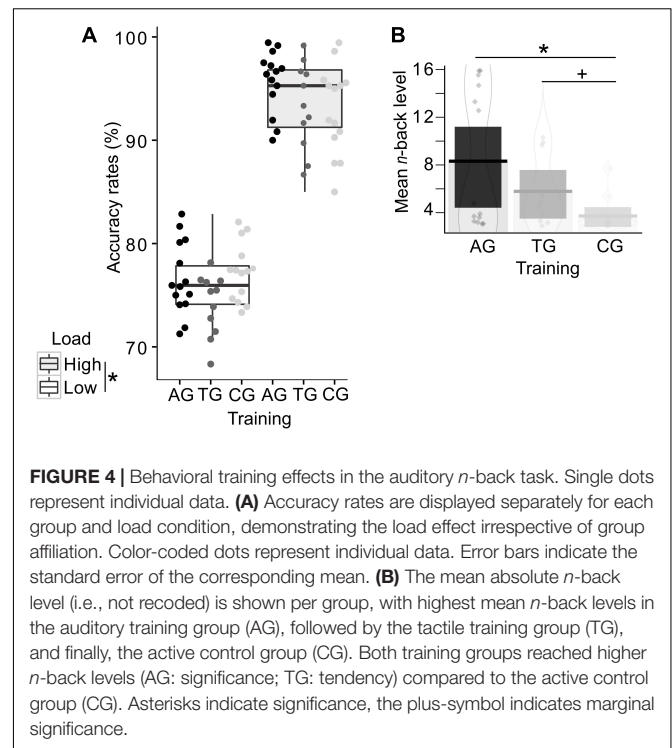
To statistically evaluate differences in WM load effects in the auditory WM task between groups (AG, TG, CG), for each of the four frequency bands, cluster-based permutation tests on load-related power changes were employed (1,000 iterations; Maris and Oostenveld, 2007), with group affiliation being randomly permuted across participants. Specifically, two planned group contrasts were performed: (1) the auditory training group was contrasted with both other groups ( $AG_{Powdiff}$  vs.  $TG_{Powdiff}$  &  $CG_{Powdiff}$ ), to analyze the impact of modality-specific WM training on the load effect; (2) both, i.e., auditory and tactile, training groups were contrasted with the active control group ( $AG_{Powdiff}$  &  $TG_{Powdiff}$  vs.  $CG_{Powdiff}$ ) to analyze the general impact of WM training on the load effect, irrespective of training modality. Spatial clusters were formed of adjacent source points with  $t$ -values below  $p < 0.05$ . The  $t$ -values within clusters were summed. Clusters, which exceeded 95% of the largest summed  $t$ -values from the permutation distribution, were considered statistically significant when  $p < 0.025$  (two-sided  $t$ -tests).

## RESULTS

### Behavioral Performance

In order to test the effects of load on WM performance, accuracy rates were compared with a GLMM, including the predicting variables *WM load* (low and high load) and *group* (AG, TG, CG) as fixed effects. The *WM load* was additionally included into the random effects structure, allowing for random intercepts and slopes for each subject. The GLMM confirmed the expected main effect of load ( $\chi^2(1) = 60.67$ ,  $p < 0.001$ ; **Figure 4A**), reflecting the expected drop in performance in the high load condition across all groups. The predicting variables *group* ( $\chi^2(2) = 4.29$ ,  $p = 0.117$ ) and the interaction between *WM load* and *group* ( $\chi^2(2) = 4.36$ ,  $p = 0.113$ ) were not significant. Note that the absent interaction effect between *WM load* and *group* provides no conclusions about training-related performance changes, as the  $n$ -back levels were adjusted. Instead, in this analysis the absent interaction effect again confirms that the adjusting procedure (cf. section “Data Analyses”) was successful and load demands, as indicated by accuracy, were comparable across groups despite differing  $n$ -back levels in the high load condition.

In order to address training-related performance differences, we analyzed the mean WM capacity, represented by the mean  $n$ -back levels reached during the adaptive high load condition.



**FIGURE 4 |** Behavioral training effects in the auditory  $n$ -back task. Single dots represent individual data. **(A)** Accuracy rates are displayed separately for each group and load condition, demonstrating the load effect irrespective of group affiliation. Color-coded dots represent individual data. Error bars indicate the standard error of the corresponding mean. **(B)** The mean absolute  $n$ -back level (i.e., not recoded) is shown per group, with highest mean  $n$ -back levels in the auditory training group (AG), followed by the tactile training group (TG), and finally, the active control group (CG). Both training groups reached higher  $n$ -back levels (AG: significance; TG: tendency) compared to the active control group (CG). Asterisks indicate significance, the plus-symbol indicates marginal significance.

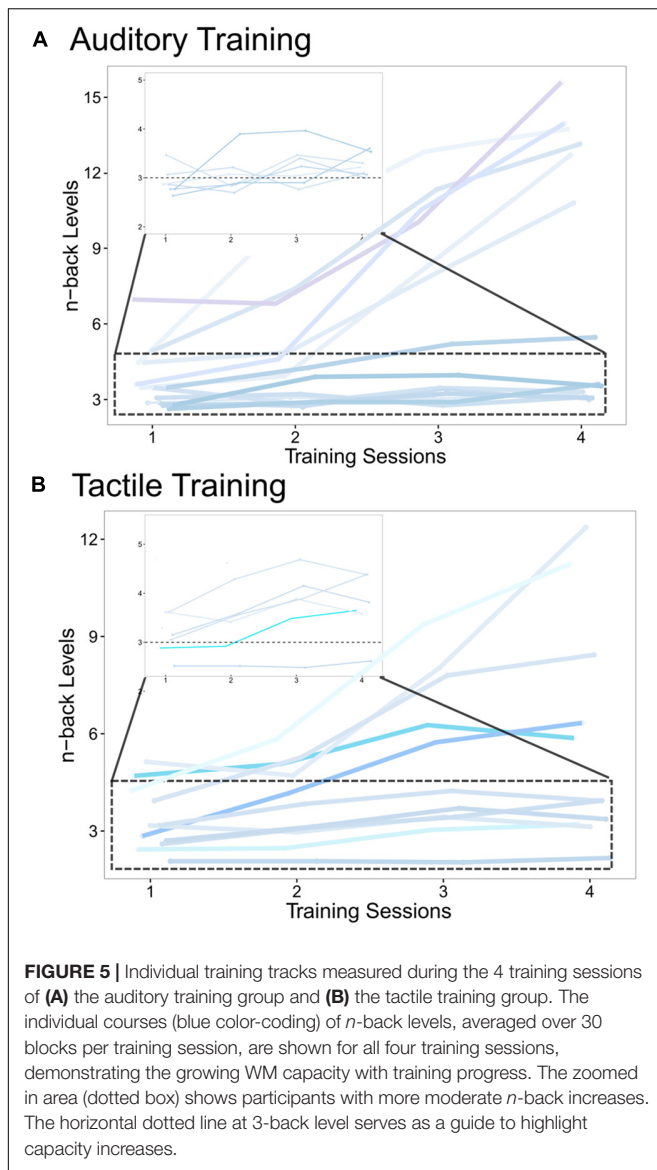
A one-way ANOVA revealed a significant effect of group ( $F(2,36) = 6.21$ ,  $p = 0.005$ ;  $\eta^2_p = 0.204$ ; **Figure 4B**), reflecting a training-related increase in WM capacity. The consecutive Bonferroni-corrected *post hoc* two-sided  $t$ -tests showed the expected performance advantages following modality-specific WM training, as indicated by higher  $n$ -back levels in the auditory training group relative to the control group (AG vs. CG:  $t(26) = 3.23$ ,  $p = 0.010$ ). The tactile training resulted in marginally significant higher  $n$ -back levels relative to the control group (TG vs. CG:  $t(23) = 2.56$ ,  $p = 0.053$ ). AG and TG did not significantly differ in their mean  $n$ -back levels (AG vs. TG:  $t(23) = 1.58$ ,  $p = 0.386$ ).

Additionally, to ensure that no baseline differences existed between groups before training, a one-way ANOVA was run on pre-training 2-back accuracy rates, revealing no significant effect of group ( $F(2,36) = 2.03$ ,  $p = 0.147$ ;  $\eta^2_p = 0.101$ ; data not shown). The individual training tracks (courses of  $n$ -back levels over time across the four training sessions) of AG and TG are shown in **Figure 5**. No training tracks can be shown for CG, since this group performed only the 1-back task with no changes in  $n$ -back levels throughout the four sessions.

### Post-training Load-Related Source Power Changes

We observed a statistical trend for reduced load effects in beta-band power in sensory processing areas in participants with modality-specific WM training relative to both other groups (AG vs. TG&CG; marginally significant cluster,  $p = 0.069$ ; **Figure 6A**). The identified cluster included the right medial temporal lobe, extending along the superior temporal sulcus. Group-specific





*post hoc* analyses (Bonferroni-corrected) confirmed that beta-band power in the identified cluster, pooled across all significant voxels, was significantly increased in the high load condition compared to the low load condition in both the TG and CG ( $t(24) = 3.33$ ,  $p = 0.003$ ; **Figure 6B**; separately for TG and CG: TG:  $t(10) = 2.21$ ,  $p = 0.052$ ; CG:  $t(13) = 2.55$ ,  $p = 0.024$ ), but not in the AG ( $t(13) = -1.58$ ,  $p = 0.137$ ; **Figure 6B**). Furthermore, in the high load condition, beta-band power in the identified cluster was negatively correlated with the maximally reached absolute *n*-back level (Bonferroni-corrected) in the TG and CG ( $r = -0.47$ ,  $p = 0.019$ ; **Figure 6C**), but not in the AG ( $r = -0.13$ ,  $p = 0.652$ ; **Figure 6C**). No further differences in load-related power changes were observed between the AG vs. TG and CG, neither in the theta- (in all identified clusters all  $p > 0.363$ ), nor in the alpha- (in all identified clusters all  $p > 0.462$ ), nor in the gamma-band (in all identified clusters all  $p > 0.155$ ). Finally, no general impact of WM training on the load effect, irrespective of the modality, was

observed when contrasting both training groups with the control group (AG&TG vs. CG; in all identified clusters all  $p > 0.197$ ).

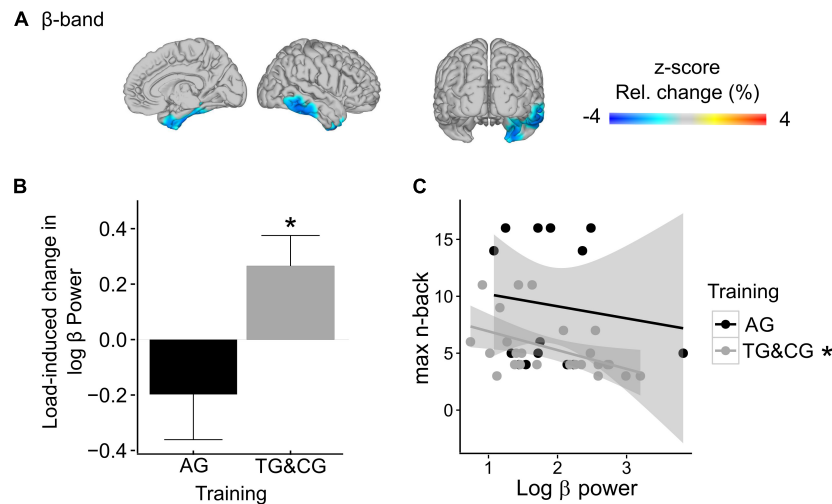
## Load-Related Modulations in Source Power

As reported above, load-related power changes were modulated by WM training only at trend-level. Thus, the main effect of WM load was additionally analyzed across all participants, irrespective of group affiliation. In the theta-band, one cluster with positive *t*-values ( $p = 0.015$ ) and one cluster with negative *t*-values ( $p = 0.002$ ) reached significance (**Figure 7A**). In the high load relative to the low load condition, theta-band power was increased at bilateral frontal poles, spreading to left dorsal and right ventral regions in the frontal lobe. The load-related theta power decrease was localized to the parietal cortex, with strongest desynchronization over bilateral inferior parietal lobules. In the alpha-band power, one significant cluster with negative *t*-values ( $p = 0.002$ ) was found (**Figure 7B**). The load-related alpha-band power was decreased over bilateral centro-parietal regions broadly spreading around pre- and postcentral gyri. In the beta-band power, a significant cluster with negative *t*-values ( $p = 0.010$ ) was observed over bilateral central parts of the cingulate gyri, including bilaterally the precuneus (**Figure 7C**). Finally, in the gamma-band power no significant clusters were observed (in all identified clusters all  $p > 0.284$ ). For full transparency, furthermore, the low-high load contrast is displayed separately for each training group (**Figure 8**).

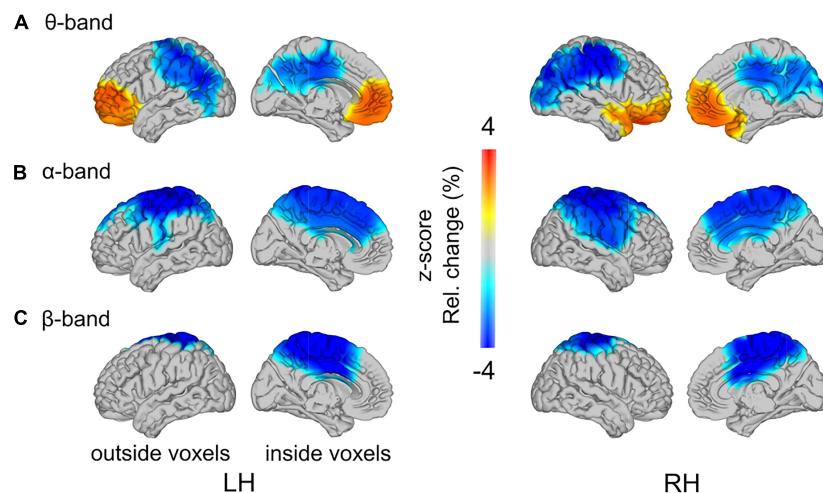
## DISCUSSION

The present study investigated the effects of WM training by comparing neural mechanisms of auditory WM performance in different training groups under high load conditions. The behavioral results confirmed that the WM training was successful compared to the active control group, showing performance increases that suggest an extended WM capacity particularly in the modality-specific (auditory) WM training group. The EEG results revealed two main findings. First, participants who received modality-irrelevant tactile WM training and those who did not receive adaptive WM training (control group), showed load-related increases in beta-band power in the right temporal lobe, when performing the auditory WM task. These changes were not observed in participants who were trained with task-relevant auditory stimuli. This provides evidence that for individuals with higher WM proficiency neural processing was modulated, possibly underlying increases in individual WM capacity. However, note that the main effect of this analysis was only observed at trend level and requires follow-up research for confirmation. Second, we found load-related power changes across all groups in the theta-, alpha-, and beta-band: With increasing load demands, theta-band power increased at bilateral frontal poles, while power decreased in posterior theta-, centro-parietal alpha-, and mid-central beta-band. This suggests that WM processing at high load demands requires both a strengthening of task-relevant processing as well as an attenuation of task-irrelevant processing.





**FIGURE 6 |** Training-related differences in WM load effects (AG vs. TG&CG)\* in beta-band power (17.5–22.5 Hz). **(A)** The observed significant interaction between training and WM load in the right medial temporal lobe is projected onto a glass brain. In interaction-sensitive voxels, color-coding represents z-scores in power change between conditions. Note, light-blue to dark-blue colors indicate reduced load effects in the AG relative to the TG and CG. **(B)** The bar graph shows the load-related power changes separately for AG and TG and CG, pooled across all interaction-sensitive voxels. **(C)** The relationship between beta-band power in the right medial temporal lobe and maximally reached *n*-back levels in the adaptive high-load condition is depicted separately for AG and TG and CG, showing a negative correlation for the tactile training group and the control group, but not for the auditory training group. Asterisks indicate significance. \* AG, auditory WM training group; TG, tactile WM training group; CG, active control group.

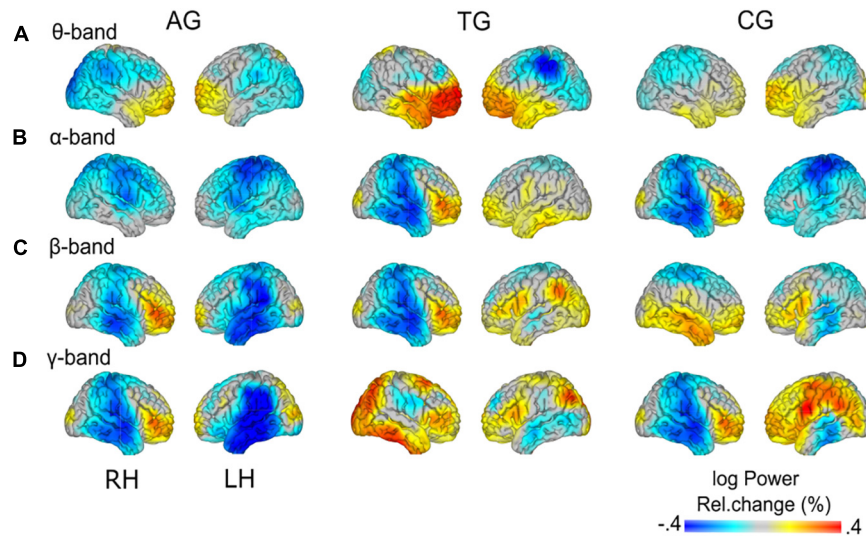


**FIGURE 7 |** Load-related power changes in **(A)** theta- (2.5–5 Hz), **(B)** alpha- (10–12.5 Hz), and **(C)** beta-band (17.5–22.5 Hz), across all training groups, are projected onto a glass brain. Outside and inside voxels are depicted for the left (LH) and right (RH) hemispheres. Color-coding represents z-scores in power change between low load and high load conditions.

## Behavioral Training Effects

As expected, relative to the control group, the increases in auditory WM capacity in the training group, as indicated by the mean *n*-back levels reached during the adaptive post-training task, was significant following modality-specific auditory training, and marginally significant following modality-irrelevant tactile training. We, thus, observed a marginal advantage of tactile training on auditory WM. The observed advantage

of modality-specificity (i.e., a significant vs. a marginally significant WM capacity increase relative to a control group) between training and task is consistent with previous findings. For example, Schneiders et al. (2011) observed a greater training gain in a visual WM task following visual training as compared to auditory training. Typically, WM training effects have been described as narrow, declining with transfer demands (for reviews see Melby-Lervag and Hulme, 2013, 2016;



**FIGURE 8 |** Group-specific load effects. Load-related power changes (log power) are depicted for **(A)** theta- (2.5–5 Hz), **(B)** alpha- (10–12.5 Hz), **(C)** beta- (17.5–22.5 Hz), and **(D)** gamma-band (60–80 Hz), separately for the three training groups (AG, auditory training group; TG, tactile training group; CG, active control group). Note the group-specific differences in load effects (cf. **Figure 7**) vs. similarities (cf. **Figure 8**) in load effects across groups.

Soveri et al., 2017; but see Au et al., 2015, 2016). Nevertheless, our findings indicate that both modality-specific auditory and modality-irrelevant tactile training paradigms improved WM processing, the latter, however, at a marginally significant level.

## Training-Related Differences in WM Load Effects

Our main focus in the present study was on modulations of WM load effects, comparing groups with different post-training WM performance. Individuals who received WM training with task-irrelevant tactile stimuli (TG) and non-adaptive training (CG), relative to those who received WM training with task-specific auditory stimuli (AG), showed load-induced beta-band increases in the right medial temporal lobe, extending along the superior temporal sulcus. This increase in beta-band power, furthermore, correlated negatively with the absolute *n*-back levels that were maximally reached under the high load condition.

These results suggest that participants who did not receive modality-specific WM training (TG and CG) required additional activation of right medial temporal regions as load demands increased. The medial temporal lobe (including hippocampal structures together with anatomically adjacent cortical regions) has been reliably associated with both encoding and retrieval of information (for a review see Simons and Spiers, 2003), being particularly critical for rapid learning of new episodic information (Patterson et al., 2007). Furthermore, the observed activation in the right medial temporal lobe extended along the superior temporal sulcus, an area known to be voice-selective, with an emphasis on the right hemisphere (Belin et al., 2000, 2002). The task-relevant features used in the auditory verbal *n*-back task of the present study were voices, which participants had to match regarding the speakers' identities. Thus, the increased activation in these memory- and voice-relevant

areas seems to reflect additional support of WM processing under high load demands. These considerations are in line with a study by Leiberg et al. (2006). In an auditory Sternberg task with syllables spoken by a natural voice, the authors found a load-induced increase in beta-band power over right temporal MEG sensors during the maintenance period. Leiberg et al. (2006) related the load-induced parametric beta-band enhancements to the maintenance of an increased stimulus set, suggesting that beta-band power codes for the representations of task-relevant stimulus features. Given that we adjusted the task-difficulty to a comparable level across individuals and groups (cf. section "Data Analyses"), these results confirm our hypothesis. They show that, despite all individuals performing at their capacity limits (adjusted load-levels), participants of the tactile training group and the control group recruited additional brain areas for successful WM processing at high load demands compared to participants with modality-specific training. Effects might result from more efficient maintenance of task-relevant stimulus representations, following modality-specific auditory training. Probably a higher experience with task-relevant auditory stimuli facilitated perceptual processing, supporting maintenance processes. The fact that the beta-band activity may be related to processing efficiency is further reflected in the negative correlation between the absolute *n*-back level and beta-band power in the right temporal lobe in participants of the tactile training group and the active control group. That is, participants with highest gains in performance, as indicated by higher *n*-back levels reached during the adaptive high load condition, showed the least load-induced beta-band increases, and thus, the least need for additional activation in the right medial temporal lobe. Possibly, a training-induced efficiency in maintenance processes results in a more efficient resource allocation from perceptual to WM-related processing.

These group-specific differences in beta-band power were right-lateralized, thus being in accordance with literature on voice-selective regions (Belin et al., 2000, 2002), and more general, in accordance with literature on hemispheric lateralization of the auditory cortex. There is a higher selectivity of the right auditory cortex for the processing of slow spectral aspects of auditory input, such as speech prosody or pitch variations, over fast temporal aspects, along with a complementary specialization of the left auditory cortex (for review: Poeppel, 2001; Zatorre et al., 2002; Assaneo et al., 2019). The right lateralization of the observed effect corresponds to this functional differentiation, as in our study voices, i.e., sounds' spectral resolution, had to be maintained and discriminated.

This finding, however, has been observed only at statistical trend level, and has to be interpreted with caution. Nevertheless, given this marginally significant result finds support in further research, these data would provide neurophysiological evidence for the previously reported "narrowness" of behavioral training effects (Melby-Lervåg and Hulme, 2013, 2016). Narrow training effects are indicated by attenuated training gains with transfer to other sensory modalities (Schneiders et al., 2011) or similar tasks (Dahlin et al., 2008). Our results demonstrate that the narrowness of training effects may be related to a training-induced increase in efficiency to maintain representations of task-relevant features, particularly following modality-specific WM training. Importantly, as reflected in group-specific WM load effects during the WM *maintenance period*, the here reported training-related benefits go beyond previously reported mere encoding advantages (e.g., Lustig and Flegal, 2008).

## WM Load Effects

Working memory load effects across all participants and groups were observed in theta-, alpha-, and beta-band power, but not in gamma-band power.

In line with previous studies (Krause et al., 2000; Jensen and Tesche, 2002; Deiber et al., 2007; Notebaert and Verguts, 2008), we found that theta-band oscillatory activity increased with WM load over bilateral prefrontal regions. The load-related increases in frontal theta-band power have been functionally related to enhanced requirements of cognitive control and executive functioning at higher task demands (Sauseng et al., 2010; Hellbernd and Sammler, 2016). The distribution of voxels with significant effects, comprising the frontal theta-band increase, spread from bilateral mid frontal poles in the left hemisphere to *dorsal* regions, and in the right hemisphere to *ventral* regions. Neuroimaging studies, aimed at classifying activation patterns among WM-related regions, have linked dorsal frontal activation to executive processes, necessary for a continuous updating and maintenance of the sequential order of items in WM (D'Esposito et al., 1998; Owen, 2000; Lisman, 2010; Heusser et al., 2016). Right-lateralized ventral frontal activation was associated with the manipulation of the stored information (Wager and Smith, 2003), particularly including selection and evaluation of WM items (Owen et al., 2005). It is reasonable to assume that under high load conditions both types of these executive functions – continuous updating and maintenance as well as manipulation – are required to a larger extent during WM performance. Thus, the

observed asymmetric spreading of theta-band increases toward left dorsal and right ventral regions possibly reflects distinct sub-functions involved in WM processing.

Additionally, we observed load-induced theta-band decreases in the parietal cortex. Theta-band decreases in regions other than frontal cortex have been previously reported (Howard et al., 2003; Meltzer et al., 2008; van Vugt et al., 2010; but see, Raghavachari et al., 2006; Sauseng et al., 2010). The strongest theta-band desynchronizations were observed over bilateral inferior parietal lobules. Bilateral supramarginal gyri, located in the inferior parietal lobules, have been shown to contribute to phonological decisions (McDermott et al., 2003; Shalom and Poeppel, 2008; Hartwigsen et al., 2010). The stimuli in the auditory *n*-back task implemented in the present study consisted of a pseudo-word spoken by different individuals. That is, the speakers' voices, thus, the sounds (i.e., phonology) of pseudo-words had to be maintained and discriminated. Hence, theta-band decreases over bilateral inferior parietal lobules are probably related to more efficient processing of voice stimuli. These considerations are in line with an fMRI study by Schon et al. (2008), who observed gradual load-induced BOLD-signal decreases in left lateral temporal regions along with prefrontal activation increases in a dual *n*-back task, in which visuospatial and auditory information had to be maintained simultaneously. The authors related this anterior-posterior shift in BOLD activity to a load-related shift from perceptually dominated processing to memory processing. Thus, similarly as Jaeggi et al. (2007), we speculate that the observed load-induced desynchronizations over inferior parietal regions may indicate a shift from phonological processing to enhanced cognitive control, as indicated by load-induced frontal theta-band increases.

In the alpha-band, oscillatory power decreased with WM load over bilateral centro-parietal regions. These load-induced alpha-band desynchronizations are in good agreement with previous studies implementing the *n*-back paradigm (Gevins et al., 1997; Deiber et al., 2007; Schuster et al., 2018). Together with previous evidence that parietal regions, in addition to frontal regions, constitute the core WM network (for a review see Eriksson et al., 2015), our data suggest that the observed desynchronizations in bilateral centro-parietal alpha-band power reflect overall enhanced engagement with high processing demands (cf. Klimesch et al., 2007).

Interestingly, while our data support the interpretation of enhanced involvement of both frontal and centro-parietal regions with increasing load demands, the underlying neural correlates differ in their spectral profiles, as indicated by load-induced frontal *theta-band increases* versus load-induced centro-parietal *alpha-band decreases*. However, the relatively low frequency resolution of 2.5 Hz applied in our analyses (cf. section "Data Analyses") may constrain our findings particularly in the low frequency range.

In the beta-band, we found WM load-induced power decreases over bilateral mid-central regions. Although in contrast to some studies, reporting load-related increases of WM-relevant oscillatory activity in *n*-back tasks (Deiber et al., 2007; Larrouy-Maestri et al., 2013), our findings are consistent with the frequent reports of load-related beta-band desynchronizations,

mainly at medial regions (along the anterior-posterior-axis) including the cingulate cortex (Pesonen et al., 2007; Brookes et al., 2011; Palomaki et al., 2012; Takei et al., 2016; Scharinger et al., 2017). Interestingly, Brookes et al. (2011) stressed the largely overlapping spatial distribution of load-induced beta-band desynchronizations in WM processing with regions comprising the default mode network (DMN). The DMN commonly involves the medial prefrontal cortex, bilateral inferior parietal lobules, and, particularly, medially located posterior and anterior cingulate cortices (Shulman et al., 1997; Raichle et al., 2001). This anatomically confined brain system has been characterized by spontaneous activity during cognitive disengagement from the external world, and complementary, by task-induced deactivations (Gusnard and Raichle, 2001; Buckner et al., 2008). Of particular relevance for the present results is that simultaneous EEG and fMRI recordings have linked the DMN, typically studied with neuroimaging methods, mainly to *beta-band* oscillatory activity (Laufs et al., 2003; Mantini et al., 2007; Brookes et al., 2011). Accordingly, we assume that the observed load-induced decreases in beta-band activity over medial centroparietal regions might reflect a task-related attenuation of default mode functioning. Similarly, in fMRI studies, activity in the DMN has been observed to decrease with increasing WM load (Esposito et al., 2006; Woodward et al., 2006; Park et al., 2011; Piccoli et al., 2015).

Apparently, when it comes to highly demanding cognitive tasks, complementary mechanisms in terms of suppression of task-irrelevant processing and strengthening of task-related processing seem to be relevant for successful performance. As for alpha-band synchronization (Klimesch et al., 2007), we hypothesize that beta-band desynchronizations are involved in inhibitory control. More specifically, alpha-band synchronizations have been linked to *exogenously* driven top-down inhibition to prevent interference from task-irrelevant sensory input or task-irrelevant sensory/external processing (Klimesch et al., 2007). For beta-band oscillations, we speculate that desynchronizations over regions defining the DMN reflect an *endogenously* driven top-down inhibitory system to prevent interference from task-irrelevant internal processes. This interpretation is consistent with the recently proposed hypothesis on beta-band functioning by Engel and Fries (2010), stressing beta's relevance for maintaining ongoing motor activity and cognitive sets, particularly when the current state is prioritized over the processing of new signals, possibly by inhibiting new sensory input.

In this study, all participants were blindfolded, as it was part of a larger study investigating WM in congenital blindness (cf. section "Experimental Procedures"). Blindfolding changes the overall dynamic pattern of oscillatory brain activity (Berger, 1929; Adrian and Matthews, 1934; Geller et al., 2014). However, these general effects are constants; the reported load effects, thus, cannot be explained by blindfolding. Whether the overall topography of the observed WM effects at high load demands was modulated by blindfolding cannot be finally excluded but the parieto-occipital topography of attentional effects on alpha oscillations, as shown in a previous study (Wostmann et al., 2020), argues against this possibility.

## CONCLUSION

Taken together, our findings on WM maintenance-related activity in the EEG highlight that neural mechanisms that are recruited when individuals perform at their capacity limits were tendency altered after modality-specific WM training. This interesting finding, however, requires further research for confirmation. Additionally, replicating and extending previous WM load effects, we showed that successful WM maintenance at highly demanding load levels requires both a strengthening of task-relevant processing as well as an attenuation of task-irrelevant processing, characterized by specific electrophysiological signatures.

## DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available because no consent for data sharing in repositories was given. Requests to access the datasets should be directed to HG-M [helene.gudi-mindermann@posteo.de](mailto:helene.gudi-mindermann@posteo.de).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the the German Psychological Association (Deutsche Gesellschaft für Psychologie, DGPs). The participants provided their written informed consent to participate in the study.

## AUTHOR CONTRIBUTIONS

All authors conceptualized the study and edited the manuscript. HG-M and JR wrote the manuscript. HG-M analyzed the data.

## FUNDING

This work was supported by the German Research Foundation (DFG: SFB 936/B2/A3 to BR and AE, SFB TRR 169/A1/B1 to BR and AE, and BR 4913/2-1 to PB), a grant "Crossmodal learning" from the Landesforschungsförderung Hamburg (FV25 to BR and AE), and the Max-Planck-Institute for Empirical Aesthetics (JR).

## ACKNOWLEDGMENTS

We thank Davide Bottari, Jonathan Schubert, and Tobias Heed for helpful suggestions and discussion during the progress of the study, and Guido Nolte for his support and helpful comments on methodological issues. We thank Krutika Gohil, Özlem-Feray Kayali, Lilly Kramer, Hatice Oe, Sabine Öhlschläger, Marc Rommel, and Hanna Stoffregen for their support conducting the training study and acquiring the EEG data, and Nicola Kaczmarek for her support in recruiting and scheduling participants.



## REFERENCES

- Adrian, E. D., and Matthews, B. H. (1934). The interpretation of potential waves in the cortex. *J. Physiol.* 81, 440–471. doi: 10.1113/jphysiol.1934.sp003147
- Ashby, F. G., Isen, A. M., and Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychol. Rev.* 106, 529–550. doi: 10.1037/0033-295x.106.3.529
- Assaneo, M. F., Rimmele, J. M., Orpella, J., Ripolles, P., de Diego-Balaguer, R., and Poeppel, D. (2019). The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. *Front. Integr. Neurosci.* 13:28. doi: 10.3389/fnint.2019.00028
- Astle, D. E., Barnes, J. J., Baker, K., Colclough, G. L., and Woolrich, M. W. (2015). Cognitive training enhances intrinsic brain connectivity in childhood. *J. Neurosci.* 35, 6277–6283. doi: 10.1523/JNEUROSCI.4517-14.2015
- Au, J., Buschkuhl, M., Duncan, G. J., and Jaeggi, S. M. (2016). There is no convincing evidence that working memory training is NOT effective: a reply to Melby-Lervag and Hulme (2015). *Psychon. Bull. Rev.* 23, 331–337. doi: 10.3758/s13423-015-0967-4
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., and Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon. Bull. Rev.* 22, 366–377. doi: 10.3758/s13423-014-0699-x
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201
- Barnes, J. J., Nobre, A. C., Woolrich, M. W., Baker, K., and Astle, D. E. (2016). Training working memory in childhood enhances coupling between frontoparietal control network and task-related regions. *J. Neurosci.* 36, 9001–9011. doi: 10.1523/JNEUROSCI.0101-16.2016
- Barnett, K. J., Finucane, C., Asher, J. E., Bargary, G., Corvin, A. P., Newell, F. N., et al. (2008). Familial patterns and the origins of individual differences in synaesthesia. *Cognition* 106, 871–893. doi: 10.1016/j.cognition.2007.05.003
- Belin, P., Zatorre, R. J., and Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res. Cogn. Brain Res.* 13, 17–26. doi: 10.1016/s0926-6410(01)00084-2
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078
- Berger, H. (1929). Über das elektroencephalogramm des menschen. *Arch. Psychiatr. Nervenkrankheiten* 87, 527–570.
- Brookes, M. J., Wood, J. R., Stevenson, C. M., Zumer, J. M., White, T. P., Liddle, P. F., et al. (2011). Changes in brain network activity during working memory tasks: a magnetoencephalography study. *Neuroimage* 55, 1804–1815. doi: 10.1016/j.neuroimage.2010.10.074
- Buchsbaum, B. R., and D'Esposito, M. (2019). A sensorimotor view of verbal working memory. *Cortex* 112, 134–148. doi: 10.1016/j.cortex.2018.11.010
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Buschkuhl, M., Jaeggi, S. M., and Jonides, J. (2012). Neuronal effects following working memory training. *Dev. Cognit. Neurosci.* 2(Suppl. 1), S167–S179. doi: 10.1016/j.dcn.2011.10.001
- Buzsaki, G., and Wang, X. J. (2012). Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* 35, 203–225. doi: 10.1146/annurev-neuro-062111-150444
- Chen, Y., and Huang, X. (2015). Modulation of alpha and beta oscillations during an n-back task with varying temporal memory load. *Front. Psychol.* 6:2031. doi: 10.3389/fpsyg.2015.02031
- Cromer, J. A., Roy, J. E., and Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66, 796–807. doi: 10.1016/j.neuron.2010.05.005
- Dahlin, E., Neely, A. S., Larsson, A., Backman, L., and Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510–1512. doi: 10.1126/science.1155466
- Dalbert, C. (1992). Subjektives wohlfinden junger erwachsener: theoretische und empirische analysen der struktur und stabilität. *Z. Differ. Diagnost. Psychol.* 13, 207–220.
- de Diego-Balaguer, R., Rodriguez-Fornells, A., and Bachoud-Levi, A. C. (2015). Prosodic cues enhance rule learning by changing speech segmentation mechanisms. *Front. Psychol.* 6:1478. doi: 10.3389/fpsyg.2015.01478
- Deiber, M. P., Missonnier, P., Bertrand, O., Gold, G., Fazio-Costa, L., Ibanez, V., et al. (2007). Distinction between perceptual and attentional processing in working memory tasks: a study of phase-locked and induced oscillatory brain dynamics. *J. Cogn. Neurosci.* 19, 158–172. doi: 10.1162/jocn.2007.19.1.158
- D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 761–772. doi: 10.1098/rstb.2007.2086
- D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., and Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Brain Res. Cogn. Brain Res.* 7, 1–13. doi: 10.1016/s0926-6410(98)00004-4
- D'Esposito, M., and Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142. doi: 10.1146/annurev-psych-010814-015031
- Donner, T. H., and Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends Cogn. Sci.* 15, 191–199. doi: 10.1016/j.tics.2011.03.007
- Dupoux, E., Sebastian-Galles, N., Navarrete, E., and Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition* 106, 682–706. doi: 10.1016/j.cognition.2007.04.001
- Emmorey, K., Giezen, M. R., Petrich, J. A. F., Spurgeon, E., and O'Grady Farnady, L. (2017). The relation between working memory and language comprehension in signers and speakers. *Acta Psychol. (Amst.)* 177, 69–77. doi: 10.1016/j.actpsy.2017.04.014
- Engel, A. K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015
- Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25. doi: 10.1016/s1364-6613(00)01568-0
- Eriksson, J., Vogel, E. K., Lansner, A., Bergstrom, F., and Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron* 88, 33–46. doi: 10.1016/j.neuron.2015.09.020
- Esposito, F., Bertolino, A., Scarabino, T., Latorre, V., Blasi, G., Popolizio, T., et al. (2006). Independent component model of the default-mode brain function: assessing the impact of active thinking. *Brain Res. Bull.* 70, 263–269. doi: 10.1016/j.brainresbull.2006.06.012
- Fliege, H., Rose, M., Arck, P., Levenstein, S., and Klapp, B. F. (2001). Validierung des "perceived stress questionnaire" (PSQ) an einer deutschen Stichprobe. *Diagnostica* 47, 142–152. doi: 10.1026/0012-1924.47.3.142
- Geller, A. S., Burke, J. F., Sperling, M. R., Sharan, A. D., Litt, B., Baltuch, G. H., et al. (2014). Eye closure causes widespread low-frequency power increase and focal gamma attenuation in the human electrocorticogram. *Clin. Neurophysiol.* 125, 1764–1773. doi: 10.1016/j.clinph.2014.01.021
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037
- Gevens, A., Smith, M. E., McEvoy, L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* 7, 374–385. doi: 10.1093/cercor/7.4.374
- Gudi-Mindermann, H., Rimmele, J. M., Nolte, G., Bruns, P., Engel, A. K., and Roder, B. (2018). Working memory training in congenitally blind individuals results in an integration of occipital cortex in functional networks. *Behav. Brain Res.* 348, 31–41. doi: 10.1016/j.bbr.2018.04.002
- Gusnard, D. A., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694. doi: 10.1038/35094500
- Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., and Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *Neuroimage* 92, 46–55. doi: 10.1016/j.neuroimage.2014.01.049
- Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., and Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16494–16499. doi: 10.1073/pnas.1008121107
- Hellbernd, N., and Sammler, D. (2016). Prosody conveys speaker's intentions: acoustic cues for speech act perception. *J. Mem. Lang.* 88, 70–86. doi: 10.1016/j.jml.2016.01.001
- Heusser, A. C., Poeppel, D., Ezyy, Y., and Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. *Nat. Neurosci.* 19, 1374–1380. doi: 10.1038/nn.4374
- Howard, M. W., Rizzuto, D. S., Caplan, J. B., Madsen, J. R., Lisman, J., Aschenbrenner-Scheibe, R., et al. (2003). Gamma oscillations correlate with

- working memory load in humans. *Cereb. Cortex* 13, 1369–1374. doi: 10.1093/cercor/bhg084
- Jaeggi, S. M., Buschkuhl, M., Etienne, A., Ozdoba, C., Perrig, W. J., and Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cogn. Affect. Behav. Neurosci.* 7, 75–89. doi: 10.3758/cabn.7.2.75
- Jausovec, N., and Jausovec, K. (2012). Working memory training: improving intelligence—changing brain activity. *Brain Cognit.* 79, 96–106. doi: 10.1016/j.bandc.2012.02.007
- Jensen, O., and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Eur. J. Neurosci.* 15, 1395–1399. doi: 10.1046/j.1460-9568.2002.01975.x
- Jokisch, D., and Jensen, O. (2007). Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *J. Neurosci.* 27, 3244–3251. doi: 10.1523/JNEUROSCI.5399-06.2007
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., and Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59, 193–224. doi: 10.1146/annurev.psych.59.103006.093615
- Kaiser, J., Lutzenberger, W., Decker, C., Wibrall, M., and Rahm, B. (2009). Task- and performance-related modulation of domain-specific auditory short-term memory representations in the gamma-band. *Neuroimage* 46, 1127–1136. doi: 10.1016/j.neuroimage.2009.03.011
- Kaiser, J., Ripper, B., Birbaumer, N., and Lutzenberger, W. (2003). Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *Neuroimage* 20, 816–827. doi: 10.1016/S1053-8119(03)00350-1
- Kapeller, C., Ogawa, H., Schalk, G., Kunii, N., Coon, W. G., Scharinger, J., et al. (2018). Real-time detection and discrimination of visual perception using electrocorticographic signals. *J. Neural Eng.* 15:036001. doi: 10.1088/1741-2552/aaa9f6
- Klimesch, W. (2012). alpha-band oscillations, attention, and controlled access to stored information. *Trends Cognit. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003
- Kopell, N., Whittington, M. A., and Kramer, M. A. (2011). Neuronal assembly dynamics in the beta frequency range permits short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3779–3784. doi: 10.1073/pnas.1019676108
- Krause, C. M., Sillanmaki, L., Koivisto, M., Saarela, C., Haggqvist, A., Laine, M., et al. (2000). The effects of memory load on event-related EEG desynchronization and synchronization. *Clin. Neurophysiol.* 111, 2071–2078. doi: 10.1016/s1388-2457(00)00429-6
- Kubicki, S., Herrmann, W. M., Fichte, K., and Freund, G. (1979). Reflections on the topics: EEG frequency bands and regulation of vigilance. *Pharmakopsychiatr. Neuropsychopharmakol.* 12, 237–245. doi: 10.1055/s-0028-1094615
- Langer, N., von Bastian, C. C., Wirz, H., Oberauer, K., and Jancke, L. (2013). The effects of working memory training on functional brain network efficiency. *Cortex* 49, 2424–2438. doi: 10.1016/j.cortex.2013.01.008
- Larrouy-Maestri, P., Leybaert, J., and Kolinsky, R. (2013). The benefit of musical and linguistic expertise on language acquisition in sung material. *Musicae Sci.* 17, 217–228. doi: 10.1177/1029864912473470
- Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., et al. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11053–11058. doi: 10.1073/pnas.1831638100
- Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. Balingen: Spitta-Verlag.
- Leiberg, S., Lutzenberger, W., and Kaiser, J. (2006). Effects of memory load on cortical oscillatory activity during auditory pattern working memory. *Brain Res.* 1120, 131–140. doi: 10.1016/j.brainres.2006.08.066
- Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., et al. (1993). Development of the perceived stress questionnaire: a new tool for psychosomatic research. *J. Psychosom. Res.* 37, 19–32. doi: 10.1016/0022-3999(93)90120-5
- Lisman, J. (2010). Working memory: the importance of theta and gamma oscillations. *Curr. Biol.* 20, R490–R492. doi: 10.1016/j.cub.2010.04.011
- Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cognit. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006
- Luethi, M., Meier, B., and Sandi, C. (2008). Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. *Front. Behav. Neurosci.* 2:5. doi: 10.3389/neuro.08.005.2008
- Lustig, C., and Flegal, K. E. (2008). Targeting latent function: encouraging effective encoding for successful memory training and transfer. *Psychol. Aging* 23, 754–764. doi: 10.1037/a0014295
- Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., and Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13170–13175. doi: 10.1073/pnas.0700668104
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- McDermott, K. B., Petersen, S. E., Watson, J. M., and Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41, 293–303. doi: 10.1016/s0028-3932(02)00162-8
- Medendorp, W. P., Kramer, G. F., Jensen, O., Oostenveld, R., Schoffelen, J. M., and Fries, P. (2007). Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cereb. Cortex* 17, 2364–2374. doi: 10.1093/cercor/bhl145
- Melby-Lervag, M., and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228
- Melby-Lervag, M., and Hulme, C. (2016). There is no convincing evidence that working memory training is effective: a reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychon. Bull. Rev.* 23, 324–330. doi: 10.3758/s13423-015-0862-z
- Meltzer, J. A., Zaveri, H. P., Goncharova, I. I., Distasio, M. M., Papademetris, X., Spencer, S. S., et al. (2008). Effects of working memory load on oscillatory power in human intracranial EEG. *Cereb. Cortex* 18, 1843–1855. doi: 2008
- Montgomery, J. W. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. *J. Speech Lang. Hear. Res.* 43, 293–308. doi: 10.1044/jslhr.4302.293
- Morrison, A. B., and Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychon. Bull. Rev.* 18, 46–60. doi: 10.3758/s13423-010-0034-0
- Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307. doi: 10.1016/j.clinph.2004.04.029
- Notebaert, W., and Verguts, T. (2008). Cognitive control acts locally. *Cognition* 106, 1071–1080. doi: 10.1016/j.cognition.2007.04.011
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/s1388-2457(00)00527-7
- Owen, A. M. (2000). The role of the lateral frontal cortex in mnemonic processing: the contribution of functional neuroimaging. *Exp. Brain Res.* 133, 33–43. doi: 10.1007/s002210000398
- Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131
- Palomaki, J., Kivikangas, M., Alafuzoff, A., Hakala, T., and Krause, C. M. (2012). Brain oscillatory 4–35 Hz EEG responses during an n-back task with complex visual stimuli. *Neurosci. Lett.* 516, 141–145. doi: 10.1016/j.neulet.2012.03.076
- Palva, J. M., Monto, S., Kulashekhar, S., and Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7580–7585. doi: 10.1073/pnas.0913113107
- Palva, S., Kulashekhar, S., Hamalainen, M., and Palva, J. M. (2011). Localization of cortical phase and amplitude dynamics during visual working memory encoding and retention. *J. Neurosci.* 31, 5013–5025. doi: 10.1523/JNEUROSCI.5592-10.2011
- Park, H. J., Chun, J. W., Park, B., Park, H., Kim, J. I., Lee, J. D., et al. (2011). Activation of the occipital cortex and deactivation of the default mode network

- during working memory in the early blind. *J. Int. Neuropsychol. Soc.* 17, 407–422. doi: 10.1017/S1355617711000051
- 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
- Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187. doi: 10.1016/0013-4694(89)90180-6
- Pesonen, M., Hamalainen, H., and Krause, C. M. (2007). Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. *Brain Res.* 1138, 171–177. doi: 10.1016/j.brainres.2006.12.076
- Piccoli, T., Valente, G., Linden, D. E., Re, M., Esposito, F., Sack, A. T., et al. (2015). The default mode network and the working memory network are not anti-correlated during all phases of a working memory task. *PLoS ONE* 10:e0123354. doi: 10.1371/journal.pone.0123354
- Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognit. Sci.* 25, 679–693. doi: 10.1207/s15516709cog2505\_3
- Polomac, N., Leicht, G., Nolte, G., Andreou, C., Schneider, T. R., Steinmann, S., et al. (2015). Generators and connectivity of the early auditory evoked gamma band response. *Brain Topogr.* 28, 865–878. doi: 10.1007/s10548-015-0434-6
- R Core Team, (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/> (accessed September 01, 2017).
- Raghavachari, S., Lisman, J. E., Tully, M., Madsen, J. R., Bromfield, E. B., and Kahana, M. J. (2006). Theta oscillations in human cortex during a working-memory task: evidence for local generators. *J. Neurophysiol.* 95, 1630–1638. doi: 10.1152/jn.00409.2005
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Rimmele, J. M., Gudi-Mindermann, H., Nolte, G., Roder, B., and Engel, A. K. (2019). Working memory training integrates visual cortex into beta-band networks in congenitally blind individuals. *Neuroimage* 194, 259–271. doi: 10.1016/j.neuroimage.2019.03.003
- Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cognit. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010
- Roux, F., Wibral, M., Mohr, H. M., Singer, W., and Uhlhaas, P. J. (2012). Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *J. Neurosci.* 32, 12411–12420. doi: 10.1523/JNEUROSCI.0421-12.2012
- R Studio Team (2015). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. Available online at: <http://www.rstudio.com/> (accessed September 01, 2017).
- Salazar, R. F., Dotson, N. M., Bressler, S. L., and Gray, C. M. (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science* 338, 1097–1100. doi: 10.1126/science.1224000
- Sauseng, P., Griesmayr, B., Freunberger, R., and Klimesch, W. (2010). Control mechanisms in working memory: a possible function of EEG theta oscillations. *Neurosci. Biobehav. Rev.* 34, 1015–1022. doi: 10.1016/j.neubiorev.2009.12.006
- Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., et al. (2009). Brain oscillatory substrates of visual short-term memory capacity. *Curr. Biol.* 19, 1846–1852. doi: 10.1016/j.cub.2009.08.062
- Scharinger, C., Soutschek, A., Schubert, T., and Gerjets, P. (2017). Comparison of the working memory load in n-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. *Front. Hum. Neurosci.* 11:6. doi: 10.3389/fnhum.2017.00006
- Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., and Bastiaansen, M. C. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage* 44, 1224–1238. doi: 10.1016/j.neuroimage.2008.08.041
- Schneiders, J. A., Opitz, B., Krick, C. M., and Mecklinger, A. (2011). Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cereb. Cortex* 21, 2555–2564. doi: 10.1093/cercor/bhr037
- Schon, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition* 106, 975–983. doi: 10.1016/j.cognition.2007.03.005
- Schuster, S., Scharinger, M., Brooks, C., Lahiri, A., and Hartwigsen, G. (2018). The neural correlates of morphological complexity processing: detecting structure in pseudowords. *Hum. Brain Mapp.* 39, 2317–2328. doi: 10.1002/hbm.23975
- Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* 146, 1082–1108. doi: 10.1016/j.neuroscience.2006.12.072
- Shalom, D. B., and Poeppel, D. (2008). Functional anatomic models of language: assembling the pieces. *Neuroscientist* 14, 119–127. doi: 10.1177/1073858407305726
- Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cognit. Neurosci.* 9, 648–663. doi: 10.1162/jocn.1997.9.5.648
- Simons, J. S., and Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nat. Rev. Neurosci.* 4, 637–648. doi: 10.1038/nrn1178
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., and Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychon. Bull. Rev.* 24, 1077–1096. doi: 10.3758/s13423-016-1217-0
- Takei, Y., Fujihara, K., Tagawa, M., Hironaga, N., Near, J., Kasagi, M., et al. (2016). The inhibition/excitation ratio related to task-induced oscillatory modulations during a working memory task: a multimodal-imaging study using MEG and MRS. *Neuroimage* 128, 302–315. doi: 10.1016/j.neuroimage.2015.12.057
- van Dijk, H., Nieuwenhuis, I. L. C., and Jensen, O. (2010). Left temporal alpha band activity increases during working memory retention of pitches. *Eur. J. Neurosci.* 31, 1701–1707.
- van Vugt, M. K., Schulze-Bonhage, A., Litt, B., Brandt, A., and Kahana, M. J. (2010). Hippocampal gamma oscillations increase with memory load. *J. Neurosci.* 30, 2694–2699. doi: 10.1523/JNEUROSCI.0567-09.2010
- Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550
- Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognit. Affect. Behav. Neurosci.* 3, 255–274.
- Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609–1612. doi: 10.1126/science.1139597
- Woodward, T. S., Cairo, T. A., Ruff, C. C., Takane, Y., Hunter, M. A., and Ngan, E. T. (2006). Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. *Neuroscience* 139, 317–325. doi: 10.1016/j.neuroscience.2005.05.043
- Wostmann, M., Schmitt, L. M., and Obleser, J. (2020). Does closing the eyes enhance auditory attention? Eye closure increases attentional alpha-power modulation but not listening performance. *J. Cognit. Neurosci.* 32, 212–225. doi: 10.1162/jocn\_a\_01403
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cognit. Sci.* 6, 37–46. doi: 10.1016/s1364-6613(00)01816-7

**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.

The reviewer CB declared a shared affiliation, with no collaboration, with one of the authors, NK, to the handling Editor at the time of review.

Copyright © 2020 Gudi-Mindermann, Rimmele, Bruns, Kloosterman, Donner, Engel and Röder. 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.



# Better Phonological Short-Term Memory Is Linked to Improved Cortical Memory Representations for Word Forms and Better Word Learning

Sari Ylinen<sup>1,2,3\*</sup>, Anni Nora<sup>4</sup> and Elisabet Service<sup>5</sup>

<sup>1</sup> CICERO Learning, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, <sup>2</sup> Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, <sup>3</sup> BioMag Laboratory, Helsinki University Central Hospital, Helsinki, Finland, <sup>4</sup> Department on Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland, <sup>5</sup> ARIEL Research Centre, Department of Linguistics and Languages, McMaster University, Hamilton, ON, Canada

## OPEN ACCESS

### Edited by:

Vitória Plai,  
Radboud University Nijmegen,  
Netherlands

### Reviewed by:

Johanna Maria Rimmele,  
Max Planck Society (MPG), Germany  
Jutta L. Mueller,  
University of Osnabrück, Germany

### \*Correspondence:

Sari Ylinen  
sari.ylinen@helsinki.fi

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 29 November 2019

**Accepted:** 08 May 2020

**Published:** 05 June 2020

### Citation:

Ylinen S, Nora A and Service E  
(2020) Better Phonological  
Short-Term Memory Is Linked  
to Improved Cortical Memory  
Representations for Word Forms  
and Better Word Learning.  
*Front. Hum. Neurosci.* 14:209.  
doi: 10.3389/fnhum.2020.00209

Language learning relies on both short-term and long-term memory. Phonological short-term memory (pSTM) is thought to play an important role in the learning of novel word forms. However, language learners may differ in their ability to maintain word representations in pSTM during interfering auditory input. We used magnetoencephalography (MEG) to investigate how pSTM capacity in better and poorer pSTM groups is linked to language learning and the maintenance of pseudowords in pSTM. In particular, MEG was recorded while participants maintained pseudowords in pSTM by covert speech rehearsal, and while these brain representations were probed by presenting auditory pseudowords with first or third syllables matching or mismatching the rehearsed item. A control condition included identical stimuli but no rehearsal. Differences in response strength between matching and mismatching syllables were interpreted as the phonological mapping negativity (PMN). While PMN for the first syllable was found in both groups, it was observed for the third syllable only in the group with better pSTM. This suggests that individuals with better pSTM maintained representations of trisyllabic pseudowords more accurately during interference than individuals with poorer pSTM. Importantly, the group with better pSTM learned words faster in a paired-associate word learning task, linking the PMN findings to language learning.

**Keywords:** magnetoencephalography, phonological short-term memory, language learning, paired-associate word learning, phonological mapping negativity

## INTRODUCTION

Phonological short-term or working memory (pSTM) has been suggested to play a critical role in language learning, contributing to the establishment of long-term memory traces (for reviews, see Gathercole and Baddeley, 1993; Baddeley et al., 1998; Service, 2013). A number of studies have shown a link between pSTM and the learning of first-language and foreign vocabulary



(e.g., Service, 1992; Service and Kohonen, 1995; Atkins and Baddeley, 1998; Gathercole et al., 1999) and syntax (French, 2006; French and O'Brien, 2008). In the working memory framework by Baddeley (1986) see also Baddeley and Hitch (1974), all speech has obligatory access to a phonological short-term store, and its contents are refreshed by a rehearsal component that prevents the decay of memoranda. According to Baddeley et al. (1998), this phonological loop is critical for word learning because it is used to maintain unfamiliar sound patterns in memory while more permanent memory representations are being constructed. An alternative view has questioned the direction of causality, suggesting instead that vocabulary size may determine pSTM capacity (e.g., Melby-Lervåg et al., 2012). Longitudinal studies of second-language learning that have followed the accumulation of vocabulary from the start have lent support to the original view by Baddeley, Gathercole and colleagues (Service and Kohonen, 1995; French, 2006). Recently, brain stimulation studies have linked the storage component of pSTM, housed in the left supramarginal gyrus, to the ability to support the maintenance of verbal order (Papagno et al., 2017; Savill et al., 2019). The ability to represent the order of phonemes in a novel word form, and the order of words in phrases, has been suggested as the mechanism relating pSTM to learning of both the phonological structure of novel word forms and grammatical phrases (Gupta and Tisdale, 2009).

At the level of neuroanatomy, the phonological loop was first suggested to rely on Broca's area and the left supramarginal gyrus (Paulesu et al., 1993; for more recent work, see Papagno et al., 2017; Savill et al., 2019). Later neuroimaging studies on pSTM point to a network involving also posterior temporal or temporo-parietal areas [posterior superior temporal gyrus (STG), posterior superior temporal sulcus (STS), posterior planum temporale (PT), or Sylvian-parietal-temporal areas (Spt)] areas (Buchsbaum et al., 2001; Hickok et al., 2003; Buchsbaum and D'Esposito, 2008; McGettigan et al., 2011; Richardson et al., 2011; Herman et al., 2013), also supported by lesion studies (Baldo et al., 2012). Extending the phonological loop model, rehearsal has been suggested to be a process of circulating information between phonological input and output buffers, involving temporo-parietal cortex and left inferior frontal cortex, respectively (Jacquemot and Scott, 2006; Herman et al., 2013). A framework for STM maintenance and language repetition by Majerus (2013) proposes that speech is encoded and phonological representations maintained in fronto-temporal language networks (dorsal and ventral speech processing streams; Scott and Johnsrude, 2003; Hickok and Poeppel, 2007) and attentional focalization is coordinated from a fronto-parietal network.

Linking the neural implementation of auditory working memory or pSTM with word learning, both short-term and longer-term changes related to establishing memory representations for novel words have been shown in auditory cortices (Davis and Gaskell, 2009; Davis et al., 2009). Thus, when novel words are maintained in pSTM, cortical responses in temporal areas reflect the quality of the phonological memory representations, with contributions from frontal motor representations (Nora et al., 2015). In addition to neocortical areas, medial temporal areas have been shown to be important for initial encoding and maintenance during word learning,

and a recent study by Kumar et al. (2016) demonstrates the involvement of hippocampus as well as fronto-temporal connections in all stages of the working memory process, namely, encoding, maintenance, and retrieval.

Encoding of speech input has been suggested to be the primary determinant for efficient pSTM functioning (Barry et al., 2009, 2011) and long-term learning (Service et al., 2007). However, also the maintenance of phonological information has been shown to exert an influence on remembering and learning words (Davachi et al., 2001). pSTM contents are thought to be affected by decay or interference but can be maintained for a longer period by rehearsal or attentional refreshing (i.e., focusing attention on memoranda for their maintenance; Camos et al., 2009; Camos and Barrouillet, 2014; Lewandowsky and Oberauer, 2015). When considering the learning of spoken words during natural communication, a typical source of interference is the auditory input following the to-be-learned word. In this case, efficient maintenance may strengthen the pSTM representation and protect it from interference. This raises questions whether individual learners differ in their ability to maintain word forms in pSTM during interfering auditory input, whether this is reflected in the cortical activation during phonological processing, and whether word learning varies as a function of this phonological maintenance ability and its neural correlates. Some previous studies (e.g., Čeponienė et al., 1999; Barry et al., 2009) have compared brain responses in participants with poorer and better pSTM. However, these studies neither used tasks with active pSTM maintenance during interference, nor actual word-learning tasks in adults.

We used magnetoencephalography (MEG) to study the maintenance of novel word forms in pSTM by rehearsal during interfering auditory input. Firstly, by comparing brain responses between participants that have better or poorer pSTM, we aimed to determine whether pSTM maintenance during interference differs between these groups. Secondly, to clarify how the ability to maintain phonological representations is the brain influences language learning, we investigated whether these groups differ in their word-learning ability. The experimental paradigm used here has been described in Ylinen et al. (2015). In each trial, participants first heard a target pseudoword that they were instructed to rehearse covertly. This target pseudoword was followed by random distractors as well as probe stimuli that fully or partially matched with the rehearsed target. This condition was compared with a control condition in which pseudoword rehearsal had been replaced by silent counting of recurring visual symbols (i.e., there was no match between pSTM contents and auditory stimulation). We assumed that the rehearsal of word forms in working memory would re-activate or refresh the phonological representations of to-be-remembered target pseudowords and protect them from interference caused by auditory distractors. Probe stimuli were used to test the level of activation and accuracy of these representations in participants with better or poorer pSTM.

The rationale of using probe stimuli is based on the findings that covert speech used in rehearsal in pSTM generates forward prediction of the rehearsed item, projected from frontal cortex speech areas to auditory cortex in the form of efference copy signals. Efference copies are internal copies of efferent commands

produced by the motor system (cf. Sperry, 1950; von Holst and Mittelstaedt, 1950). The forward prediction appears to regulate the activation of auditory cortex (Houde et al., 2002; for covert speech see Tian and Poeppel, 2010, 2012; Ylinen et al., 2015). Auditory input matching overt or covert speech has been found to suppress responses in auditory cortex, whereas mismatching input enhances the responses (Chang et al., 2013; Ylinen et al., 2015). Thus, when covert rehearsal is combined with matching or mismatching auditory probe stimuli, brain responses can be used to index pSTM maintenance.

We were particularly interested in detecting neural activity that previous studies have linked with a discrepancy from phonological expectations. In tasks involving listening to wordlike stimuli while phonological expectations are active, specifically the event-related potential or field (ERP/ERF) component named the phonological mapping negativity (PMN, formerly phonological mismatch negativity) has been observed (Connolly and Phillips, 1994; D'Arcy et al., 2004; Kujala et al., 2004). PMN is elicited at about 200–250 ms after an unexpected phoneme is encountered and it has been located to anterior temporal cortex (Kujala et al., 2004). Enhanced responses are seen when phonological expectations based on sentence or phonotactic context or covert speech are not met. Therefore, PMN can be used to index pSTM maintenance by covert rehearsal. The PMN has been associated with phonology because it is similarly elicited for words and pseudowords, ruling out dependence on lexical-semantic processing for its elicitation (Newman and Connolly, 2009).

The PMN is thought to reflect mapping of auditory input onto phonemes in speech recognition, yet the results for distinguishing children with language or literacy disorders from typically developing children based on the PMN have been mixed (Bonte and Blomert, 2004; Desroches et al., 2013; Malins et al., 2013). However, it is noteworthy that previous work has been limited to inspecting mismatches in the first syllable. As discussed by Connolly et al. (2001), the PMN response is likely not limited to onset processing. Here we report responses to both salient onsets and less salient third syllables in trisyllabic stimuli. We predicted that auditory pseudowords mismatching the rehearsed pseudoword would elicit a stronger magnetic PMN response compared with pseudowords matching it. Moreover, we thought that the PMN process might be sensitive to pSTM abilities. We, therefore, hypothesized that the brain responses of groups with better and poorer pSTM might differ from each other at the PMN latency as activated phonological representations are thought to be necessary for PMN elicitation. When hearing distractors, participants with better pSTM were expected to show more accurate maintenance of phonological items due to more persistent and resilient phonological representations. In contrast, the groups' responses were not expected to differ from each other in the control condition with no resemblance between internal speech and the auditory stimuli. We further thought that the less salient third syllables may be more sensitive to group differences (cf. Service and Maury, 2003). Finally, based on behavioral studies (Service and Kohonen, 1995), we also hypothesized that the neural correlates of pSTM should be linked to indices of language learning, such that if there are differences in PMN between the

groups, similar differences should be found also in a paired-associate novel word learning task.

## MATERIALS AND METHODS

### Pre-test

#### Ethics Statement

All subjects signed a written informed consent form before participation in the experiment. The pre-test was approved by the Ethical Committee of the Faculty of Behavioural Sciences, University of Helsinki.

#### Participants

Fifty-one university students [28 females, 23 males; age 19–34 years, mean age 23.8 years (*y*), *SD*: 4.09] volunteered for the pre-test. The inclusion criteria were right-handedness, normal hearing, no early bilingualism, Finnish as the native language, no language or speech disorders including dyslexia, and no neurological or psychiatric disorders or drug addiction. Three females were excluded from analysis because of these criteria.

#### Procedure

The pre-test included three cognitive tasks, with order of presentation counterbalanced within the participant group: (1) pseudoword pair repetition, (2) pseudoword memory span, and (3) paired-associate novel word learning. The pseudoword repetition task (Čeponienė et al., 1999) included 20 pairs of 4- and 5-syllabic pseudowords with relatively complex, but for Finnish legal, phonological structure (e.g., /sohraelma-/nahterkop:io/). Participants were instructed to listen to the pseudoword pairs, then say “toistan” (I repeat) and then repeat aloud the pseudoword pairs. The initial “toistan” was intended to wipe echoic memory content. The pseudoword span task (Numminen et al., 2002) included lists of spoken disyllabic CVCV pseudowords for immediate recall. Participants were first presented with 10 lists of three pseudowords, then 10 lists of four pseudowords, and finally 10 lists of five pseudowords. After hearing a list, the participants' task was to repeat back the list in the order of presentation. Word learning was studied with a paired-associate learning task with eight familiar Finnish words each paired with a Finnish-sounding pseudoword (Čeponienė et al., 1999). Four items were disyllabic and four trisyllabic. After reading each of the eight pairs, participants saw one Finnish word at a time and were instructed to say aloud the corresponding pseudoword. The task had four trials during which the same eight word-pseudoword pairs were presented in random order. Out of these tasks, pseudoword pair repetition and pseudoword memory span were used to assign participants to two pSTM groups for the MEG experiment (better or poorer pSTM). The paired-associate word learning task was, in turn, used to compare language learning ability between these groups.

### Statistical Analysis of Paired-Associate Word Learning

Paired-associate word learning scores were submitted to a mixed ANOVA including within-subjects factors Word length (short,

long), Trial (1, 2, 3), and a between-subjects factor Group (better pSTM, poorer pSTM).

## MEG Experiment

### Ethics Statement

All subjects signed a written informed consent form before participation in the experiment. The MEG experiment was approved by the Research Ethics Committee of Helsinki University Central Hospital. All experiments were carried out according to the Declaration of Helsinki.

### Participants

Based on pSTM performance in the pre-test, a standard compound score was formed by transforming pseudoword repetition and pseudoword span raw scores (the number of items repeated correctly and the number of lists repeated correctly, respectively) into z-scores for each task and participant and then averaging the z-scores across the tasks for each participant. Thirteen participants with the lowest scores and 13 participants with the highest scores were invited to take part in the MEG recording. They formed a poorer and a better pSTM group, respectively. Participants were not informed of the group they belonged to. One participant in each group was unavailable for the MEG session, resulting in 12 participants in the poorer pSTM group (six females and six males; mean age 23.08 years, *SD*: 2.98) and 12 in the better pSTM group (seven females and five males; mean age 23.67 years, *SD*: 4.62). The MEG participants' standard pSTM score ranged from  $-2.61$  to  $-0.58$  in the poorer pSTM group and from  $0.48$  to  $1.64$  in the better pSTM group. The average raw score was 8.33 for pseudoword repetition and 6.67 for pseudoword span in the poorer pSTM group and 17.33 for pseudoword repetition and 16.83 for pseudoword span in the better pSTM group.

### Stimuli

The auditory and visual stimuli in the MEG experiment were the same as those in Ylinen et al. (2015; see **Figure 1A**). The auditory stimulus material included 30 different pseudowords, each having two variants (60 stimuli in total). The pseudowords had a CVCVCCV structure (e.g., /pukot:o/, /tavek:o/, /konata:/) with geminate stop consonants including a silent phase before the release burst in the third syllable. The pseudowords complied with the phonotactic structure of Finnish but were unfamiliar to the participants. The stimuli were produced at a normal speaking rate by a female native speaker of Finnish and digitally recorded with a Eurorack MX1604A Mixer and a Rode NT2-A microphone in an acoustically shielded room. The final experimental stimuli were chosen from several variants on the basis of judgments of three naïve native speakers of Finnish, who assessed the goodness of the stimuli with respect to their native language. The chosen pseudowords were further modified with Praat 5.0.40. (Boersma and Weenink, 2008) as follows: the intensity of the stimuli was scaled to 90% and the durations of the syllables within the stimuli were equalized preserving their typical ratio (the 1<sup>st</sup> and 2<sup>nd</sup> syllable excluding the final consonant 260 ms; the silent phase of the geminate stop 220 ms; the 3<sup>rd</sup> syllable (excluding the silent phase of the stop) 120 ms; in total 600 ms, see **Figure 1A**). In

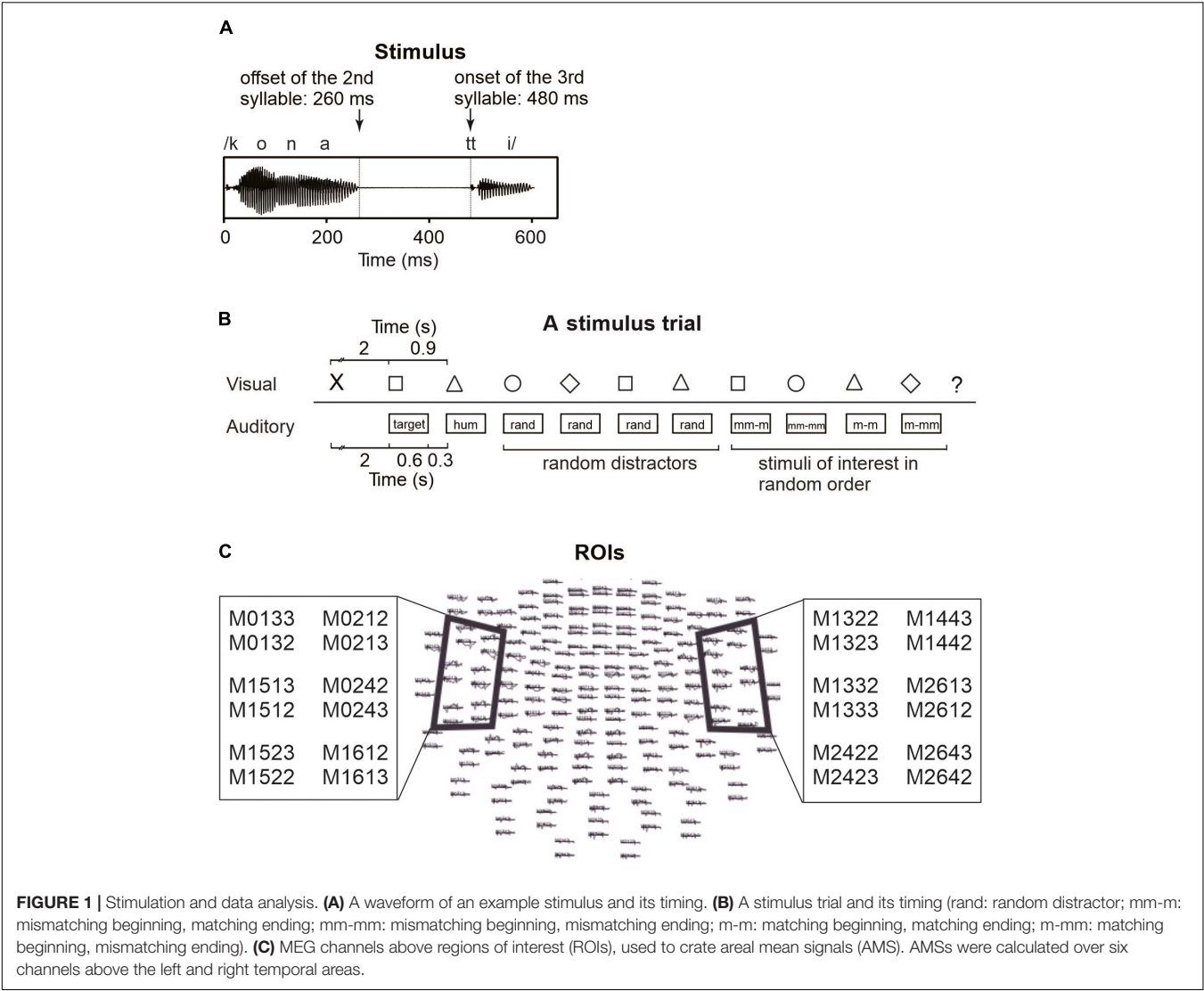
addition to speech stimuli, the stimulus set included a humming sound that was created by filtering a pseudoword stimulus [pamup:a] with a 250 Hz low-pass filter and a harmonic tone of 75 ms duration and 500 Hz fundamental frequency (with harmonic partials of 1000 and 1500 Hz). Responses to these non-speech sounds were not analyzed.

The participants were also presented with visual stimuli on a screen in front of them (see **Figure 1B**) simultaneously with the auditory stimuli. The stimuli were geometric shapes (a square, a circle, a triangle, a diamond) displayed in black on a gray background. The stimulus presentation was commanded by a script written in Presentation 12.2 (Neurobehavioral Systems, Albany, NY, United States).

### Procedure

The MEG experiment followed the procedure of Ylinen et al. (2015). There were two task conditions, rehearsal and control, in both of which participants heard stimulus pseudowords from loudspeakers and simultaneously saw visual symbols on the screen. In the rehearsal condition, participants were instructed to covertly rehearse the first auditory pseudoword of the trial each time an auditory stimulus was heard (and a simultaneous visual symbol was shown). To ensure that the participants rehearsed the heard items as instructed, they had to say the rehearsed pseudoword aloud at the end of the trial when a question mark was shown on the screen. In the control condition, the participants' task was to count the number of occurrences of the visual symbol that had been presented first in that trial. To ensure that the participants performed the task as instructed, they had to say aloud the number of counted symbols at the end of the trial when a question mark was shown on the screen. The two conditions were run in counterbalanced order within the two participant groups. The 100 trials of each condition were divided into five blocks, and 10 s breaks were inserted between the blocks. Participants were instructed to blink extensively during breaks to reduce blinking during the experimental trials. Each block started with the presentation of 20 repetitions of a harmonic tone, after which the task began. Participants were allowed to take a break between conditions. Instructions for the task in question were given immediately before each condition.

The tasks differed between the two conditions, but the stimulation was identical (with the exception that the order of the 30 pseudowords was randomized separately for the conditions). Each trial (see **Figure 1B**) consisted of a sequence as follows. First, a cross showed up on the screen as a signal to get ready to perform the task. After 2 s, the participants heard a pseudoword and saw the first geometric shape symbol. Depending on the condition, they were to remember and covertly rehearse the pseudoword, or to silently count occurrences of the symbol during the trial. Then a humming sound (a low-pass filtered pseudoword) was presented to set the rhythm for the covert rehearsal. A hum instead of another pseudoword was used to avoid immediately erasing the to-be-remembered pseudoword from phonological memory before the participants had got started with the rehearsal. After the hum, four random pseudowords that did not resemble the to-be-remembered word



**TABLE 1 |** Time windows used for quantifying areal mean signals (AMS).

Rehearsal condition								
1 <sup>st</sup> syllable match		1 <sup>st</sup> syllable mismatch		3 <sup>rd</sup> syllable match		3 <sup>rd</sup> syllable mismatch		
	LH	RH	LH	RH	LH	RH	LH	RH
Better pSTM	147–197	190–240	188–238	192–242	648–698	646–696	666–716	656–706
Poorer pSTM	155–205	152–202	155–205	192–242	650–700	645–695	665–715	653–703

Control condition								
1 <sup>st</sup> syllable “match”		1 <sup>st</sup> syllable mismatch		3 <sup>rd</sup> syllable “match”		3 <sup>rd</sup> syllable mismatch		
	LH	RH	LH	RH	LH	RH	LH	RH
Better pSTM	150–200	183–233	185–235	177–227	663–713	651–701	666–716	661–711
Poorer pSTM	153–203	200–250	148–198	192–242	699–749	653–703	655–705	665–715

Note that in the control condition, there is no actual match to pSTM contents like in the rehearsal condition, yet the auditory stimuli are the same in the two conditions. LH, left hemisphere; RH, right hemisphere.

were presented as distractors. These were followed by four stimuli in random order. These included (1) the same pseudoword as the rehearsed word (but not identical recording), (2) a minimal-pair pseudoword with a different final vowel (e.g., for rehearsed pseudoword [pukot:a], minimal pair [pukot:o]), (3) a different pseudoword but with the same ending (e.g., for



rehearsed pseudoword [pukot:a], a pseudoword with the same ending [konat:a]), and (4) a random pseudoword not resembling the rehearsed pseudoword (e.g., for rehearsed pseudoword [pukot:a], a random pseudoword [kilep:o]). All presentations of auditory pseudowords were accompanied by the simultaneous presentation of visual symbols, but the auditory and visual stimuli were not otherwise associated with each other. The same visual symbol was presented 2–4 times in random order during a trial. After 10 simultaneous presentations of auditory and visual stimuli, a question mark was shown on the screen, indicating that participants should say aloud, depending on the task, either the rehearsed pseudoword or the number of counted symbols in the trial. This was to make sure the participants were performing the tasks as instructed. Since the to-be-remembered pseudowords were followed by four auditory distractors and four other stimuli (see **Figure 1B**), it is unlikely that the participants could have remembered and said the pseudowords aloud, if they had not rehearsed them in pSTM (i.e., without active maintenance of the to-be-remembered pseudowords, the distractors would have erased them from phonological memory). A new trial started 2.5 s after the presentation of the question mark. Within each trial, the interstimulus interval was 300 ms.

### MEG Recording and Analysis

ERFs were recorded with a 306-channel Vectorview MEG device (Elekta Neuromag, Elekta Oy, Helsinki, Finland) with 204 planar gradiometers and 102 magnetometers. Simultaneously, electroencephalography (EEG) was recorded from three scalp sites (Fz, Cz, and Pz, referenced to the left mastoid; EEG analysis is not reported here). The participants sat in a magnetically and acoustically shielded chamber with their head covered by the helmet of the MEG device. They were instructed to avoid blinking except during the breaks between the blocks and not to move their head (even during the breaks). Before the experiment, four head-position indicator coils were attached to each participant's head and their location with respect to anatomical landmarks of the head (nasion and pre-auricular points) was determined by an Isotrak 3D digitizer (Polhemus, Colchester, VT, United States). The position of the head within the helmet of the MEG device was determined by feeding current to the coils and measuring their locations in the helmet. MEG and EEG signals were recorded with a 600 Hz sampling rate and filtered with a band pass of 0.1–200 Hz.

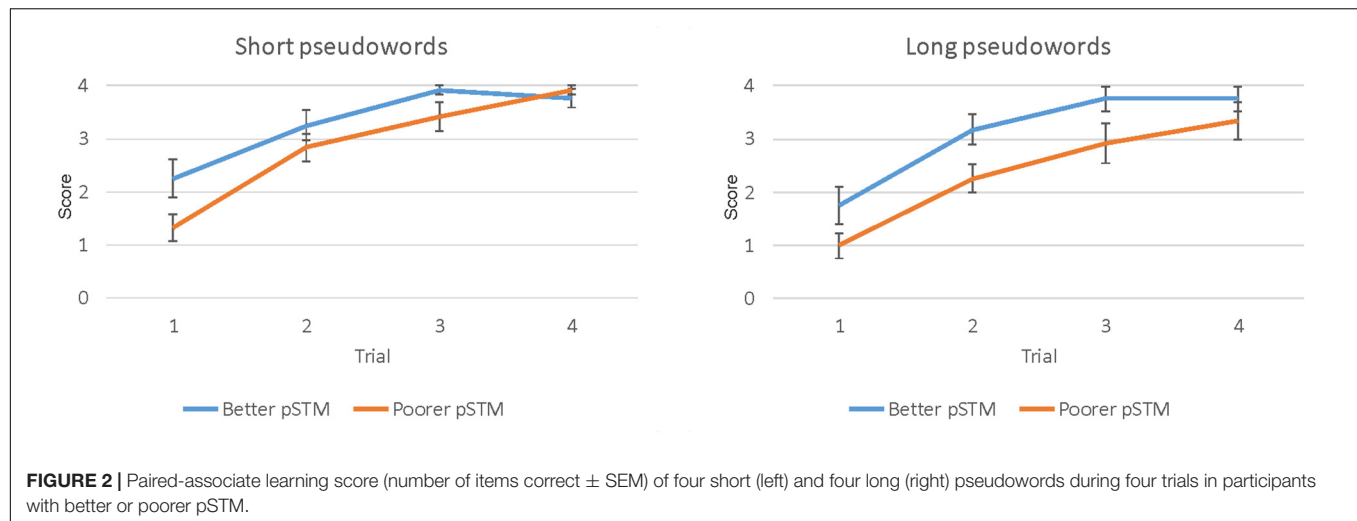
The data were off-line filtered with band pass of 0.5–30 Hz (slope 12 and 24 dB/octave, respectively) and artifacts exceeding 1200 fT/cm on gradiometers were rejected. Baseline was set to 100 ms windows preceding the onset of the analyzed syllables (–100 to 0 ms for the first syllable and 380–480 ms for the third syllable; cf. Barry et al., 2009). To determine response strength, areal mean signals (AMSs) were calculated from six gradiometer pairs above the temporal lobe of each hemisphere (see **Figure 1C** for channel locations) where the responses of interest were expected to be elicited (Ylinen et al., 2015). Time windows for analysis were selected based on the latencies of the highest grand-average AMS peaks in the time window of 150–300 ms from syllable onset (i.e., around PMN latency), determined

separately for the four stimulus types (for time windows in each condition, see **Table 1**). 50 ms time windows were centered at the latencies of these peaks to calculate response strength for each experimental condition. These conditions included first-syllable match, first-syllable mismatch, third-syllable match, and third-syllable mismatch (i.e., rather than using mismatch-minus-match difference waves, response strengths were calculated from the AMS for each condition). The AMS for the first syllable matching with the target was the average across the responses to all the stimuli with the matching beginning (i.e., including pseudowords that were the same as the rehearsed target, e.g., target [pukot:a] vs. [pukot:a], and the stimuli that had the same beginning but mismatching ending, e.g., target [pukot:a] vs. [pukot:o]). The AMS for a mismatching first syllable was the average across the responses to all the stimuli with a mismatching beginning (i.e., including the pseudowords with mismatching beginning and ending, e.g., rehearsed target [pukot:a] vs. [kilep:o], and those with mismatching beginning but matching ending with respect to the target, e.g., target [pukot:a] vs. [konat:a]). The AMSs for the third syllable included responses to match (e.g., [pukot:a] vs. [pukot:a]) or mismatch (e.g., [pukot:a] vs. [pukot:o]) with respect to the third syllable of the rehearsed target (note that in both cases, the pseudoword beginnings matched the target until the third-syllable onset at 480 ms). We expected the items that had a mismatching first/third syllable with respect to the rehearsed target to elicit a stronger response at the PMN latency compared with the matching items. Moreover, we expected stronger PMN responses in the group with better pSTM capacity.

To control for group differences in overall engagement in the rehearsal task, we also inspected the suppression effect of N1 caused by covert rehearsal in the third syllable. If participants were performing the rehearsal task, the covert rehearsal of items matching the auditory stimuli should induce suppressed N1 responses as compared to the control condition (Ylinen et al., 2015). If the N1 suppression effect was different between the groups, then the groups' effort or engagement in rehearsal could have been different. AMSs were calculated from the same six gradiometer pairs as included for PMN, but only in the left hemisphere, where suppression effects at the syllable level were expected to occur (Ylinen et al., 2015). Time windows for analysis were selected based on the latencies of the highest grand-average AMS peaks at 100–140 ms from the 3<sup>rd</sup> syllable onset. AMS peaks were determined separately for the different stimulus types, and 50 ms time windows were centered at the latencies of these peaks to calculate response strength for each experimental condition (time windows for N1 ranged from 575–625 to 586–636 ms, i.e., from 120 to 131 ms from the 3<sup>rd</sup> syllable onset).

### Statistical Analysis of AMS

For statistical analysis of AMS strength, we used mixed ANOVA with repeated factors Syllable (first, third), Task (rehearsal, control), Match [matching, mismatching with the rehearsed syllable (or equivalent stimulus in the control condition)], Hemisphere (left, right) and the between-subjects factor Group (better pSTM, poorer pSTM). In addition, we



**TABLE 2 |** Paired-associate word learning scores ( $\pm$ SD) averaged across trials 1–3 in participants with better or poorer phonological short-term memory (pSTM).

	Short words	Long words
Better pSTM	3.13 ( $\pm$ 0.78)	2.89 ( $\pm$ 0.89)
Poorer pSTM	2.53 ( $\pm$ 0.78)	2.06 ( $\pm$ 0.81)

report step-down analyses with repeated factors Task (rehearsal, control), Match [matching, mismatching with the rehearsed syllable (or equivalent stimulus in the control condition)], Hemisphere (left, right) and the between-subjects factor Group (better pSTM, poorer pSTM). Consequent interactions were followed up with Bonferroni-corrected pairwise comparisons. To ensure that both groups were engaged by the rehearsal task in a similar manner, the suppression effect of N1 that is caused by covert rehearsal was compared between the groups by submitting N1 AMSs for the third-syllable match condition to an ANOVA including factors Task (rehearsal, control) and Group (better, poorer pSTM).

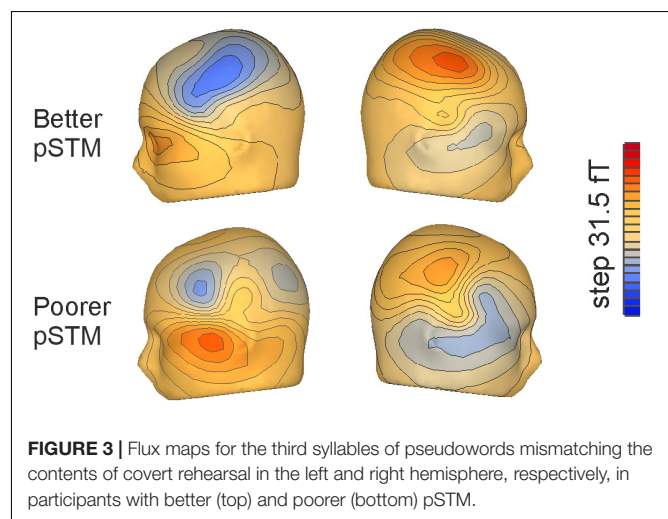
## RESULTS

### Paired-Associate Word Learning

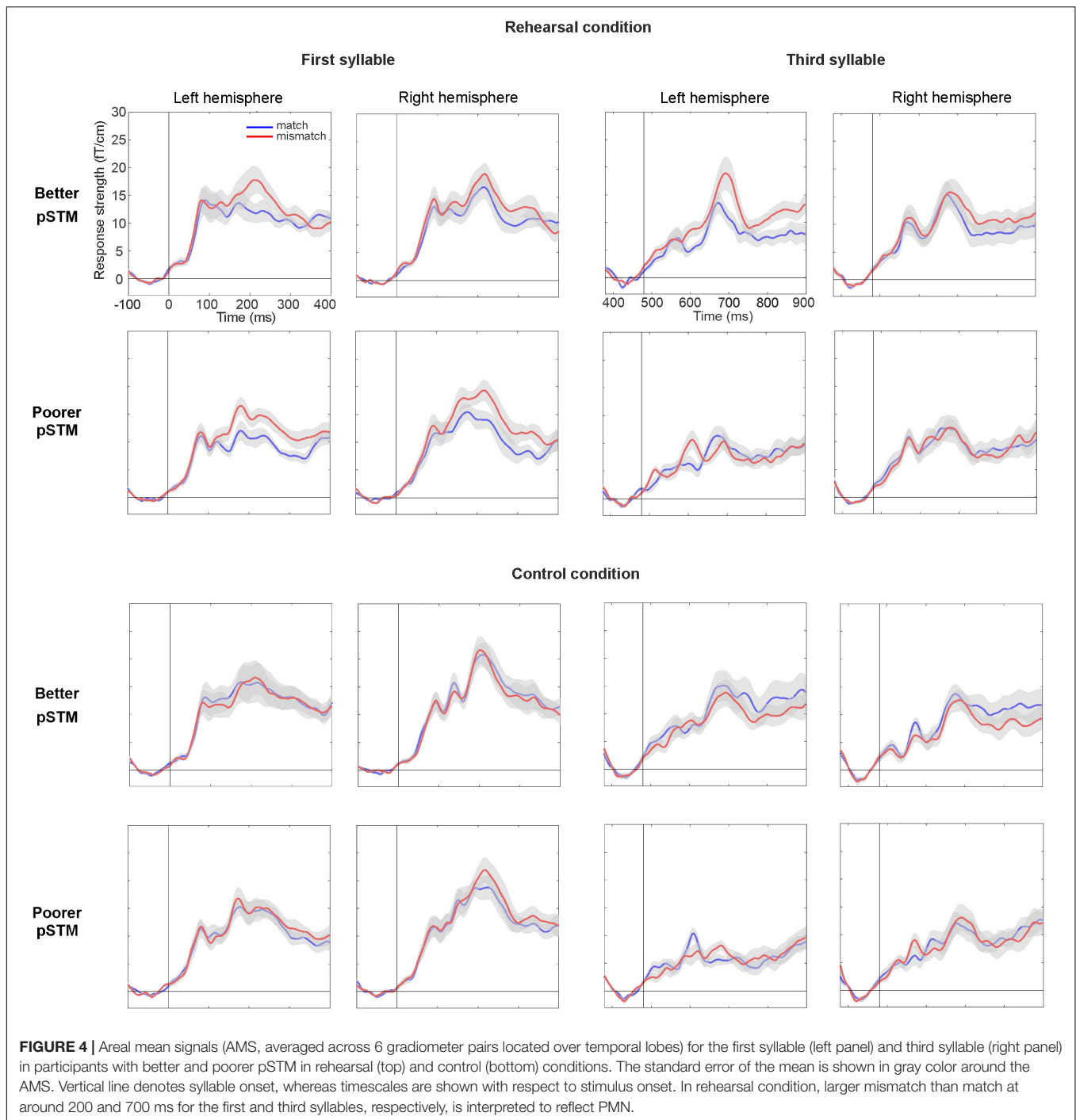
The paired-associate word learning task had four trials. However, in the last trial both groups performed close to ceiling, indicating that this trial could not show group differences accurately (see **Figure 2**). Therefore, the results of trials 1–3 were used in the analysis. The mixed ANOVA showed the main effects of Word length [ $F(1,22) = 5.78$ ,  $p = 0.025$ ], with higher scores for shorter than longer words, Trial [ $F(2,22) = 71.93$ ,  $p < 0.001$ ], with higher scores on later than earlier trials, and Group [ $F(1,22) = 5.83$ ,  $p = 0.025$ ], with higher scores in the better than the poorer pSTM group (see **Table 2** and **Figure 2**).

### Areal Mean Signals

In line with previous PMN literature (Kujala et al., 2004), syllables mismatching the contents of covert rehearsal induced



activity over anterior temporal cortex (see **Figure 3**). At the PMN latency, a five-way ANOVA with the AMS as dependent variable showed a significant four-way interaction of Group  $\times$  Syllable  $\times$  Task  $\times$  Match [ $F(1,22) = 4.81$ ,  $p < 0.039$ ]. This was further explored by separate ANOVAs of responses to the first and the third syllable. In the first-syllable analysis, the main effect of Match was significant [ $F(1,22) = 23.51$ ,  $p < 0.001$ ] due to stronger responses to mismatching than matching stimuli. The main effects of Task [ $F(1,22) = 17.17$ ,  $p < 0.001$ ] and Hemisphere [ $F(1,22) = 9.35$ ,  $p = 0.006$ ] were also significant due to stronger responses for the control than rehearsal task and stronger responses over the right than left hemisphere, respectively. There was also a significant interaction of Task  $\times$  Match [ $F(1,22) = 12.21$ ,  $p = 0.002$ ]. According to pairwise comparisons, the responses were significantly stronger to mismatch than match in both tasks (for control,  $p = 0.015$ ; for rehearsal,  $p < 0.001$ ). No interactions or effects involving Group were observed for the first syllable (see **Figure 4**).



In contrast, for the third syllable we found an interaction of Group  $\times$  Task  $\times$  Match [ $F(1,22) = 9.59$ ,  $p = 0.005$ ], which was due to significantly stronger responses to mismatching than matching stimuli in the better pSTM group in the rehearsal task ( $p < 0.024$ ), but not in the control task (n.s.). Furthermore, a significant interaction of Group  $\times$  Task  $\times$  Hemisphere  $\times$  Match [ $F(1,22) = 6.85$ ,  $p = 0.016$ ] showed that the rehearsal effect in the better pSTM group was driven by significantly stronger responses to mismatch than match in the left

hemisphere ( $p = 0.013$ , see **Figure 4**). All other pairwise comparisons for this interaction were non-significant. In the group with poorer pSTM, Match comparisons were not significant in either task.

An ANOVA for the N1 component for the third syllable in the matching condition was run to establish that there were no group differences in the auditory effects of rehearsal as compared with the control task. Rehearsal would be expected to result in auditory cortex suppression effects in the matching

condition. The analysis showed a significant main effect of Task [ $F(1,22) = 10.86$ ,  $p = 0.003$ ], but the effects of Group [ $F(1,22) = 0.057$ , n.s.] and Task  $\times$  Group [ $F(1,22) = 0.143$ , n.s.] did not approach significance.

## DISCUSSION

By presenting auditory probes matching or mismatching the word forms rehearsed in pSTM, the present study aimed to determine how pSTM ability affects the maintenance of word forms during interference and whether this ability and its neural correlates are linked to language learning. Firstly, two groups with different pSTM capacities were compared in paired-associate word learning of word-pseudoword pairs. Although both groups performed close to ceiling on the fourth trial, those with better pSTM had significantly higher learning scores during the first three trials. Thus, those with better pSTM learned the associations faster, with fewer repetitions. Secondly, comparison of AMSs over the temporal cortices did not suggest rehearsal-related differences between the pSTM groups in the processing of the first syllable of the pseudowords, yet only the better pSTM group showed a significant PMN response for a mismatch in the third syllable.

The maintenance of the phonological form of pseudowords in pSTM by covert rehearsal modulated responses peaking around the typical PMN latency, that is, about 200 ms from 1<sup>st</sup> and 3<sup>rd</sup> syllable onsets (about 200 and 680 ms from stimulus onset, respectively). In the rehearsal condition, the effect of covert rehearsal on the processing of the first syllable was reflected in a significantly stronger response to mismatching than matching stimuli in both groups. This is in line with earlier PMN findings (e.g., Connolly et al., 2001) as well as findings showing that matching covert speech suppresses auditory responses (Numminen and Curio, 1999; Kauramäki et al., 2010; Tian and Poeppel, 2012; Ylinen et al., 2015), whereas mismatching input elicits enhanced responses (Chang et al., 2013; Ylinen et al., 2015). However, only the better pSTM group showed an enhanced PMN response to a mismatch in the third syllable in the rehearsal condition. This result indicates that pSTM ability modulated the phonological processing accuracy of the endings of trisyllabic word forms. Different effects with respect to pSTM between the first and third syllables suggest that the role of pSTM in the processing of phonological sequences may differ between word beginnings and endings or between shorter and longer words, which is in line with previous results on phonological memory (Service and Maury, 2003). The pattern of results suggests that pSTM ability determines the accuracy of phonological representations for all phonemes of novel words during interfering input. Those with poorer pSTM may be able to represent accurately only short word forms or beginnings of longer word forms and be challenged to fully represent novel multi-syllabic word forms.

The group differences in representing trisyllabic word forms in pSTM as reflected by the PMN can be accounted for by differences in either pSTM maintenance (see Buchsbaum et al., 2005; Buchsbaum and D'Esposito, 2008)

or encoding input into pSTM (Barry et al., 2011). Regarding possible group differences in pSTM maintenance, one may ask whether participants with poorer pSTM rehearsed the words less extensively despite the same instructions given to the two groups. Given that both N1 suppression and PMN enhancement may reflect the ongoing rehearsal process, such a difference does not seem likely because N1 for the probe that matched the rehearsed pseudoword appeared similarly diminished in both groups in the rehearsal condition. This suggests that the participants with poorer pSTM rehearsed the pseudowords covertly as requested and must have had at least some kind of active representations for the rehearsed pseudowords that could modulate N1. Why, then, would those with poorer pSTM be unable to maintain strong enough pSTM representations to elicit PMN in the third syllable despite rehearsal? One possible answer is that their pSTM capacity was overloaded by rehearsal of a trisyllabic novel form and processing of an incoming form with also three syllables. Together this task requires memory for six ordered syllables, if aligned third syllables are to elicit a mismatch response.

Another possibility is that since participants were requested to rehearse pseudowords covertly along with the rhythm of regularly presented stimuli, brain responses may have been modulated by the participants' ability to synchronize their rehearsal with auditory input. A recent study by Assaneo et al. (2019) has suggested that individuals' spontaneous ability to synchronize their speech to an isochronous train of auditorily presented syllables is linked to differences in white matter and brain-to-stimulus synchronization over frontal areas. However, rather than spontaneous synchronization, our task more closely resembles metronome-beat synchronized speech, where participants have been very accurate in keeping the external rhythm, with mean differences in actual and expected time between the productions being within 10 ms (Davidow et al., 2010). Although it is not clear to which extent PMN might be modified by synchronization abilities, previous studies have shown that synchronized rehearsal is not a prerequisite for PMN elicitation. The PMN is often elicited in a task where a word is first manipulated in one's mind and then an auditory stimulus is presented afterward (see, e.g., Connolly et al., 2001). Thus, poor synchronization skills cannot fully account for the lack of third-syllable PMN response in the poorer pSTM group.

Besides pSTM maintenance by rehearsal, group differences in third-syllable PMN might also be influenced by differences in the encoding process of auditory stimuli to pSTM for rehearsal. Previous research by Barry et al. (2011) has suggested that encoding words into memory results in larger hemodynamic responses in individuals with better non-word repetition (pSTM ability). In another study, Barry et al. (2009) found that in an oddball paradigm, those with poorer non-word repetition had smaller late discriminative negativity (LDN) responses for pseudoword-internal third syllables of auditory stimuli, interpreting this to reflect less efficient encoding. In particular, they suggested that in poor non-word repeaters, syllable recognition is not rapid enough and, therefore, earlier syllables interfere with the processing of later syllables of longer words. Consistent pSTM effects in the processing of the third syllable across studies (i.e., the current study and Barry et al., 2009)



support the view that memory capacity is linked to these effects via word length. We do not, however, find in our data any consistent differences between better and poorer pSTM groups in the pace of processing (see **Table 1** and **Figure 3**). In addition, we found the group differences in responses to the final syllable. According to Barry et al. (2009), the processing of the final syllable should have recovered from a cumulative memory load effect in those with poorer pSTM, if their problem was a slower rate of the encoding process.

Nevertheless, it is still possible that the group differences were related to encoding, for example via the code used in pSTM maintenance. Although we have previously argued that the code of covert rehearsal in our task is most likely phonological (Ylinen et al., 2015), recent literature suggests that pSTM may use both acoustic storage and categorical representations (Joseph et al., 2015) and that items can be maintained by rehearsing phonologically or by using domain-general attentional refreshing (i.e., focusing attention on memoranda for their maintenance; Camos et al., 2009; Camos and Barrouillet, 2014; Lewandowsky and Oberauer, 2015). These studies suggest that phonological rehearsal is not a necessity for the maintenance of verbal material in pSTM. In a similar vein, our inner speech may vary with respect to the detail of its phonological formulation. Therefore, one possible account for our pattern of results is that participants in the poorer pSTM group used less phonological means of maintenance during pSTM tasks, for example by occasionally (or consistently) maintaining acoustic-phonetic representations via attentional refreshing. The code used in pSTM maintenance, in turn, could either be due to the efficiency of phonological encoding process or the efficiency of the maintenance process itself. Although this account is speculative in the sense that it was not part of our original hypothesis, it could explain the lack of third-syllable PMN in participants with poorer pSTM while at the same time they showed similar N1 effects for the third syllable as the better pSTM group. Further research is needed to clarify the effect of pSTM ability on the code used in pSTM.

Unexpectedly, ANOVA suggested stronger neural activation for mismatching than matching first syllables of the pseudowords also in the control condition. As illustrated by **Figure 4**, however, this difference is more subtle than in the rehearsal condition (particularly in the participants with better pSTM). Note that since the control condition included no rehearsal of pseudowords, the stimuli could not actually match memoranda maintained in pSTM. Therefore, there is no match and mismatch in the same sense as in the rehearsal task. However, in each trial there were two kinds of stimuli, the beginnings of which were phonologically identical (i.e., a stimulus that, in the context of rehearsal, would have had a matching beginning and ending or a matching beginning and mismatching ending; this design was necessary to study the third syllable), which might contribute to the effect. We can only speculate why responses to the two stimulus types differed in the control condition, but one possibility is that the presence of these two pseudowords with phonologically identical beginnings in the same trial interacted with attentional control. A previous study by Engell et al. (2016) has shown that sounds preceded by maskers with similar frequencies resulted in more reduced activation

when participants attended to the auditory modality compared to when they attended to the visual modality. Perhaps, then, if our participants did not properly inhibit auditory stimuli that were irrelevant to the control task, repetition of phonologically identical pseudoword beginnings in close succession may have caught their attention, which in turn may have modulated their responses.

## CONCLUSION

The comparison of MEG responses in individuals with better or poorer pSTM suggested that pSTM capacity affected the ability to maintain pseudowords in phonological memory during interference, as reflected in PMN responses. Specifically, the maintenance of the third syllables but not the first syllables differentiated between poorer and better pSTM groups. It seems that tri-syllabic words challenge pSTM and, therefore, PMN responses to these longer words can reveal differences in pSTM capacity. We also found that those with better pSTM and stronger third-syllable responses learned words faster in a paired-associate word learning task, suggesting a link between pSTM maintenance (or encoding and maintenance) and language learning. This might be related to use of a phonological code in the maintenance of spoken word forms and their phoneme order in pSTM.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of Helsinki University Central Hospital. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SY and ES designed the research. SY and AN performed the research. SY, AN, and ES analyzed the data and modified the manuscript. SY wrote the original draft of the manuscript.

## FUNDING

This work was supported by the Academy of Finland (grant numbers 131963 and 274058) and the Emil Aaltonen Foundation.

## ACKNOWLEDGMENTS

The authors thank Mr. Tommi Makkonen for technical assistance.

## REFERENCES

- Assaneo, M. F., Ripolles, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., and Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat. Neurosci.* 22, 627–632. doi: 10.1038/s41593-019-0353-z
- Atkins, P. W. B., and Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Appl. Psycholinguist.* 19, 537–552.
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon Press.
- Baddeley, A. D., and Hitch, G. J. (1974). “Working memory,” in *The Psychology of Learning and Motivation*, ed. G. A. Bower (New York: Academic Press), 47–89.
- Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi: 10.1037/0033-295x.105.1.158
- Baldo, J. V., Katseff, S., and Dronkers, N. F. (2012). Brain regions underlying repetition and auditory-verbal short-term memory deficits in aphasia: evidence from voxel-based lesion symptom mapping. *Aphasiology* 26, 338–354. doi: 10.1080/02687038.2011.602391
- Barry, J. G., Hardiman, M. J., and Bishop, D. V. (2009). Mismatch response to polysyllabic nonwords: a neurophysiological signature of language learning capacity. *PLoS One* 4:e6270. doi: 10.1371/journal.pone.0006270
- Barry, J. G., Sabisch, B., Friederici, A. D., and Brauer, J. (2011). Encoding: the keystone to efficient functioning of verbal short-term memory. *Neuropsychologia* 49, 3636–3647. doi: 10.1016/j.neuropsychologia.2011.09.018
- Boersma, P., and Weenink, D. (2008). *Praat: Doing phonetics by Computer [Computer Program]*. Version 5.0.40. Available online at: <http://www.praat.org/> (accessed November 17, 2008).
- Bonte, M. L., and Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Brain Res. Cogn. Brain Res.* 21, 360–376. doi: 10.1016/j.cogbrainres.2004.06.010
- Buchsbaum, B. R., and D’Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778. doi: 10.1162/jocn.2008.20501
- Buchsbaum, B. R., Hickok, G., and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* 25, 663–678.
- Buchsbaum, B. R., Olsen, R. K., Koch, P., and Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48, 687–697. doi: 10.1016/j.neuron.2005.09.029
- Camos, V., and Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: the executive and phonological loops. *Front. Hum. Neurosci.* 8:900. doi: 10.3389/fnhum.2014.00900
- Camos, V., Lagner, P. L., and Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *J. Mem. Lang.* 61, 457–469.
- Čeponiene, R., Service, E., Kurjenluoma, S., Cheour, M., and Näätänen, R. (1999). Children’s performance on pseudoword repetition depends on auditory trace quality: evidence from event-related potentials. *Dev. Psychol.* 35, 709–720. doi: 10.1037/0012-1649.35.3.709
- Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., and Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2653–2658. doi: 10.1073/pnas.1216827110
- Connolly, J. F., and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266. doi: 10.1162/jocn.1994.6.3.256
- Connolly, J. F., Service, E., D’Arcy, R. C., Kujala, A., and Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Neuroreport* 12, 237–243. doi: 10.1097/00001756-200102120-00012
- D’Arcy, R. C., Connolly, J. F., Service, E., Hawco, C. S., and Houlihan, M. E. (2004). Separating phonological and semantic processing in auditory sentence processing: a high-resolution event-related brain potential study. *Hum. Brain Mapp.* 22, 40–51. doi: 10.1002/hbm.20008
- 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
- Davidow, J. H., Bothe, A. K., Richardson, J. D., and Andreatta, R. D. (2010). Systematic studies of modified vocalization: effects of speech rate and instatement style during metronome stimulation. *J. Speech Lang. Hear. Res.* 53, 1579–1594. doi: 10.1044/1092-4388(2010/09-0173)
- Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111
- Davis, M. H., Di Betta, A. M., Macdonald, M. J., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *J. Cogn. Neurosci.* 21, 803–820. doi: 10.1162/jocn.2009.21059
- Desroches, A. S., Newman, R. L., Robertson, E. K., and Joanisse, M. F. (2013). Electrophysiological indices of phonological impairments in dyslexia. *J. Speech Lang. Hear. Res.* 56, 250–264. doi: 10.1044/1092-4388(2012/10-0351)
- Engell, A., Junghofer, M., Stein, A., Lau, P., Wunderlich, R., Wollbrink, A., et al. (2016). Modulatory effects of attention on lateral inhibition in the human auditory cortex. *PLoS One* 11:e0149933. doi: 10.1371/journal.pone.0149933
- French, L. M. (2006). *Phonological Working Memory and L2 Acquisition: A Developmental Study of Quebec Francophone Children Learning ENGLISH*. New York, NY: Edwin Mellen Press.
- French, L. M., and O’Brien, I. (2008). Phonological memory and children’s second language grammar learning. *Appl. Psycholinguist.* 29, 463–487.
- Gathercole, S. E., and Baddeley, A. D. (1993). *Working Memory and Language*. New York, NY: Psychology Press.
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A., and Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Appl. Cogn. Psychol.* 13, 65–77.
- Gupta, P., and Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *J. Mem. Lang.* 61, 481–502.
- Herman, A. B., Houde, J. F., Vinogradov, S., and Nagarajan, S. S. (2013). Parsing the phonological loop: activation timing in the dorsal speech stream determines accuracy in speech reproduction. *J. Neurosci.* 33, 5439–5453. doi: 10.1523/JNEUROSCI.1472-12.2013
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J. Cogn. Neurosci.* 15, 673–682. doi: 10.1162/089892903322307393
- Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 14, 1125–1138. doi: 10.1162/089892902760807140
- Jacquemot, C., and Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* 10, 480–486. doi: 10.1016/j.tics.2006.09.002
- Joseph, S., Iverson, P., Manohar, S., Fox, Z., Scott, S. K., and Husain, M. (2015). Precision of working memory for speech sounds. *Q. J. Exp. Psychol.* 68, 2022–2040. doi: 10.1080/17470218.2014.1002799
- Kauramäki, J., Jääskeläinen, I. P., Hari, R., Möttönen, R., Rauschecker, J. P., and Sams, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *J. Neurosci.* 30, 1314–1321. doi: 10.1523/JNEUROSCI.1950-09.2010
- Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., and Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Brain Res. Cogn. Brain Res.* 21, 106–113. doi: 10.1016/j.cogbrainres.2004.05.011
- Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., and Griffiths, T. D. (2016). A brain system for auditory working memory. *J. Neurosci.* 36, 4492–4505. doi: 10.1523/JNEUROSCI.4341-14.2016
- 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
- Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Front. Hum. Neurosci.* 7:357. doi: 10.3389/fnhum.2013.00357

- Malins, J. G., Desroches, A. S., Robertson, E. K., Newman, R. L., Archibald, L. M., and Joanisse, M. F. (2013). ERPs reveal the temporal dynamics of auditory word recognition in specific language impairment. *Dev. Cogn. Neurosci.* 5, 134–148. doi: 10.1016/j.dcn.2013.02.005
- McGettigan, C., Warren, J. E., Eisner, F., Marshall, C. R., Shanmugalingam, P., and Scott, S. K. (2011). Neural correlates of sublexical processing in phonological working memory. *J. Cogn. Neurosci.* 23, 961–977. doi: 10.1162/jocn.2010.21491
- Melby-Lervåg, M., Lervåg, A., Lyster, S. A., Klem, M., Hagtvet, B., and Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychol. Sci.* 23, 1092–1098. doi: 10.1177/0956797612443833
- Newman, R. L., and Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biol. Psychol.* 80, 114–121. doi: 10.1016/j.biopsycho.2008.04.008
- Nora, A., Renvall, H., Kim, J. Y., Service, E., and Salmelin, R. (2015). Distinct effects of memory retrieval and articulatory preparation when learning and accessing new word forms. *PLoS One* 10:e0126652. doi: 10.1371/journal.pone.0126652
- Numminen, H., Service, E., and Ruoppila, I. (2002). Working memory, intelligence and knowledge base in adult persons with intellectual disability. *Res. Dev. Disabil.* 23, 105–118. doi: 10.1016/s0891-4222(02)00089-6
- Numminen, J., and Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci. Lett.* 272, 29–32. doi: https://doi.org/10.1016/s0304-3940(99)00573-x
- Papagno, C., Comi, A., Riva, M., Bizzi, A., Vernice, M., Casarotti, A., et al. (2017). Mapping the brain network of the phonological loop. *Hum. Brain Mapp.* 38, 3011–3024. doi: 10.1002/hbm.23569
- Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature* 362, 342–345. doi: 10.1038/362342a0
- Richardson, F. M., Ramsden, S., Ellis, C., Burnett, S., Megnin, O., Catmur, C., et al. (2011). Auditory short-term memory capacity correlates with gray matter density in the left posterior STS in cognitively normal and dyslexic adults. *J. Cogn. Neurosci.* 23, 3746–3756. doi: https://doi.org/10.1162/jocn\_a\_00060
- Savill, N. J., Cornelissen, P., Pahor, A., and Jefferies, E. (2019). rTMS evidence for a dissociation in short-term memory for spoken words and nonwords. *Cortex* 112, 5–22. doi: 10.1016/j.cortex.2018.07.021
- Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107. doi: 10.1016/S0166-2236(02)00037-1
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *Q. J. Exp. Psychol. A* 45, 21–50. doi: 10.1080/14640749208401314
- Service, E. (2013). “Working memory in second language acquisition: phonological short-term,” in *The Encyclopedia of Applied Linguistics*, ed. C. A. Chapelle (Hoboken, NJ: Wiley Blackwell), doi: 10.1002/9781405198431.wbeal1287
- Service, E., and Kohonen, V. (1995). Is the relation between phonological memory and foreign-language learning accounted for by vocabulary acquisition? *Appl. Psycholinguist.* 16, 155–172.
- Service, E., and Maury, S. (2003). All parts of an item are not equal: effects of phonological redundancy on immediate recall. *Mem. Cognit.* 31, 273–284. doi: 10.3758/bf03194386
- Service, E., Maury, S., and Luotoniemi, E. (2007). Individual differences in phonological learning and verbal STM span. *Mem. Cognit.* 35, 1122–1135. doi: 10.3758/bf03193483
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* 43, 482–489. doi: 10.1037/h0055479
- Tian, X., and Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychol.* 1:166. doi: 10.3389/fpsyg.2010.00166
- Tian, X., and Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Front. Hum. Neurosci.* 6:314. doi: 10.3389/fnhum.2012.00314
- von Holst, E., and Mittelstaedt, H. (1950). Das Reafferenzprinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften* 37, 464–476.
- Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huottilainen, M., Shtyrov, Y., et al. (2015). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cereb. Cortex* 25, 1576–1586. doi: 10.1093/cercor/bht351

**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 Ylinen, Nora and Service. 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](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## 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