Journal of Memory and Language 140 (2025) 104573

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

An embedded computational framework of memory: Accounting for the influence of semantic information in verbal short-term memory

Dominic Guitard [a,*], Jean Saint-Aubin [b], J. Nick Reid [c], Randall K. Jamieson [d]

[a] Cardiff University, Cardiff, Wales, United Kingdom
[b] Université de Moncton, Moncton, NB, Canada
[c] University of Northern British Columbia, Prince George, BC, Canada
[d] University of Manitoba, Winnipeg, MB, Canada

ARTICLE INFO

Keywords:
Semantic relatedness
Immediate serial recall
Order reconstruction
Distributional Semantic models
Instance theory
DRM false recall

ABSTRACT

We introduce the Embedded Computational Framework of Memory (eCFM), a model that integrates structured semantic word representations with an instance-based memory model to account for the influence of semantic information in verbal short-term memory. The eCFM combines principles from the episodic MINERVA 2 model and the semantic Latent Semantic Analysis model. After reviewing how semantic information impacts verbal short-term memory performance, we demonstrate eCFM's ability to reconcile various phenomena within a common computational framework. Our model captures key findings, such as the influence of semantic information in serial recall, its reduction in serial reconstruction, and the impact of task difficulty on semantic information. In five experiments, we tested predictions derived from the eCFM. Experiments 1 and 2 manipulated list organization, with Experiment 1 using alternating lists of related or unrelated words and Experiment 2 using blocked lists. Experiment 3 varied presentation rates, Experiment 4 revisited the detrimental effect of semantic information on order information, and Experiment 5 explored false recall. We found that semantic information interacts with list composition, presentation rate affects the magnitude of its influence, and semantic information impacts order information contrary to the dominant view. Additionally, increasing the number of related study words to a non-studied semantic lure boosts false recall. The eCFM captured these findings as well as memory at the item level. Our demonstration provides insight into the cognitive mechanisms underlying verbal short-term memory and the interplay of semantic and episodic memory processes in recall.

Introduction

In verbal short-term memory tasks, participants encode and recall a series of words in the order they were studied. Much prior research indicates that participants' prior linguistic experience influences performance in this task (see Oberauer et al., 2018 for a review). One significant demonstration of this influence is the effect of semantic information. Semantic information refers to knowledge about verbal information accumulated over the lifespan, enabling people to understand word meaning, make associations, and categorize information. The influence of semantic information on verbal short-term memory performance has yielded a large number of robust findings that have been challenging to reconcile under a common framework (Kowialiewski et al., 2023; Poirier & Saint-Aubin, 1995; Neath et al., 2022; Neale & Tehan, 2007; Tehan, 2010; Saint-Aubin & Poirier, 1999a,b; Saint-Aubin et al., 2005; 2014). In this investigation, we address this challenge by accounting for the influence of semantic information on verbal short-term memory using a new computational model called the Embedded Computational Framework of Memory (eCFM).

The eCFM is a memory model with a large lexicon (50,000+ words) that bridges the gap between episodic and semantic memory models by integrating word representations from the Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) model of semantic memory into the MINERVA 2 (Hintzman, 1986) model of episodic memory. The aim of this study is to demonstrate that by combining these models, we can explain a range of key phenomena concerning how semantic information influences verbal memory at both the overall and item levels—a challenge that has eluded traditional computational memory theories, which often rely on arbitrary vectors to represent words. Additionally, we used eCFM to make novel, testable predictions that are evaluated via

* Corresponding author at: School of Psychology, Cardiff University, Tower Building, 70 Park Place, Cardiff CF10 3AT, United Kingdom.
E-mail address: guitardd@cardiff.ac.uk (D. Guitard).

https://doi.org/10.1016/j.jml.2024.104573
Received 14 August 2023; Received in revised form 13 September 2024; Accepted 26 September 2024
Available online 4 October 2024

empirical investigation.

Semantic relatedness

In recent years, semantic information has played an important role in understanding the interactions between people's lexicon (i.e., the corpus of word representations) and short-term memory performance (e.g., Ishiguro & Saito, 2021; Kowialiewski & Majerus, 2020; Kowialiewski et al., 2021, 2023; Neath et al., 2022). A classic empirical benchmark on the influence of semantic information on verbal short-term memory performance is the semantic relatedness effect, that is, superior immediate serial recall for lists of related words compared to unrelated words. For example, when participants are tasked with recalling a sequence of words in their presentation order, they remember those words better when they are related (e.g., STEEL, COPPER, BRASS…) than when they are unrelated (e.g., STEEL, MAGAZINE, SERGEANT…). The findings are robust and have been observed across both laboratories and languages (e.g., Kowialiewski & Majerus, 2020) with both young and older adults (Neale & Tehan, 2007), providing a critical basis for evaluating and constraining models of memory.

Operationalization. Over the years, semantic relatedness has been operationalized differently: (a) category membership (e.g., fruits: apple, orange, banana; Murdock & Vom Saal, 1967), (b) association based on co-occurrence in language corpora (e.g., APPLE, TREE; Tse, 2009), (c) meaning (e.g., GOOD, NICE; Crowder, 1979), and (d) introspective empirical judgments of word properties like valence, arousal, and dominance (Ishiguro & Saito, 2021, 2024). Despite the diverse operationalizations, a recent study by Neath et al. (2022) revealed a strong and consistent pattern of serial recall performance for semantically related over semantically unrelated word lists, whether semantic relatedness was defined by category membership, association, or meaning. In another innovative study, Ishiguro and Saito (2021) defined semantic relatedness based on people's judgments of word valence, arousal, and dominance, using a variant of the method for deriving meaning in a high-dimensional vector space for word representations inspired by Osgood et al. (1957). However, empirical evidence has accumulated that challenges the predictive validity of this operationalization for word relatedness (see Kowialiewski et al., 2023; Ishiguro & Saito, 2024; Sonier et al., 2024). Therefore, at the current moment, it appears that a successful computational model is needed to explain the influence of semantic relatedness across these three well-established operationalizations (i.e., category, association, and meaning) in immediate serial recall.

Across these various operationalizations, the influence of semantic information has produced a rich set of empirical findings that are challenging to reconcile under a common framework. In the following sections, we briefly review these key findings that we attempt to reconcile with our model.

Semantic relatedness affects immediate serial recall but not immediate serial reconstruction of order

It is well established that participants' immediate serial recall performance is superior for semantically related relative to semantically unrelated lists of words. However, when faced with a serial reconstruction task, where participants study a list of words and then are tested on their ability to put them back in order at test, participants perform equally well at reconstructing the presentation order of words from semantically related as semantically unrelated lists (see Neath et al., 2022, for a review). Consistent with that fact, Neath et al. (2022) demonstrated the difference across three experiments when semantic relatedness was defined by category membership (Experiment 1), association (Experiment 2), or meaning (Experiment 3); a difference to results where stimulus similarity is based on appearance and phonology that typically hinders rather than benefits recall (Guitard & Cowan, 2020; Logie et al., 2016; Poirier et al., 2007; Conrad & Hull, 1964; Roodenrys et al., 2022). Naturally, this puzzling dissociation between the influence of semantic similarity in serial recall and serial reconstruction has been of great interest to the research community and has provided important insights into the impact of semantic information on verbal short-term memory.

Redintegration hypothesis. The *redintegration hypothesis* provides a key account on the influence of semantic similarity in verbal short-term memory (e.g., Lewandowsky, 1999; Neath et al., 2022; Neale & Tehan, 2007; Saint-Aubin & Poirier, 1999a,b; Saint-Aubin et al., 2005). This hypothesis, originally derived from Crowder's (1979) sophisticated guessing hypothesis, suggests that at recall, phonological representations are degraded and must undergo a reconstruction process based on long-term knowledge of the memoranda (Hulme et al., 1991; Schweickert, 1993). By corollary, semantic relatedness increases the availability of the appropriate long-term representations in the individual lexicon by activating recall candidates differentially in memory, thus restricting the pool of recall candidates (e.g., APPLE, BANANA, and so forth for a list of *fruits*). With a pool of activated and thus accessible recall candidates, the probability of a successful recall is higher and performance at the immediate serial recall task improves. However, in a serial reconstruction task, semantic relatedness does not provide a recall advantage because the items are already provided to the participant at test, alleviating the need to recall the words but requiring them to put the words back into studied order. Although the redintegration hypothesis provides a useful overarching framework for understanding why there is a serial recall advantage for related over unrelated word lists without a corresponding advantage in serial reconstruction, it leaves the mechanistic details ambiguous. In the description of our model, we will address this issue and provide a clear computational expression to resolve these ambiguities.

Present Study. In the present study, to explore the intriguing dissociation between immediate serial recall and immediate serial reconstruction, we simulated the results of Neath et al. (2022) across the three different operationalizations of semantic relatedness using their materials (see Demonstration 1).

Semantic relatedness and intralist error

In this section, we focus on the impact of semantic similarity on order recall. To do so, we define order information as the serial position that a word was presented so we can measure intralist errors where a studied word is recalled but in the wrong serial position (e.g., a word studied in serial position 5 but recalled in serial position 3). Understanding what constrains order information is often challenging. Therefore, we will divide this section into two parts, focusing on the influence of semantically related information on intralist errors in immediate serial recall and examining how semantically related information can affect specific patterns of migration errors (e.g., a word in position 5 being recalled in error as having appeared earlier rather than later in the list).

Serial Recall and Intralist Error. The prevailing view in the short-term memory literature is that semantically related information primarily affects item information (the ability to recall which words were presented) with limited influence on order information (the ability to recall when each word was presented). This partially explains why the influence of semantic relatedness is limited in serial reconstruction tasks that alleviate participants' responsibility to recall the items themselves (e.g., Neath, 1997; Guitard et al., 2021, 2022; Guitard & Cowan, 2023). However, serial recall tasks that necessitate both item (information about the words) and order information (information about the position of the words in the list) for optimal performance have yield mixed findings (Neath et al., 2022; Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011). A recent review by Neath et al. (2022) suggests that most experimental results favor a beneficial effect on item information, with a small or null effect on order information. Despite the elegance of Neath et al.'s work, there are notable exceptions that are difficult to dismiss. For example, Saint-Aubin et al. (2005) conducted one of the largest experiments on semantic relatedness defined by category membership in French ($n = 252$) and observed a detrimental effect of semantic relatedness on order memory, despite controlling for the number of items recalled between related and unrelated lists. In Saint-Aubin et al.'s (2005) experiment, semantic relatedness was manipulated between participants. Additionally, the same detrimental effect on order was observed in participants performing concurrent articulation (repeating aloud the word "mathématiques" at a rate of about three repetitions every 2 s). Echoing their results, Tse (2009; Tse et al., 2011) found a detrimental effect on order information in two studies using different manipulations of semantic relatedness (i.e., category membership and association) despite controlling for the number of items recalled. It is also worth highlighting that previous studies with fewer participants have often also found a trend for a small detrimental effect on order (e.g., Guérard & Saint-Aubin, 2012; Neale & Tehan, 2007; Murdock, 1976; Saint-Aubin & Poirier, 1999a,b). This mixed evidence makes the situation suboptimal for evaluating the models and leaves it unclear whether these phenomena are based on specific stimulus characteristics; thus, requiring further empirical evidence.

Migration Error. Another important phenomenon concerning the influence of semantic relatedness is that semantic information can exert an influence on order recall by affecting the pattern of order errors (i.e., the pattern serial recall position errors in which a specific word is most likely to be incorrectly recalled). For example, in their landmark study, Poirier et al. (2015) strategically positioned semantically related words in the first three serial positions of a study list (e.g., CANARY, MUSTARD, BANANA) and a word in the 5th serial position related to the initial triplet (e.g., YELLOW) in experimental lists or another unrelated word (e.g., JUNGLE) in control lists. They found that participants were more likely to recall the 5th word earlier in the sequence compared to its counterpart in the control list, indicating an important influence of semantic association on memory retrieval. This pattern of "semantic migration" has been of interest in recent computational investigations (see Kowialiewski et al., 2021 for a review and model) because it represents one of the most important finding on the influence of semantic information on order information. The results have been reinforced and extended with different list organizations, including alternate and mixed lists by Kowialiewski et al. (2024).

Summary. Overall, in most studies, semantic information has a more significant influence on item information relative to order information. However, semantic information has affected order information in serial recall (e.g., Tse, 2009; Tse et al., 2011), notably in Saint-Aubin et al. (2005) with a between-participants design using a large sample size. Also, order information can be constrained by semantic information, affecting the serial position in which related words are erroneously recalled (e.g., Kowialiewski et al., 2021, 2024; Poirier et al., 2015).

Present Study. In the present study, we simulate Saint-Aubin et al.'s (2005) results (see Demonstration 2). Given that our eCFM model includes a lexicon, we demonstrate that it can track people's performance. This approach allows us to determine whether the model accurately reflects the patterns observed in Saint-Aubin et al.'s specific memoranda and experimental design. We will also illustrate how different lexicons derived from different models of semantic memory can be interchanged within the model. Regarding migration error, we will simulate the benchmark findings of Poirier et al. (2015) that have already been subject to computational investigation (Kowialiewski et al., 2021) to investigate and understand the impact of semantic information on order information (see Demonstration 3). Empirically, to further contribute to this important debate, in Experiment 4, we will reassess the influence of semantic information on intralist error patterns in serial recall with a large sample of 160 participants and a new set of stimuli (see Experiment 4 and Demonstration 7).

Semantic relatedness and task difficulty

Manipulating task difficulty is of great interest in relation to the redintegration hypothesis, as it allows testing of key predictions. For example, the hypothesis suggests that the magnitude of the semantic relatedness effect—the difference in memory performance between related and unrelated lists—should be more pronounced when encoding is poor. This is because the more degraded representations have a greater opportunity to benefit from redintegration due to semantic information, compared to when encoding is nearly perfect and redintegration is less crucial (Schweickert, 1993).

Consistent with the redintegration hypothesis, challenging people's encoding of study words by concurrent articulation (i.e., participants repeating an irrelevant word like 'mathematic' during study), the influence of semantic relatedness increases (Neale & Tehan, 2007; Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999a,b; Saint-Aubin et al., 2005). Similarly, Neale and Tehan (2007) demonstrated that manipulating list length and the delay between study and recall yields a memorial advantage for words from related compared to unrelated word lists increases linearly with task difficulty. These insightful findings underscore the importance of redintegration when memoranda are poorly encoded.

Present Study. To address these facts, we report simulations that apply the eCFM to the materials and conditions in Saint-Aubin et al. (2005), both with and without concurrent articulation (see Demonstration 2). To ensure that our results are not confined to a specific simulation, we conducted another test in Experiment 3 (see Experiment 3 and Demonstration 6) by manipulating the presentation rate—a factor well-established to affect task difficulty (e.g., Coltheart & Langdon, 1998; Dauphinee et al., 2024; Guitard & Cowan, 2023).

Semantic relatedness and list structure

It is becoming increasingly clear that a range of factors influence semantic relatedness. One such factor, which has received growing interest, is the influence of list structure and organization—how we arrange related and unrelated information within a list. List organization is a crucial experimental manipulation in the study of short-term memory (e.g., Saint-Aubin et al., 2021). Specifically, when semantically related information is presented in adjacent serial positions (e.g., words in positions 3 and 4), memory performance is typically superior compared to when that information is spaced apart. This phenomenon was first reported by Saint-Aubin et al. (2014), who observed improved recall performance for pairs of related and unrelated words presented either at adjacent positions or separated by one or two unrelated items, with better recall noted when the paired members were adjacent (see also Brooks & Watkins, 1990).

These findings were further replicated and expanded upon by Kowialiewski and Majerus (2020; Kowialiewski, et al., 2021, 2022, 2024) who explored list organization with two semantic categories either blocked together (e.g., AAABBB, where 'A' represents one semantic category, and 'B' represents a different semantic category) or interleaved (e.g., ABABAB, where 'A' and 'B' are different semantic categories). They also tested configurations with related and unrelated triads (RRRUUU, UUURRR; where 'U' = unrelated and 'R' = related), reporting results consistent with Saint-Aubin et al. (2014).

Present Study. Given the importance of list organization, in our study we aimed to expand on these findings using word lists that incorporate words from related and unrelated categories either interleaved in Experiment 1 or blocked in Experiment 2. This approach will serve as the empirical basis to test our model's capability of capturing the influence of list organization on the influence of semantic information (see Experiments 1 and 2 and Demonstrations 4 and 5).

Semantic relatedness and false memory

The last empirical finding we will examine is the influence of semantic information on false memory. Human memory is reconstructive by nature, and therefore, inherently flawed. One of the most compelling demonstrations of the impact of semantic information on verbal short-term memory is the phenomenon of false memory, where individuals erroneously recall or recognize un-studied yet semantically related verbal information. Researchers often investigate this type of memory error using the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In this paradigm, participants study a list of words such as BED, NAP, SNORE, AWAKE, SNOOZE, TIRED, WAKE, REST, BLANKET, and DREAM that are all semantically related to an unpresented but implied critical lure such as SLEEP. Participants are then asked to recognize or recall, or both, the presented items. False memory is particularly intriguing in this context because it allows us to directly evaluate, at the item level, if the model can predict specific recalls of an unstudied item in the participants' lexicon, an analysis that is impossible without word-specific lexical representations.

One of the first studies to investigate the influence of false recall in immediate and delayed serial recall was conducted by Tehan (2010). Using the DRM paradigm, they found that participants recalled critical related lures even after just a few seconds. This phenomenon and the ability to make item-level predictions are of great interest for computational models such as the generative model of memory construction and consolidation by Spens and Burgess (2024) that initially succeeded in accounting for one of the classic finding in false memory literature in recognition and free recall: the likelihood of recalling the critical lure increases with the number of related study words (Robinson & Roediger, 1997). However, this classical finding has not received the same level of attention in serial recall. We aimed to overcome this limitation with a new demonstration of false recall of critical lures in serial recall.

Present Study. To investigate the issue, we conducted an experiment to measure false recall of a critical lure in the context of serial recall. Mirroring Spens and Burgess (2024; see also Robinson & Roediger (1997), we manipulated the number of related words associated with a specific non-presented critical lure and tracked the serial position in which the critical lure was falsely recalled. This experimental and computational demonstration provides a different look at the classic DRM phenomena and will serve as an articulate basis for model evaluation (see Experiment 5 and Demonstration 8).

Summary semantic relatedness

In this section, we briefly summarize the key empirical findings that we focus on throughout the remainder of this manuscript and lay the groundwork for evaluating the eCFM model, highlighting the value of combining episodic and semantic models of memory. The primary goal of this manuscript is to demonstrate how the influence of semantic information on verbal memory performance can be explained. To achieve this, we systematically investigated several key phenomena: 1) the semantic relatedness paradox, which highlights the beneficial effect of semantically related words in serial recall and its absence in serial reconstruction; 2) the impact of semantic relatedness on intralist error, whether semantic relatedness has a detrimental effect on order information; 3) the influence of semantic relatedness on migration errors; 4) the effect of task difficulty on the magnitude of the semantic relatedness; 5) the role of list organization on the semantic relatedness effect; and, 6) the influence of semantic relatedness on false memory in serial recall.

Modelling background

Over the years, numerous models have been developed to enhance our understanding of recall (e.g., Brown et al., 2007; Burgess & Hitch, 1999, 2006; Farrell & Lewandowsky, 2002; Franklin & Mewhort, 2002, 2015; Henson, 1998; Hurlstone et al., 2014; Kowialiewski, et al., 2021; Nairne, 1988; Murdock, 1974, 1982, 1993; Oberauer et al., 2012; Oberauer & Lewandowsky, 2011; Page & Norris, 1998; Raaijmakers & Shiffrin, 1980, 1981; Saint-Aubin et al., 2021). These models have provided clear theoretical frameworks and successfully captured key processes linked to specific human behaviors. A common technique in these models involves using randomly generated vectors to represent studied items to simulate a lexicon and tame the volatility of predictions conditional on specific memoranda presented in a target experiment. However, while this method provides a convenient basis for knowledge representation in semantic memory, it falls short of representing the semantic relationships people share (Johns & Jones, 2010; Osth et al., 2020). To address the issue, we represent words in a structured mental lexicon, derived from LSA (Landauer & Dumais, 1997), an established theory of semantic memory. To simulate recall with those representations, we embed those representations within an established model of episodic and memory. Naturally, we are not the first to adopt structured representations in a computational model of cognition and so this work extends on a larger disciplinary effort (Chang & Johns, 2023; Criss & Shiffrin, 2004, 2005; Franklin & Mewhort, 2002, 2015; Johns et al., 2012; Kimball et al. 2007; Nosofsky et al., 2018a, 2018b; Mewhort et al., 2018; Lewandowsky & Murdock, 1989; Osth et al., 2020; Osth & Zhang, 2023; Reid & Jamieson, 2023; Reid et al., 2023; Sirotin et al., 2005).

For example, Sirotin et al. (2005) and Kimball et al. (2007) have represented pre-experimental, pairwise semantic associations by incorporating similarity scores obtained from latent semantic analysis (LSA; Landauer & Dumais, 1997) and word association space (WAS; Steyvers, Shiffrin, & Nelson, 2005) into Raaijmakers and Shiffrin's eSAM and fSAM models (1980, 1981). In those demonstrations, the mental lexicon was artificially constrained to suit the characteristics of the task, rather than having a complete lexicon (e.g., representations for a subset of 750 words). Nevertheless, even with a small lexicon, those models accounted for an impressive range of phenomena in free recall. Overcoming some of those limitations, Mewhort et al. (2018) used a large lexicon equipped with 39,076 words with BEAGLE vectors representing word meanings within a holographic model for recall to account for the Hebb effect, the von Restorff effect, and the release of proactive interference.

To build on this tradition, we developed the eCFM by linking a classic instance-based model of episodic memory, MINERVA 2 (Hintzman, 1986), that provides a sound and articulate account of the redintegration hypothesis that has served as an excellent but informal account of how semantic information influences performance in verbal short-term memory and that, if successful, would link our account to a wider range of work in which MINERVA 2 has already been successfully applied to problems including frequency judgement, recognition, categorization, cued recall, implicit learning, associative learning, and heuristic decision making (e.g., see Jamieson et al., 2022).

Embedded Computational Framework of Memory

The eCFM is a memory model with a large embedded lexicon of vector-based word representations derived from the LSA model of semantic memory (50,797 words in the simulations reported here). This model is an extension of Cowan's embedded processes model of information processing, in which short-term memory consists of an activated subset of long-term memory (Cowan, 1988, 2019; Cowan et al., 2024). It bridges the gap between models of episodic memory (MINERVA 2: Hintzman, 1986) and semantic memory (LSA: Landauer & Dumais, 1997).

By integrating word representations derived from a LSA, eCFM offers a clear definition of semantic relations grounded in well-established semantic memory theory. This specificity is an important step forward from models that rely on randomly generated vectors to represent words that often lack the nuanced understanding of semantic relationships (Johns & Jones, 2010). This approach allows us to progress towards a more comprehensive model of memory that integrates advances from the research siloes of semantic and episodic memory and that supports memory modelling for the same word lists presented in experiments and for the words that people recall, correctly or incorrectly, in those experiments.

Lexical Representations. A key aspect in our model is the lexicon (or semantic memory). Accounting for the precise impact of semantic relatedness in serial recall requires a memorial representation of the study list in which words are represented both in terms of their meaning and the order in which they are studied. To accomplish that goal, we represented a lexicon within the eCFM by borrowing Landauer and Dumais' (1997) theory of Latent Semantic Analysis (LSA), one of several established Distributional Semantic Models (DSMs). We chose to use LSA in the simulation work that follows because it represents a familiar framework for readers, but we could have used other models (e.g., Bojanowski et al., 2017; Jamieson et al., 2018; Jones & Mewhort, 2007; Mikolov et al., 2013; Mitchell & Lapata, 2010) or methods (e.g., De Deyne et al., 2019; Just et al., 2010; McRae et al., 2005).

DSMs construct representations of word meaning by "reading" a record of language (i.e., a corpus) and deriving vector-based word representations based on how those words are used (Jones, 2019; Lenci, 2018; Reid & Katz, 2018). Because DSMs represent word meaning based on how words are used, they implement Firth's (1957) dictum that, "You shall know the meaning of a word by the company it keeps" (see also Harris, 1954, and Rubenstein & Goodenough, 1965).

LSA is a benchmark DSM that derives vector-based word representations by counting the number of times that words appear in different documents over a large corpus of language. The counts are recorded in a word-by-document matrix and both local and global weighting procedures are applied to the counts. For the local weighting, each element in the matrix (i.e., each word-by-document frequency count) is converted to its natural logarithm. For the global weighting, each row of the matrix is weighted according to the word's entropy across the documents (see Berry & Browne, 1999; Martin & Berry, 2007). This is referred to as a "global weight" because the same weight is applied to the entire row in the word-by-document matrix (i.e., each of the word's frequency counts is weighted in the same way). Entropy is an information-theoretic measure that can assess the informativeness of a word's occurrence. For instance, a common word such as BANK occurs in many different contexts, whereas a word such as COGNITION is specific to a few contexts (this word likely appears in documents about *psychology* but not in documents about *sports* or *fashion*). Therefore, a low-entropy word like COGNITION is a more informative word than a high-entropy word like BANK. Ultimately, the entropy weighting emphasizes more informative words in the word-by-document matrix and lessens the influence of less informative words.

After the weighting procedures are applied to the word-by-document matrix, its dimensionality is then reduced using the statistical technique of singular value decomposition. This results in a matrix where each row represents a word's meaning as a vector defined by its weight on each principal component that encodes a latent statistical dimension of word meaning. Although the dimensions of word meaning are psychologically undefined (though see Hollis & Westbury, 2016), this transformation captures the latent structure of word associations and significantly improves the signal-to-noise ratio in the resulting word representations. For example, the initial untreated word-by-document frequency matrix achieves only 15.7 % correct on a synonym judgement test whereas the final matrix after weighting and singular value decomposition improved performance to 52.7 % correct (Landauer, Foltz, & Laham, 1998).

Importantly, the dimension reduction step allows for some words that do not directly co-occur in the same documents (such as synonyms) to become more similar based on their common patterns of use across different documents in the corpus (Landauer & Dumais, 1997). For example, BEAGLE and DOG may become more similar because they are interchangeable in natural language use, even if BEAGLE and DOG themselves do not co-occur in the same document. In this way, the model captures higher-order association that signals semantic correspondence. Once singular value decomposition has been performed, the word vectors stand in as numerical analogues for word representations. Note that the values in the vectors are difficult to interpret on their own (Kintsch, 2000), but when these values are compared to the values in other word vectors, semantic structure of words in the full lexicon emerges, with similar words having correlated patterns of values over the latent dimensions (see Hollis & Westbury, 2016).

Despite the simplicity of the method, LSA vectors predict a range of lexical behaviour including rate of language learning, judgements about word meaning, and lexical judgement (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). Furthermore, using these representations within computational models can account for more complex language phenomena such as causal inferences, homonym disambiguation, metaphor interpretation (Kintsch, 2000), false memory (Reid & Jamieson, 2023), and implicit learning of semantic categories (Chubala et al., 2016). A more complete description and analysis of LSA can be found in the model's flagship theoretical paper (Landauer & Dumais, 1997).

In the following simulations, we use 300-dimension LSA word representations derived from the Touchstone Applied Science Associates Inc. (TASA) corpus as reported in Günther et al.'s 2015 article. However, in one last step we filter Günther et al.'s set of vectors to include only the items that also appear in the SUBTLEXus database (Brysbaert, New, & Keuleers, 2012). Our assumption is that this final lexicon is representative of a typical person's lexicon who participates in the memory experiments that we model in the remainder of this manuscript.

Serial position representation. To model serial recall, we assume that participants encode words at their studied serial positions and then use these positions as cues to recall corresponding words at test (for a related idea, see Nairne, 1990; Saint-Aubin et al., 2021). Serial position information is a critical feature of serial recall that must be incorporated to extend the capabilities of MINERVA 2 (Hintzman, 1986) and bridge the gap between recognition and recall.

Currently, our efforts are focused on demonstrating the importance of embedding a lexicon, which necessitates a method for representing serial positions to effectively account for serial recall performance. To this end, we have adopted methodologies well-established in item-independent context models (e.g., see Logan & Cox, 2023; Osth & Hurlstone, 2023; reviews and models). More exactly, to represent serial positions in a study list we, first, generate a random vector of dimensionality n for the first position. That representation stores an n-dimensional vector of random deviates sampled from a Normal distribution with mean 0 and standard deviation $1/\sqrt{n}$ (see Jones & Mewhort, 2007, and Murdock, 1982, for precedence on that decision). Second, we generate a new vector for each successive serial position by iteratively copying the representation from the preceding serial position and sampling a new random deviate from the same normal distribution to each dimension with probability d. Thus, d denotes the degree of similarity/dissimilarity of successive serial position representations in memory of a study list.

The scheme generates an ordered sequence of vector-based serial position representations whose similarities vary as a function of serial distance. To illustrate, refer to the 4 bottom panels of Fig. 1 that plots the cosine similarity between representations at all serial positions in a six-item list. When $d = 0$, the representation of serial position p is identical to itself, but it is also identical to the representations for all other serial positions, meaning that position information is perfectly confusable, leading to inevitable order errors (retrieving an item in the wrong serial position). Conversely, when $d = 1$, the representation of serial position p is orthogonal, resulting in no order errors. In cases where $0 < d < 1$, the representation of serial position p remains identical to itself but varies in similarity to the representations of other serial positions. The degree of similarity changes in a graded manner as a function of distance. This scheme aligns with the representation of serial positions in other models; however, our computational method for developing the similarity structure of serial position representations differs from the approaches used in these other models (e.g., Brown et al., 2000; Howard & Kahana, 2002).

In sum, when $0 < d < 1$, this representational scheme implements the idea that serial position representations are identical to themselves and vary in similarity (i.e., confusability) with others as a function of distance. As will become apparent from the description for the model of

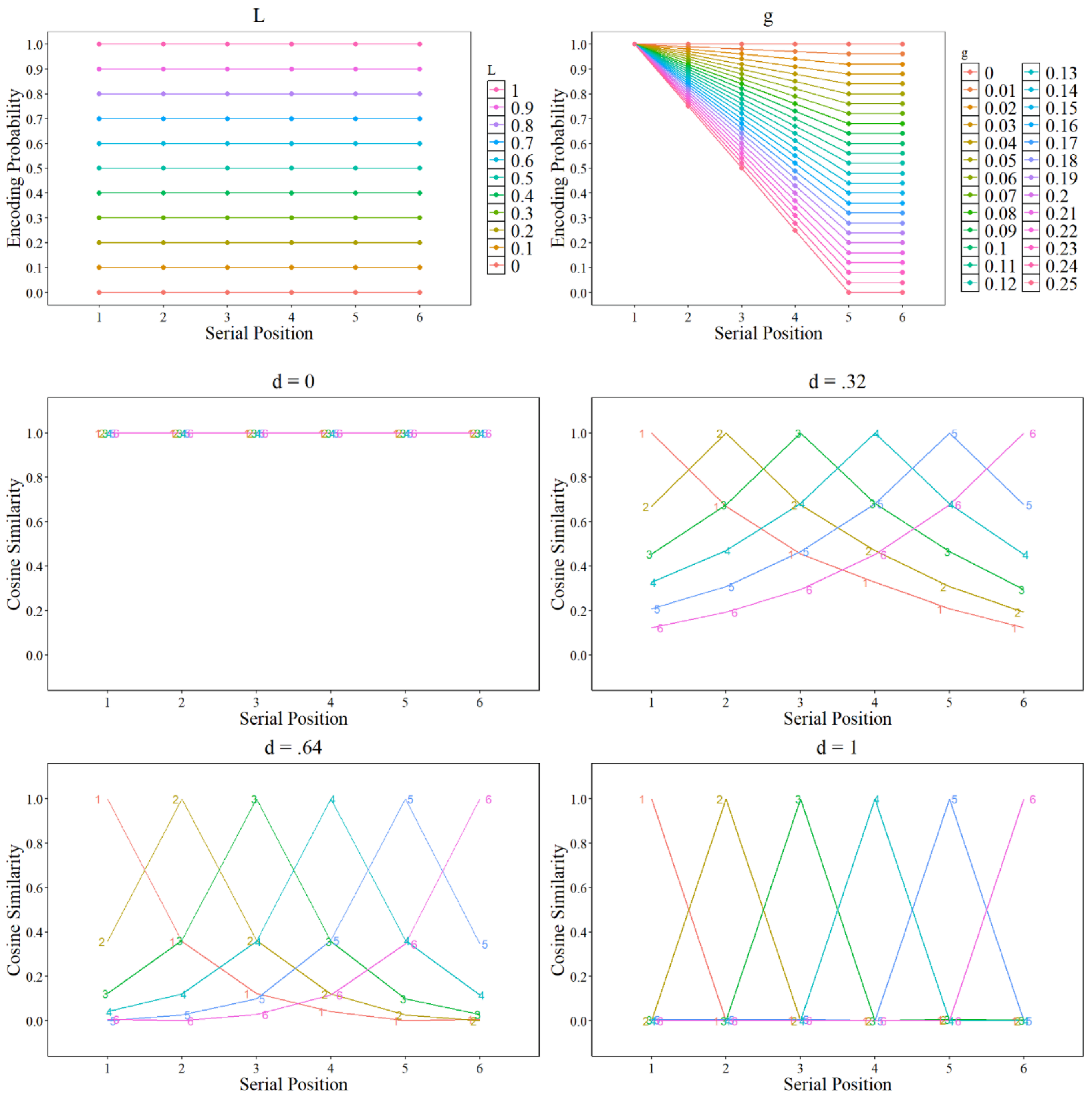

Fig. 1. Illustration of the free parameters of the models. *L* represents the learning base rate, *g* is the slope at which encoding declines as a function of serial position, and *d* indicates the similarity/dissimilarity of successive serial positions for memory of a study list. The top left panel illustrates the effect of *L*, the learning base rate, when *g* is fixed at 0 for each position. The top right panel illustrates the effect of *g* when *L* is fixed at 1. The four bottom panels show the cosine similarity of serial position representations as a function of *d*. The four bottom panels of the figure can be interpreted by cross-referencing the numbers in the body of the graph with the serial positions on the x-axis. For example, the profile of cosine similarities at serial position 1 within each graph, represented by the red line with points labelled "1″, shows the similarity of each serial position representation from 1 through 6 to the representation of the first serial position.

retrieval that follows, this means that the model will suffer from retrieval interference of events encoded at immediately adjacent serial positions most and suffer from retrieval interference from events at distant serial positions least with the value of *d* determining the degree of retrieval interference as a function of serial distance.

Study. We assume that people encode a list of to-be-remembered words as a series of traces in memory, where each trace encodes both the relevant serial position representation and the word that was presented at that serial position. To represent that assumption computationally, we represent memory as a two-dimensional matrix, **M**. Each row in memory is a 600-dimensional vector. The first 300 dimensions encode the serial position information and the last 300 dimensions encode the word's lexical representation (i.e., the word's LSA vector). Because the LSA representations for words have dimensionality 300, we define serial position representations at the same dimensionality. Thus, for a six-item study list, **M** is a 6×600 dimensional matrix. In the model and in line with recent findings on the relationship between item and order information (Guitard et al., 2021, 2022; Guitard & Cowan, 2023; Majerus, 2019), item and order information are independent traits essential to represent items-in-order.

We assume that memory for a studied word and its corresponding order information is imperfect and that items in earlier serial positions are encoded better than items at later serial positions, with the exception that items in the last two serial positions are encoded equally well. The assumption represents the fact that people have more opportunities to rehearse items earlier in the list (e.g., Bhatarah et al., 2009; Rundus, 1971) and later items are protected from interference (e.g., Nairne, 1990). To represent that assumption in the model, we copy each feature in a trace at serial position p with probability L_p,

$$L_p = \begin{cases} L - (p-1)g & ,p < LL \\ L - (p-2)g & ,p = LL \end{cases} \tag{1}$$

where L is the learning base rate (i.e., how well the item in the first position is encoded), p is the serial position, g is the slope at which encoding declines as a function of serial position, and LL is the number of items in the studied list. As indicated, each item is encoded less well than its predecessor at rate g with the exception that the last item in the list is encoded as well as the second last item in the list, similar to other implementation (e.g., Nairne, 1990).[1] To facilitate reader comprehension regarding the impact of various parameters in the model, we illustrate in the top left panel of Fig. 1 how the parameter L affects the encoding probability when g if fixed at 0. Conversely, in the top right panel in Fig. 1 we demonstrate the impact of changing the parameter g when L is fixed at 1. It is evident that as L increases, the likelihood of encoding information also rises. Inversely, an increase in g results in a lower encoding probability as a function of serial position. In our simulations, the parameter g is set relatively low and by itself, cannot fully explain the recency effect observed in serial recall. In the current model, the classic primacy effect, better memory for earlier presented items, and recency effect, small benefit for last present item, of the serial position function arises due to the interaction between encoding and differences in retrieval interference due to edge effects; edge effects refer to the fact that items first and last serial positions suffer less retrieval interference than items between those extremes because the first and last items have only one immediately adjacent trace whereas items in between have two (see Brown et al., 2007).

Serial recall. In this section, we described how we simulate serial recall. Readers are encouraged to refer to Fig. 2, which demonstrates how serial recall and serial reconstruction retrieve specific words from the lexicon for both related and unrelated lists while following the description of model.

To simulate recall of words at each of the LL serial positions in a study list, we present the representation for the relevant serial position to memory as a cue, $\mathbf{q}$, and retrieve a corresponding echo from memory, $\mathbf{e}$, that represents the information retrieved about the word that occurred at that serial position,

$$\mathbf{e} = \sum_{i=1}^{i=m} \left(\frac{\sum_{j=1}^{j=n/2} q_j \times M_{ij}}{\sqrt{\sum_{j=1}^{j=n/2} q_j^2}\sqrt{\sum_{j=1}^{j=n/2} M_{ij}^2}} \right)^3 \times M_{i\bullet} \tag{2}$$

where q_j is feature j in the cue, M_{ij} is feature j in trace i in memory, $n/2$ is the dimensionality of the serial position cue, and m is the number of traces in memory (i.e., the length of the study list in the simulations that follow). In psycholinguistics, the echo is a mental representation and would be referred to as a lexeme that represents the thought that underlies a language expression. Typically, the echo, $\mathbf{e}$, resembles the word at the cued serial position but by virtue of related studied words in the list having a similar representation to the presented word (as well as unstudied words that are also related to the studied word), the information retrieved from memory has potential to introduce an intralist (recalling an item from the list in the wrong position) or extralist (recalling an item not presented in the studied list) error. By the same token, and probably in contradiction to intuition, weak retrieval of studied items over the entire list into the echo can also help retrieval of the word at position p when studied words are related. This occurs because the word at serial position p shares semantic features with its competitors and the information retrieved about semantically related neighbours serves to reinforce and/or substitute for features unavailable in the word that the memory system is trying to retrieve (see the redintegration hypothesis; Neath et al., 2022; Saint-Aubin & Poirier, 1999a,b; Saint-Aubin et al., 2005).

Second, we compute the similarity of the information retrieved into the last 300 dimensions of the echo (i.e., the dimensions that encode word's lexical information) to each word in the lexicon (see Fig. 2 panels A and B). If the most similar word in the lexicon is greater than a report threshold, T, that word is reported (see Fig. 2 panels C and D). If no word in the lexicon has a similarity greater than T, no word is reported (an omission). Critically, and because the lexicon includes both words appearing both in and not in the study list, a reported word has opportunity to be the recall target (i.e., strict correct), a different word from the study list (i.e., an intralist intrusion), or a word from outside the study list (i.e., an extralist intrusion). In cases of false recall, like in the DRM False Recall procedure, the model can also falsely recall a critical lure.

Finally, we score the model's recall to match the way that people's recall is typically scored. If the word reported at serial position p matches the word presented at serial position p, the trial is scored as correct (strict scoring). If the word reported at serial position p does not match the word presented at serial position p but the word was in the list, the trial is scored as an intralist error (also known as an order error). If the word reported for serial position p does not match the word presented at serial position p and the word was not presented in the study list, the trial is scored as an extralist error. If no word is reported at serial position p because no word in the lexicon was similar enough to the echo, the trial is scored as an omission error.

For the simulations reported in this manuscript, we simulated fictional participants completing recall tasks for lists within each condition (i.e., each type of list and task). We then reported averaged performance over those simulated participants. This approach allows us to present averaged serial position functions and associated 95 % credible intervals, as well as the average proportions of recall errors: omissions, intralist errors, and extralist errors. Thus, the results of our simulations are presented for simulated participants in the same manner as experimental results are presented for real participants.

Serial reconstruction. We simulate serial reconstruction in nearly the same way that we simulate serial recall. However, the experimental procedure in a serial reconstruction task differs in two ways. First, participants in a serial reconstruction task are presented with the studied words for report at test; thus, participants do not need to compare to all the words in their lexicon only a lexical subset including the studied words (see Fig. 2 panels E and F). To simulate that difference to serial recall, the information retrieved into the echo is compared to the response list composed of just the words in the studied list for report. Second, because all words from the studied list are presented at recall, people select the best fitting word from the presented list; thus, omission errors in experimental performance are rare if committed at all. To match those facts, we force the model to select the best fitting word from the studied list at each serial position. That means that in serial reconstruction, the model can make correct recalls and intralist intrusion errors but cannot make omission or extralist intrusion errors. To prevent the repeated retrieval of a word once it has been recalled (known as a repetition error), the model suppresses the reporting of an already reported word at a rate s. The suppression parameter, s, was set to 1 during reconstruction, reflecting the task's characteristics where repetition

[1] This is a simplifying assumption in our current implementation of the model, and we will investigate the best form for this learning rate function in the future.

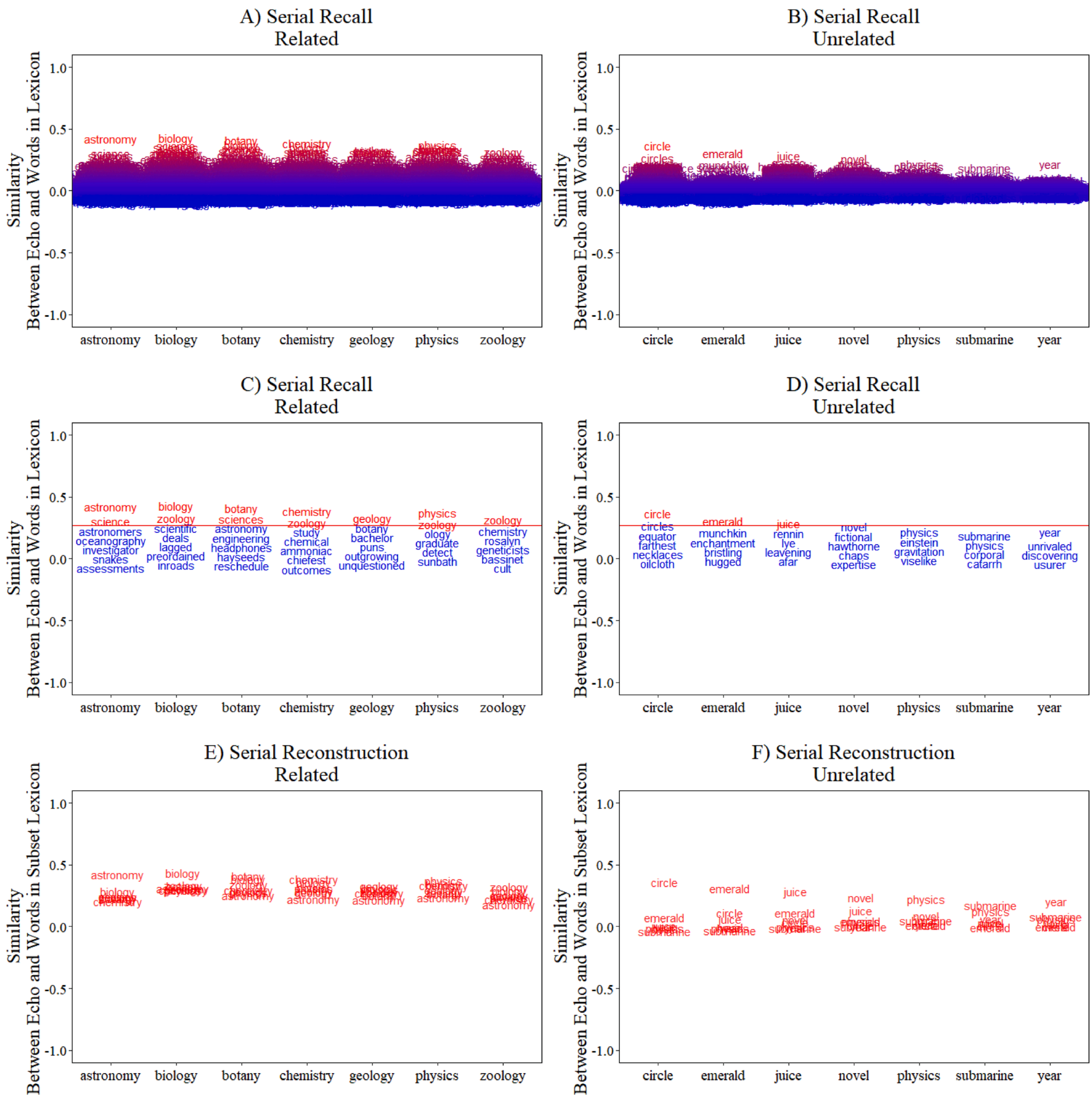

Fig. 2. This illustration shows the influence of semantic information on semantically related words (ASTRONOMY, BIOLOGY, BOTANY, CHEMISTRY, GEOLOGY, PHYSICS, ZOOLOGY) and unrelated words (CIRCLE, EMERALD, JUICE, NOVEL, PHYSICS, SUBMARINE, YEAR) in serial recall (Panels A to D) and serial reconstruction (Panels E and F). All figures were generated from 100 simulations of the model. Panels A and B display the similarity for each studied word (x-axis) between the echo and all words in the lexicon. Words that show the highest level of activation are highlighted in red, representing words considered to be in activated long-term memory according to Cowan's model (1988, 2019; Cowan et al., 2024). Words with the lowest level of activation are shown in blue. It is evident that related words exhibit a higher level of similarity compared to unrelated words. In Panels C and D, overlapping words have been removed to enhance visibility, and a horizontal red line indicates the report threshold. Words above this line in red are those likely to be recalled; the word higher on the y-axis is the one that will be recalled at that position. Words below the threshold (T) will not be recalled. If no red words are present for a given position (suggesting insufficient level of similarity between the echo and the words in the lexicon), this would result in an omission. If a word different from the one at a specific position is higher, it would be an extralist error if the word is not from the presented list, and an intralist error if it was a word from that list but not in that position. For order reconstruction, similarity is compared to the words in the list (subset lexicon), and the one with the highest level of activation will be reported regardless of similarity, as no omissions are possible in order reconstruction. See the text for more details.

errors are not possible and 0.01 for all other simulation unless otherwise specified. To ensure understanding of this simple implementation, consider the following example where the word CAT has a similarity score of 0.40 for the first position and 0.38 for the second position. In serial reconstruction, if the word CAT is selected, its value will be adjusted to 0.38 – 1 = -0.62, making it functionally impossible to be selected again. In contrast, in serial recall, its value will be adjusted to 0.38 – 0.01 = 0.37, making it still possible to be selected but less likely.

In summary, the eCFM assumes that people encode words-in-position at study to varying degrees of fidelity. At test, they use a serial position to cue memory and retrieve an echo (i.e., a lexeme). In serial recall, the echo is compared to all words in the lexicon and the best matching word is reported. In serial reconstruction, the echo is compared to just the words in the study list. Each reported word is then compared to the studied list to be scored as a correct or incorrect recall. Because the model reports words, the words it recalls can be scored in the same way as the words people recall in experiments: omissions, intralist errors, extralist errors, and even false recall of critical lures in the DRM protocol. In serial recall, the echo is compared to all the words in the lexicon and the word with the greatest similarity to the echo, above threshold, is reported. In serial reconstruction, the process is the same but only the words in the study list are available for report. Thus, any difference that arises in the two cases are due to that difference alone.

Demonstration 1: Simulation of semantic relatedness in immediate serial recall and immediate serial reconstruction of order

In our first demonstration, we applied the eCFM to Neath et al.'s (2022) design that documents the serial recall advantage for semantically related word lists coupled with the absence of that effect in serial reconstruction. Our decision to apply the model to Neath et al.'s study was driven by the comprehensive nature of their research: they provided the specific word lists they used so we could implement them in the eCFM simulations, made their data and detailed methodologies available, and were the first to systematically investigate the semantic relatedness paradox across three dimensions of word relationships using the same experimental procedure (category membership, associations, meaning). Thus, this set of experiments represents a hallmark collection of findings and an excellent empirical target. Of particular interest to us, it also allows us to investigate the model's ability to capture the phenomenon across different materials. The literature is full of demonstrations for phenomena driven by idiosyncratic stimulus properties (e.g., Guitard et al., 2018, 2019; Bireta et al., 2021, 2023), and Neath et al.'s experiments present an opportunity to test the eCFM in a set of internal replications using different materials.

In Neath et al.'s (2022) three experiments, participants performed serial recall and serial reconstruction for lists of six related or six unrelated words. On each serial recall trial, participants studied six successively presented words and then attempted to report the words in studied order. On each serial reconstruction trial, participants studied six successively presented words and then selected words presented on the screen to reproduce the order in which they were studied. As per common practice, participants in Neath et al.'s (2022) study were allowed to produce an omission error by choosing to "skip" recall of the word. The nature of the word relationships was varied across the three experiments. In Experiment 1, words in related lists were related by category membership (e.g., DIAMOND, RUBY, EMERALD, SAPPHIRE, PEARL, OPAL). In Experiment 2, words in related lists were related by association (e.g., MAD, FEAR, HATE, RAGE, TEMPER, FURY). In Experiment 3, words in related lists were related by meaning (e.g., AFFABLE, AMIABLE, GENIAL, CORDIAL, AGREEABLE, CIVIL). The reader can refer to Neath et al. for other details of their experimental procedure.

Method

To simulate Neath et al.'s (2022) experiments, we conducted simulations for 500 participants in each experiment, each completing 32 lists: 8 randomly sampled lists of each list type (i.e., Related and Unrelated) and for each recall task (i.e., serial recall and serial reconstruction) in each experiment. Because our model has a lexicon of 50,797 words, we were able to present the model with almost the identical lists that Neath et al.[2] presented to their participants and, in the context of serial recall, allowed the model to respond with any of those 50,797 words. To match their design, we constructed a related list on each recall trial by sampling one of Neath et al.'s related lists and presented the words in a randomized serial position order. When simulating performance for an unrelated list, we sampled six words from the full stimulus set (across lists) and presented those words in a randomized order.

On each simulated serial recall or serial reconstruction trial, we encoded each word of the list to memory in its presented serial position and then performed either serial recall or serial reconstruction for the list (depending on the task condition). On serial recall trials, the echo is compared to all words in the lexicon and then it reported a word at each successive serial position (see Fig. 2 panels A to D for illustration). On serial reconstruction trials, the echo was compared to the lexicon subset (i.e., just the six words in the studied list).

For all demonstrations, we fit the model by assessing its performance across a range of values in the model's joint parameter space, which included the rate at which each word in a list was encoded as a function of serial position, L and g, the threshold for issuing a response rather than 'skipping' or omitting to report a word during a serial recall trial, T, and the rate at which serial position representations differ over the six serial positions, d. We used successive grid searches to fit the model against Neath et al.'s data seeking the best overall fit across the three experiments (e.g., $L = .15$ to $.40$, $g = .01$ to $.04$, $d = .10$ to $.40$, $T = .20$ to $.40$).

Our parameters for this, and subsequent demonstration are reported in Table 1. As mentioned, s, the suppression mechanism, was set to 1 in simulations of serial reconstruction to reflect the task's characteristics where repetition errors are not possible, and to 0.01 for serial recall to reduce the number of repetition errors (i.e., the model reporting the same word at multiple serial positions).

Scoring. In these simulations and all experiments, we used the following scoring procedure. Strict scoring: A response was deemed correct if the identical word was recalled in its originally presented position only (e.g., word A presented in position 1 and word A recalled in position 1). Free scoring: A response was deemed correct if the identical word was recalled, regardless of its presented position (e.g., word A presented in position 1 and word A recalled in any position). Intralist error: An intralist error was scored based on each recalled position. It occurred when the word recalled at a position was not identical to the word presented at that position, was not an omission (a "skip"), and was a word presented in that list (e.g., the first word recalled was not equal to the word presented in position 1, was not an omission, but was a word from the list). Extralist error: An extralist error was scored when the recalled word was not one of the words presented in the list. For both intralist and extralist errors, if a word was repeated (e.g., **CAT**, TABLE, **CAT** in the same list), only the first occurrence was considered. Omission: An omission occurred when a word was not recalled or when the participant skipped as instructed. For fit between model and data across all demonstrations we reported R^2.

Availability. The simulations codes and materials for this and all subsequent demonstration are available on the OSF (Open Science Framework) page associated with the manuscript (OSF).

Results

To assess the model's performance as a function of task (i.e., serial

[2] Unfortunately, some words in Neath et al.'s (2002) lists were not available in the LSA semantic memory vectors. In those cases, we deleted the lists from our simulations resulting in five fewer lists in Experiment 1 and one fewer list in Experiment 3. All stimuli were available for Experiment 2. For details on the specific stimuli used in our simulations, please refer to the OSF page associated with the manuscript.

Table 1
Parameters and vectors used for each demonstration.

Demonstration	Study	Task	Vectors	L	d	g	T	s
0	Fig. 2	Serial Recall	LSA English	0.16	0.35	0.018	0.27	–
0	Fig. 2	Serial Reconstruction	LSA English	0.16	0.35	0.018	–	–
1	Neath et al. (2022)	Serial Recall	LSA English	0.16	0.35	0.018	0.27	0.01
1	Neath et al. (2022)	Serial Reconstruction	LSA English	0.16	0.35	0.018	–	1
2	Saint-Aubin et al. (2005)	Serial Recall Control	LSA French	0.21	0.25	0.018	0.27	0.01
2	Saint-Aubin et al. (2005)	Serial Recall Concurrent articulation	LSA French	0.16	0.25	0.018	0.27	0.01
3	Poirier et al. (2015)	Serial Recall	LSA English	0.19	0.3	0.018	0.27	0.01
4	Experiment 1	Serial Recall	LSA English	0.16	0.35	0.018	0.27	0.01
5	Experiment 2	Serial Recall	LSA English	0.16	0.35	0.018	0.27	0.01
6	Experiment 3	Serial Recall 114 ms	LSA English	0.125	0.3	0.014	0.27	0.01
6	Experiment 3	Serial Recall 500 ms	LSA English	0.165	0.3	0.014	0.27	0.01
7	Experiment 4	Serial Recall	LSA English	0.165	0.3	0.014	0.27	0.01
8	Experiment 5	Serial Recall	LSA English	0.18	0.3	0.018	0.27	0.10
8	Experiment 5	Serial Recall	fastText	0.18	0.3	0.018	0.27	0.10

recall versus serial reconstruction), word relatedness (i.e., related versus unrelated word list), serial position (1 to 6), and Experiment (1 to 3), we computed the mean performance and associated 95 % credible intervals (e.g., see Saint-Aubin et al. 2021 for similar method) over all 500 simulated participants and report the averaged serial position functions. In the case of serial recall, we also report the rates of omission, intralist, and extralist intrusions errors. To facilitate a quick comparison of Neath et al.'s (2022) empirical results to our simulations, we present the empirical results one row above each simulation for each experiment. As shown, and despite the differences in how word relatedness was defined in each the three experiments (i.e., category membership in Experiment 1, association in Experiment 2, and meaning in Experiment 3), Neath et al. reported the same overall outcomes. Participants exhibited better memory of related than unrelated lists in serial recall with no such difference in serial reconstruction. The serial recall error patterns for related versus unrelated lists also remained generally consistent across the three experiments. We now report on our model's capacity to track participants' performance across the three experiments.

Overall. Fig. 3 shows the model's performance in each of Neath et al.'s (2022) three experiments where the relatedness of word lists was defined by category membership in Experiment 1, word association in Experiment 2, and word meaning in Experiment 3. The top panel shows the results averaged over all three experiments, independent of how word relatedness was defined. Overall, as shown in Fig. 3, the model had a good overall fit to the data and captures many of key features of the data with only minor discrepancy, $R^2 = 0.90$.

More exactly, first and foremost, when collapsed across all experiments, the model exhibited better memory for related than unrelated lists in serial recall along with absent or a small reversed difference in serial reconstruction. Based on that distinction, the model predicts the dissociation that motivated Neath et al.'s (2022) study and that motivated our interest in explaining that difference. Second, the results are relatively consistent over the three experiments independent of how word relatedness was defined (i.e., category membership, association, or meaning in Experiments 1, 2, and 3, respectively). Third, the model predicts the general form of the empirically observed serial position functions. Both serial recall and serial reconstruction is best for words in serial position one (i.e., the primacy effect), declines over the middle serial positions, and exhibits a recency effect in the last serial position (i.e., better memory for words that end a studied list). Fourth, the model predicts a steeper negative serial position function in serial recall than in serial reconstruction. All these features in the model's performance match the corresponding features from Neath et al.'s empirical data; notably, with fits in each of the three experiments handicapped by using the same model parameters to fit performance over all three. Fifth, the model produces a reasonable facsimile of the distribution of error types on serial recall trials with a higher omission error rate for unrelated than related study lists, a modestly higher intralist intrusion error rate for related than unrelated study lists, but it does not capture the very modest difference in the extralist intrusion rate for related versus unrelated study lists. Although the model captures the major differences observed in Neath et al.'s experiment, there are some minor differences to the experimental data. For example, the slopes of the serial recall position functions of the related and unrelated lists in all three experiments are not so strikingly parallel as in the empirical data and serial reconstructions of words from unrelated lists is slightly better than serial reconstruction of words from related list intralist errors. In some instances, there is a small detrimental difference for related over unrelated (e.g., Experiment 1). However, those differences are minor relative to the overwhelming similarity between the model's and participants' performance overall.

Lexicon Size. In this section, we examined whether the size of the lexicon in serial recall (50,797 words) and in serial reconstruction (6 studied items plus distractors: unstudied items based on the lexicon size) affects the results. To explore this question, we conducted 500 simulations of Experiment 2 (in which all the words were available) and explored the effects of various lexicon sizes from 0 (only the studied words) to 40,000 randomly chosen words for all scoring procedures for both serial recall and serial reconstruction. As shown in Fig. 4, in serial recall, the pattern of results is stable across lexicon sizes for all measures except extralist intrusions which increased with the number of words in the lexicon. In contrast, as lexicon size increases in serial reconstruction, an interaction emerges from small detrimental effects of relatedness at very small lexicon sizes to a small beneficial effect of relatedness arising with a lexicon of 100 words or more. As expected, there is an increase in extralist errors more important for unrelated that related, a decrease in intralist errors and no omission because they are not possible without a recall threshold. This analysis of serial reconstruction provided valuable insight into the influence of semantic relatedness in our model. Unlike serial recall where we have a recall threshold in serial reconstruction, the model is forced to recall the word with the highest level of similarity between the echo and the word in the lexicon. This shows that when the lexicon is small it is often slightly harder to retrieve the specific words giving the similarity while when the lexicon is larger it is the opposite, similarity help retrieve the specific words consistent with the redintegration hypothesis. In sum, the advantage of related words relative to unrelated words in serial recall is consistent across all lexicon sizes but without a lexicon the model cannot produce extralist intrusions. The results of serial reconstruction shed light on the apparent dissociation between the tasks and might suggest that even without a recall threshold we can observe an advantage if participants on each trial have activated words in their lexicon of approximately 100 to 1000 words which is sufficient to produce an advantage on strict scoring, free scoring, and extralist errors, but would not produce omission errors, which occurs with recall threshold.

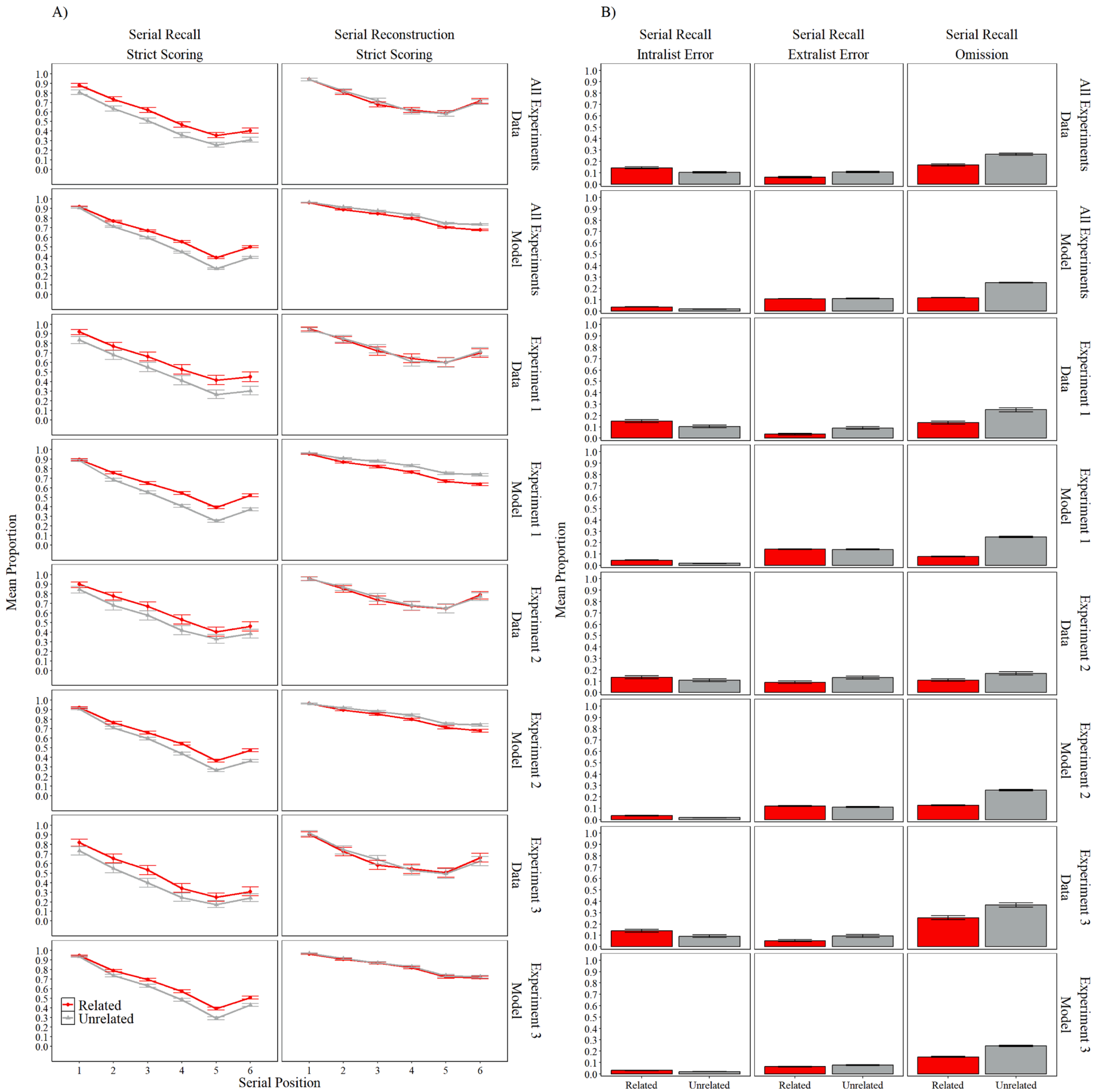

Fig. 3. The model's serial recall and serial reconstruction performance (identified by model) for related and unrelated lists for each of Neath et al.'s (2022) three experiments (identified by data). Error based corresponds to 95% credible intervals. Panel A corresponds performance across serial positions according to strict scoring criterion, and Panel B corresponds to performance collapsed across serial position for related and unrelated intralist, extralist and omission error.

Discussion

Taken together, we conclude that the model successfully accounts for the semantic similarity paradox by capturing the dissociation observed by Neath et al. (2022): a relatedness advantage in serial recall but not in serial reconstruction. The model also replicates key characteristics of serial position functions (i.e., primacy and recency), although, as mentioned, further work is needed to understand the underlying mechanisms.

While the model is not without flaws, with some minor discrepancies between serial recall and serial reconstruction noted in the first demonstration—such as small detrimental effects in certain instances—these are generally minor. In the supplementary analysis, we examined the proportion of times each participant and each simulation exhibited a semantic relatedness effect. This effect could be beneficial (related > unrelated), detrimental (related < unrelated), or show no difference (related = unrelated) across experiments. The distribution of these effects (beneficial, detrimental, no difference) is captured with only minor discrepancies (see Figure S1).

More importantly, in our model, the relatedness difference as a function of task does not reflect a difference in how encoding and recall proceed in those two tasks. Rather, the difference reflects an interaction between how memory functions, according to the model, and the different constraints imposed by these two tasks at test. Serial recall requires the retrieval of a word from the lexicon, whereas serial reconstruction only requires the retrieval and recognition of a word from the

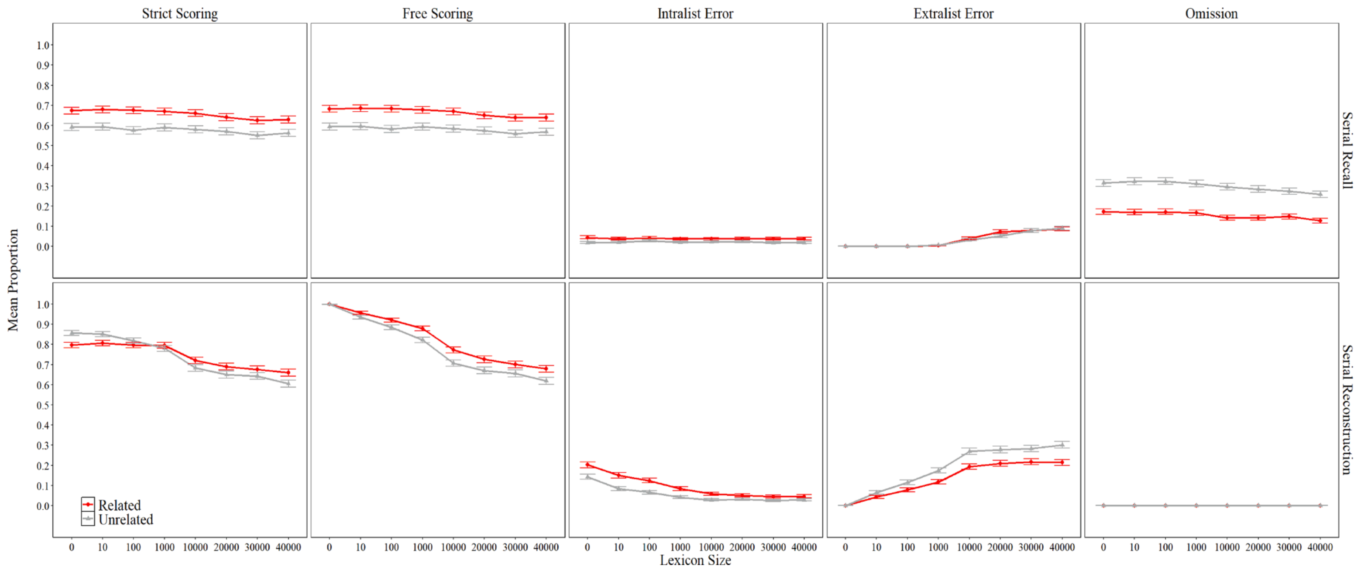

Fig. 4. The model's performance in serial recall and serial reconstruction with distractors (unstudied words from the lexicon) for related and unrelated lists in Experiment 2 of Neath et al. (2022) with various lexicon sizes (0 to 40,000 words) is presented. Error bars correspond to the 95% credible intervals. The top row corresponds to serial recall, and the bottom row corresponds to performance in serial reconstruction with distractors. Each panel corresponds to a different scoring method: strict scoring, free scoring, intralist errors, extralist errors, and omission errors. The x-axis represents the lexicon size (number of words in the lexicon), and the y-axis represents the mean proportion.

studied list. We also show in Fig. 4, that by increasing the number of words in the lexicon it does not affect serial recall other than extralist errors suggesting that the results were stable across lexicon sizes. In fact, the benefit emerges from the similarity between the echo and these related words in the lexicon that are easier to retrieve because they are more similar. However, unlike previous accounts, similarity is not arbitrarily defined. Instead, here similarity is defined by a well-established model of semantic memory (LSA, Landauer & Dumais, 1997). In addition, by manipulating the size of the lexicon in serial reconstruction we show that when the lexicon size increased the likelihood of retrieving the specific words also increased showing how the effect of relatedness in serial reconstruction goes from small or negative to positive. Importantly, we also note that the results emerge from simulations that not only assume a lexicon of word representations in which memory is embedded but also where almost the identical word lists presented to the model were also presented to experimental participants. Thus, our model not only predicts recall as a function of word relatedness in general but also predicts recall as a function of word relatedness in particular. Our integration of similarity and vector-based representations for words into a model for memory represents a degree of computational specificity that is relatively rare at present, albeit increasingly common in the field (Chang & Johns, 2023; Criss & Shiffrin, 2004, 2005; Franklin & Mewhort, 2002, 2015; Johns et al., 2012; Kimball et al. 2007; Nosofsky et al., 2018a, 2018b; Mewhort et al., 2018; Lewandowsky & Murdock, 1989; Osth et al., 2020; Osth & Zhang, 2023; Reid & Jamieson, 2023; Reid et al., 2023; Sirotin et al., 2005).

Demonstration 2: Simulation of the detrimental effect of semantic relatedness on order information in immediate serial recall

In our second demonstration, we aimed to investigate the detrimental effects of semantic relatedness on order information. As discussed in the introduction, the consensus is that semantic information does not affect order information. However, there has been opposing evidence that are often disregarded (Saint-Aubin et al., 2005; Tse, 2009; Tse et al., 2011). In this demonstration, we will investigate the influence on order information by simulating the results of Saint-Aubin et al. (2005). As previously mentioned, Saint-Aubin et al. (2005) conducted one of the largest experiments on semantic relatedness defined by category membership in French with 252 participants. Two groups of participants (one studying related lists and another studying unrelated lists) performed the task under concurrent articulation (repeating aloud the word 'mathématiques' at a rate of about three repetitions every 2 s). Additionally, two other groups (one studying related lists and another studying unrelated lists) performed the task without concurrent articulation. The effect of semantic relatedness (the difference between related and unrelated lists) was slightly larger when the task was more difficult (i.e., under concurrent articulation than in the control condition). More importantly, there was a detrimental effect on order.

Method

To simulate the experiment conducted by Saint-Aubin et al. (2005), we conducted simulations for 500 participants in each condition, each completing 14 lists in the same fixed order that Saint-Aubin et al. used for each group of participants (related control, unrelated control, related concurrent articulation, unrelated concurrent articulation). Since the participants and the studied lists were in French, we developed new French LSA vectors using the same techniques employed in the previous simulation. To derive the French LSA representations, we randomly selected 40,000 articles from French Wikipedia with the contingency that each article contained 250 words or more. From each article, we then randomly selected a 250-word section. We then built the word-by-document matrix by counting word occurrences across these 40,000 snippets of 250 words each. The vectors are available on the OSF page associated with this manuscript.

In the subsequent simulations, we used 300-dimension French LSA word representations derived from these 40,000 articles from French Wikipedia. However, similar to our approach with the English lexicon, in a final step, we filtered the vectors to include only items that appear in the Lexique database (New et al., 2004). The final lexicon included 40,862 words. This lexicon enabled us to present the model with almost identical words to those used by Saint-Aubin et al.[3] The parameters we used are presented in Table 1. To simulate the influence of participants under concurrent articulation, we provided them with a lower encoding probability ($L = 0.16$) relative to participants in the control group ($L = 0.21$).

Scoring. Due to the limited availability of the original data, we focused on strict scoring, free scoring, and conditional order error. The latter corresponds to the number of intralist errors for a given list divided by the number of words recalled according to a free scoring procedure. We applied this latter procedure because it was used by Saint-Aubin et al. (2005).

Results

To facilitate the reader's comparison to our simulations, we have redrawn the results of Saint-Aubin et al. (2005) next to our data in Fig. 5. As shown in Fig. 5, the model accurately tracks all key characteristics of the results from Saint-Aubin et al. (2005). Specifically, performance was superior in the control condition compared to the concurrent articulation condition. The semantic relatedness (the difference between the related and unrelated conditions) was more pronounced when the task was more difficult under concurrent articulation compared to the control condition. More importantly, there was a detrimental effect on order information for related lists compared to unrelated lists for both the concurrent articulation and control groups simulations. Overall, the model demonstrated a good fit to the data and captured many of the key features, with an $R^2 = 0.80$.

Discussion

The results of Saint-Aubin et al. (2005) have often been disregarded in the literature and have received little computational effort to examine their ability to track these important findings. Here, we have shown that by changing the lexicon to reflect that of the participants (from English to French) and using their materials to simulate the results, the model captures all main features. More importantly, it demonstrates a slightly larger effect of semantic relatedness when the task was more difficult, and a consistent detrimental effect on order. Many studies have dismissed these results by suggesting that this was a between-participants manipulation and the results might not be reproducible under different experimental conditions (e.g., Neath et al., 2022). It is always possible that their results are due to the particularity of their stimuli (e.g., Guitard et al., 2018, 2019; Bireta et al., 2021, 2023), the use of fixed-order lists (e.g., Guitard et al., 2023), or a combination of both. However, we believe that additional efforts are warranted to understand the exceptions rather than dismissing them. The evidence from our simulation appears to suggest that both empirically and computationally, with their particular design and almost identical lists, the results are consistent. In Experiment 4, we investigated further the influence of semantic of information on order information with a design that addressed the methodological concerns of Saint-Aubin et al. (2005).

Demonstration 3: Simulating the impact of semantic relatedness on migration errors

In our third demonstration, we aimed to investigate the influence of semantic relatedness on the pattern of order errors—specifically, where a specific word is most likely to be incorrectly recalled—by simulating the results of Poirier et al. (2015). In their study, they positioned three semantically related words in the first three positions (e.g., CANARY, MUSTARD, BANANA). In the control condition, the subsequent words were unrelated, whereas in the experimental condition the 5th word (YELLOW) was strongly associated with the initial trio. They found that participants were more likely to recall the 5th word earlier in the sequence compared to its counterpart in the control list, indicating a significant influence of semantic association on memory retrieval. This pattern of results was captured at the qualitative level by Kowialiewski et al. (2021) with randomly generated vectors. Here, we demonstrate the model's ability to capture the pattern of migration error and also the ability of the model to assess item level performance across a series of measures.

Method

Due to the availability of materials of the experiment conducted by Poirier et al. (2015) we were able to simulate the exact same 40 participants with their exact words in their same presentation position. We concluded 400 simulations (10 simulations of each of the 40 participants) with the English LSA representation as was done in Demonstration 1. The simulations parameters are presented in Table 1.

Scoring. Due availability of the original data we were able to compute all scorings for both the overall and item level performance.

Results

Overall. To facilitate the reader's comparison to our simulations, we have reanalyzed the results of Poirier et al. (2015) and plotted the results alongside our data in Figs. 6, 7, and 8. As shown in Fig. 6, at the overall level, the model, like the simulations of Kowialiewski et al. (2021), misses some unusual features of Poirier et al.'s results, such as the lack of a recency effect. However, it tracks many features, such as relatively equivalent performance between the control and experimental conditions across all scoring methods, except for fewer omissions of the target item in position 5 than the experimental data. It captures the small detrimental effects at position 4 and the slight advantage for the item in position 5, more apparent with free scoring. Overall, the fit of the results presented in Fig. 7 is relatively good, with $R^2 = 0.92$.

Migration. In Fig. 7, we explored the pattern of migration errors of the target item studied in position 5 and show that the model tracks many features of Poirier et al.'s (2015) results. Like Poirier et al., participants, when they made an error, were more likely to recall the target item in earlier positions compared to the control condition (e.g., positions 4 and 3). However, there are some minor discrepancies; the model made slightly fewer order errors, and performance in position 5 was slightly better in the model. Overall, the fit of the results presented in Fig. 7 is relatively consistent with the initial results, with $R^2 = 0.97$.

Items Level. In Fig. 8, due to the embedded representation and the availability of all the words in the lexicon, the model was able to directly compare the proportion of times each studied word was accurate or inaccurate. Specifically, we show the proportion of times each studied word was scored as strict scoring, free scoring, intralist error, extralist error, and omission. This demonstration is not possible without representations and allows an additional level of specificity. Overall, as shown in Fig. 8, across all scoring methods, the model provides a relatively good account of the results, with $R^2 = 0.85$. However, for intralist and extralist errors, the model has more difficulty, mostly due to the very small number of these errors in the experiment and some overprediction in instances where words were more likely to be replaced by another word in the lexicon (e.g., "music"), likely attributed to the high level of similarity with other words in the lexicon.

Discussion

In this demonstration, we have shown that the model captured many important features of the classic migration pattern of errors established

[3] Unfortunately, the following 4 words were not available in our lexicon, (céleri, yogourt, bicycle, muffin) and we replaced by the following 4 semantically related words (radis, mousse, vélo, brioche). For further details on the specific stimuli used in our simulations, please refer to the OSF page associated with the manuscript.

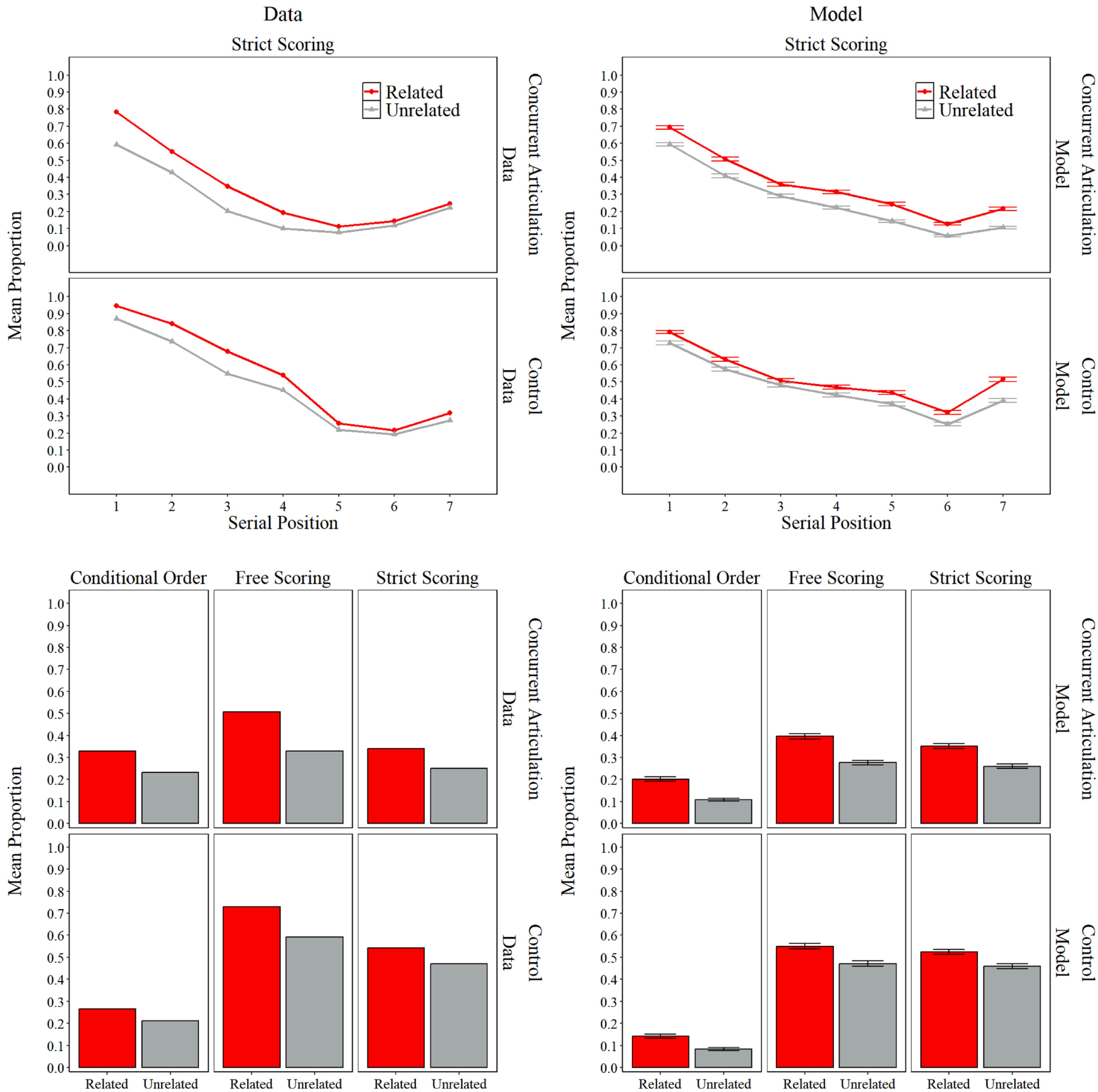

Fig. 5. The model's serial recall performance (identified by model) for related and unrelated lists for the control and concurrent articulation groups of Saint-Aubin et al. (2005) (identified by data). Error bars correspond to the 95% credible intervals (not available for the data). The top row shows performance across serial positions according to strict scoring criteria, while the bottom rows display performance collapsed across serial positions for related and unrelated lists under conditional order error, free scoring, and strict scoring.

by Poirier et al. (2015). Similar to Kowialiewski et al. (2021) without an embedded lexicon as defined here, the pattern shows some deviations but captures key elements, such as the migration of the target item towards the other similar items presented in earlier positions. Unlike previous computational demonstrations, we show that the model can account for memory performance with a higher degree of specificity than previous accounts of serial recall. Specifically, because the model includes a lexicon, we were able to compare the degree of overlap between the model and the data at the word level while capturing the exact relationship at the list level which have been shown to be critical in serial recall (e.g., Guitard et al., 2023). This allows us to evaluate the model at both macro and micro levels without any additional assumptions. This demonstration opens interesting avenues beyond veridical recall. For example, in Experiment 5, we will demonstrate how the model can track performance to produce words that were not studied by investigating false recall.

Present empirical study

Overall, our computational model successfully accounted for a range of important findings with a novel level of specificity that provides both empirical and theoretical insights. Specifically, the model captured the differential influence of semantic relatedness in serial recall and reconstruction reported by Neath et al. (2022) across different

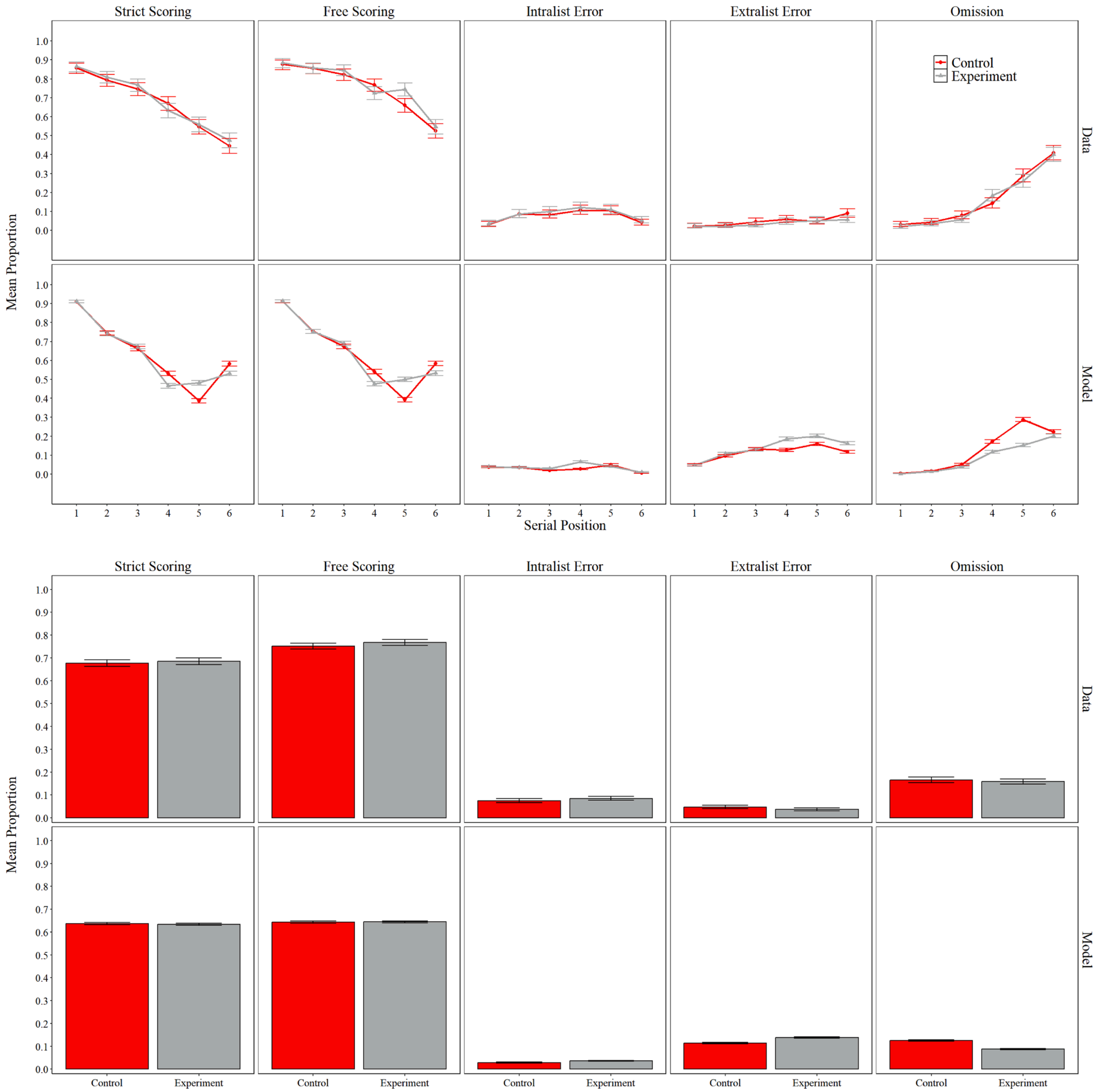

Fig. 6. The model's serial recall performance (identified by model) for control and experimental lists of Poirier et al. (2015) (identified by data). Error bars correspond to the 95% credible intervals The top two rows show performance across serial positions according to strict scoring, free scoring, intralist error, extralist error, and omission error, while the bottom rows display performance collapsed across serial positions.

operationalizations of semantic relatedness in the experimental lists (i.e., category membership, association, or meaning).

The results were stable across different lexicon sizes, but a careful analysis of serial reconstruction with distractors (studied and unstudied words) shows that lexicon size is an important factor in explaining how semantic relatedness becomes better than unrelated lists in our model. The model also successfully accounted for the detrimental effect of semantic information on order information observed by Saint-Aubin et al. (2005), which has often been dismissed. Using almost identical materials, the model predicted the pattern of results and captured the effect of task difficulty. In the third demonstration, we showed that the model was able to capture the pattern of migration errors identified by Poirier et al. (2015) with some success, a benchmark finding of the influence of semantic information on verbal memory performance. Furthermore, the model could evaluate at the item level, which has eluded previous models of serial recall without meaningful representations.

Beyond accounting for previous results, the model also makes predictions for serial recall of mixed semantic lists. In the following studies, we test these predictions in two experiments. In Experiment 1, semantically related words are presented in alternating serial positions within the list (e.g., **HOUR**, GENERAL, **MINUTE**, JOURNAL, **DAY**, JADE), while in Experiment 2, semantically related words are presented in blocked positions within the list (e.g., **HOUR**, **MINUTE**, **DAY**, GENERAL, JOURNAL, JADE). This design was engineered to evaluate whether the model can capture details on how semantic relatedness affects serial recall performance with better precision.

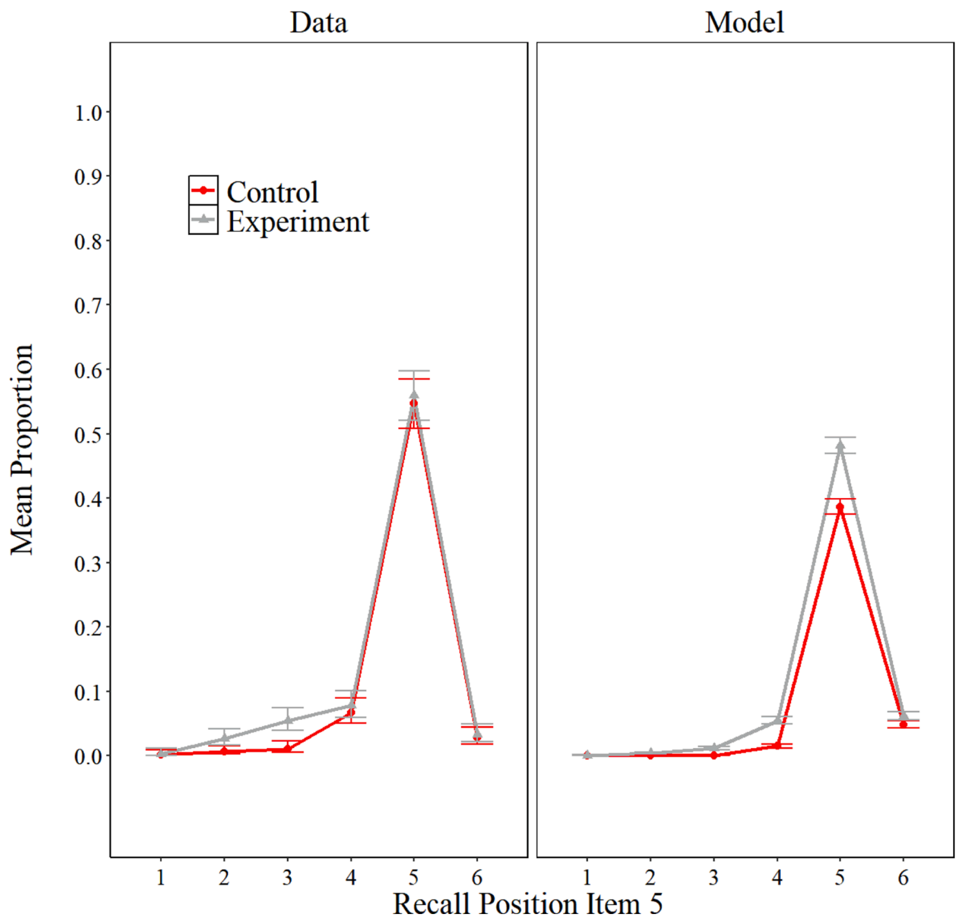

Fig. 7. The mean proportion of time the target word presented in position 5 was recalled at each serial position. The model's serial recall performance (identified by model) for control and experimental lists of Poirier et al. (2015) (identified by data). Error bars correspond to the 95% credible intervals.

In addition to investigating the influence of list organization, we revisited the influence of task difficulty on semantic relatedness in Experiment 3 by manipulating presentation rate. In Experiment 4, based on the results from our simulations of Saint-Aubin et al. (2005), we re-evaluated the influence of semantic relatedness on order information using 160 participants and a new set of stimuli. Lastly, in Experiment 5, we assessed whether the model could track specific extralist errors by manipulating the number of words related to a target word, leveraging the lexicon's capability to predict such errors.

Demonstration 4: Simulation of serial recall for pure and alternating mixed lists of related and unrelated words

We have, thus far, demonstrated that our model postdicts numerous key phenomena of semantic information on verbal short-term memory performance. This is a meaningful success demonstrating that a computational model of recall equipped with a lexicon can capture at least some aspects of human memory performance by presenting similar lists to the model. However, readers might be underwhelmed as we reverse-engineered our model to explain the results, albeit in a way that follows directly from cued recall in MINERVA 2 (Hintzman, 1986). In the work that follows, we put serial reconstruction aside and focus in on demonstrations that our model can track details of serial recall performance.

A cornerstone experimental manipulation in the study of short-term memory is mixed list recall. In mixed list designs, half of the words in a list are related in some way (e.g., low arousal words) and the other half are related in another (e.g., high arousal words; Landry et al., 2022). One benchmark example of differential recall for word types is the word frequency effect: memory is better for high-frequency than low-frequency words (e.g., Guérard & Saint-Aubin, 2012; Hulme et al., 1997; Roodenrys et al., 1994). However, whereas that difference holds true in pure list designs (i.e., where all words are of the same kind), the word-frequency effect is abolished in mixed lists designs (e.g., Hulme et al., 2003; Morin et al., 2006). These results have been central to advancing the field's understanding of how long-term memory factors interact with short-term memory performance.

As reviewed in the introduction, list organization is an important phenomenon on the influence of semantic relatedness. There are some demonstrations of the issue. For example, Brooks and Watkins (1990) provide an important investigation of list organization in a memory span task where lists are composed of both digits and words (Experiment 1) or just words from the same or different semantic domain (Experiment 2). Unfortunately, only overall memory span was reported in their study. Therefore, the pattern of results as a function of serial position remains unknown. Saint-Aubin et al. (2014) proposed one of the first attempts to investigate the influence of mixed list organization in serial recall. More exactly, they presented a pair of related and unrelated words within a list either at adjacent serial positions or separated by one or two unrelated items. They found better recall performance when the pair members were presented in adjacent serial positions. Kowialiewski and Majerus (2020) also investigated list organization; however, they used two categories of related words (e.g., **LEAF**, **TREE**, **BRANCH**, CLOUD, SKY, RAIN). In subsequent works, they examined related and unrelated words similar to our interest but omitted important control conditions for pure lists (e.g., all related words) (Kowialiewski et al., 2021, 2022).

Despite the importance of their work, none of Brooks and Watkins (1990), Saint-Aubin et al. (2014), or Kowialiewski and Majerus (2020;

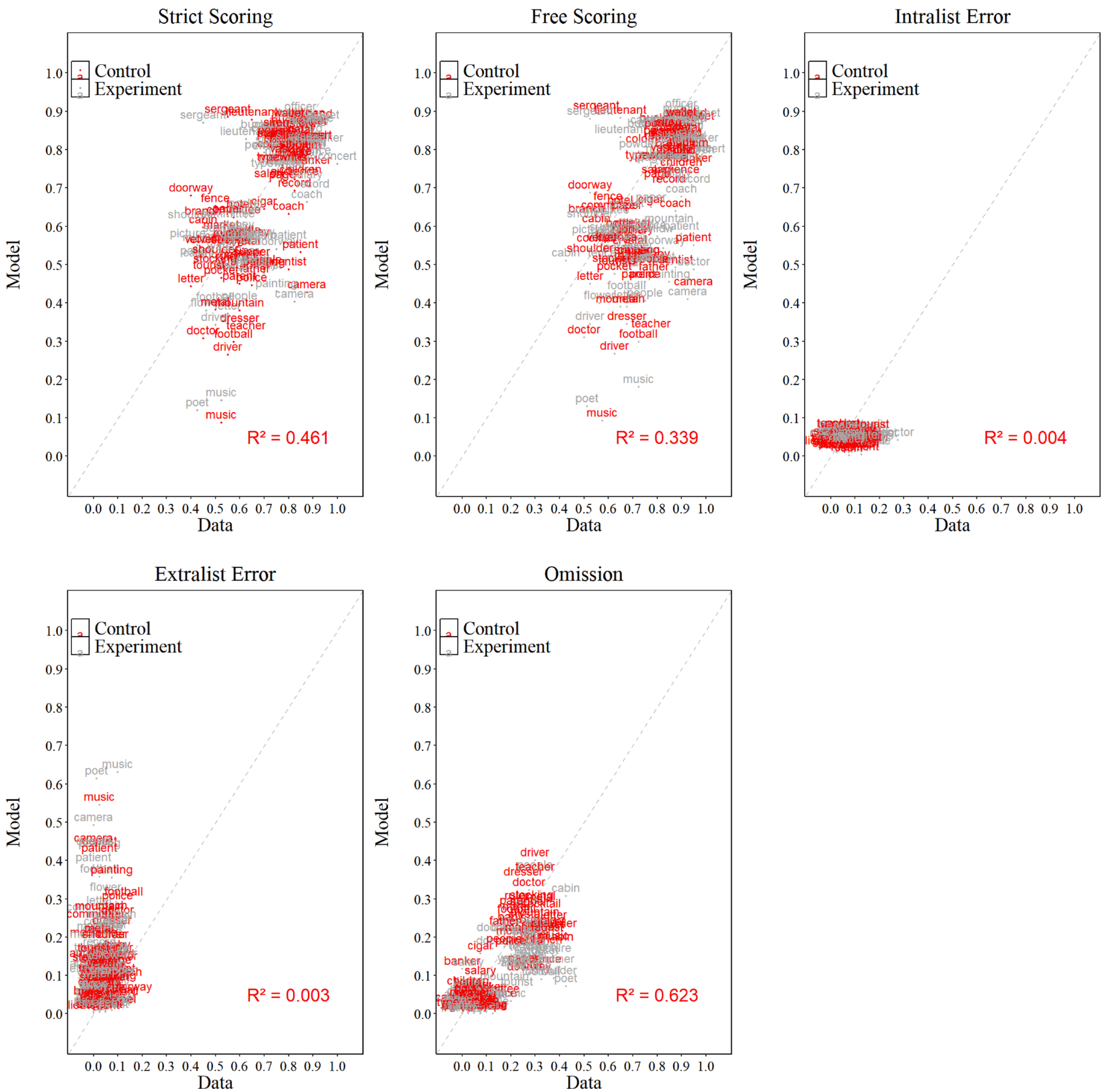

Fig. 8. The mean proportion of time each word was scored as strictly correct, freely correct, intralist error, extralist error, and omission, along with the overall fit for each scoring procedure. The model's performance is on the y-axis, and the results from Poirier et al. (2015) are presented on the x-axis.

Kowialiewski et al., 2021, 2022) provided clear predictions or empirical assessments to establish precise expectations of serial recall for semantically related and unrelated words in a mixed list design; at least not in the same way that Hulme et al. (2003) did for word frequency. Here we bridge this gap by, first, establishing predictions for mixed list serial recall of related and unrelated words with the eCFM and then testing those predictions in a corresponding experimental procedure.

In a first test, we applied the model to performance in serial recall for pure-related (i.e., RRRRRR) and pure-unrelated (i.e., UUUUUU) word lists, like in the Neath et al. (2022) study. However, we also tested the model's serial recall for otherwise equivalent alternating mixed-list conditions. In an RURURU mixed-list condition, lists were composed of three related words and three unrelated words, where three related words appeared in odd numbered serial positions and three unrelated words appeared in even numbered serial positions (e.g., **HOUR**, GENERAL, **MINUTE**, JOURNAL, **DAY**, JADE). To complement that, we also tested serial recall for words in a URURUR mixed-list condition where three related words appeared in even numbered serial positions and three unrelated words appeared in odd numbered serial positions (e.g., HOUSE, **ARM**, SWORD, **ELBOW**, PEACH, **KNEE**). Our predictions of interest were focused on measuring a benefit for related over unrelated words in both pure and mixed lists and also asking how the shape of the serial position functions would differ over the RRRRRR, UUUUUU, RURURU, and URURUR list structures.

To bring some precision to our empirical test, we generated specific word lists for all four of the wordlist test conditions: 24 RRRRRR pure lists, 24 UUUUUU pure lists, 24 RURURU mixed lists, and 24 URURUR mixed lists. For completeness, our word lists in all four test conditions are presented in Appendix A. In conformity with good experimental design, the same 144 words have equal probability of occurring in all conditions and only the list-wise arrangement of words differs between test conditions.

The parameters for our simulation are presented in Table 1. To simulate performance, we conducted simulations for 100 simulations of 24 serial recall trials for each of the RRRRRR, UUUUUU, RURURU, and URURUR list structures which is equivalent to 400 fictional participants. Then, we averaged performance over those simulated participants to report the serial position curves based on all scoring methods. Our predictions of interest are on strict serial scoring (i.e., words had to be recalled in their presented position to be correct) and free scoring (i.e., words were scored as correct if they were recalled in either their presented position or any other), but for transparency we reported all of them. Mean simulated performance for RRRRRR, UUUUUU, RURURU, and URURUR lists is presented in Fig. 9.

The five leftmost columns show serial position functions for recall of words in RRRRRR, UUUUUU, RURURU, and URURUR lists according to strict scoring, free scoring, intralist error, extralist error and omission. The five rightmost columns show average recall of R and U words in the pure list conditions (RRRRRR and UUUUUU lists) and in the mixed list conditions (RURURU and URURUR lists). For clarity, the R recall estimate in the mixed list conditions is equal to mean recall of R words from the odd numbered serial positions in **R**U**R**U**R**U mixed lists and the even numbered serial positions of U**R**U**R**U**R** lists with the U recall estimate equal to mean recall of U words from the even numbered serial positions in R**U**R**U**R**U** lists and the odd numbered serial positions of **U**R**U**R**U**R lists.

The model made two key predictions that lend themselves to empirical evaluation. First, the model predicts a memorial advantage for R over U words in both pure and mixed lists, with the size of that advantage being slightly larger for pure over mixed lists. Second, and although a very high precision prediction, the model predicts complementary sawtooth patterns in the two mixed list conditions (i.e., RURURU and URURUR) with recall not only being better for R over U items in aggregate but also being better (generally) for R over U items at each serial position. We now turn to an experimental test of the model predictions.

Experiment 1

In Experiment 1, we examined the influence of semantic relatedness in a serial recall task while manipulating list composition to test the predictions derived from the model. More exactly, we tested people's serial recall of related and unrelated words in pure related lists (RRRRRR), pure unrelated lists (UUUUUU), and alternating mixed lists (RURURU and URURUR).

Method

Participants

In the absence of prior demonstrations examining the semantic relatedness effect in serial recall with mixed lists like we are investigating, we based our calculation on the effect size of Experiment 1 of Neath et al. (2022) with pure lists. This study was chosen because our experiment was modeled after their experiment and our stimuli were drawn from their set of words. In their experiment, the size of the semantic relatedness effect between related and unrelated lists was $d = 0.805$. We used their effect size to estimate the sample size via G*Power 3.1.9.7 (Faul et al., 2007). An a priori bidirectional paired-sample t-test power analysis with the effect size of Neath et al. and default parameters revealed that with 23 participants we would have at least a power of 0.95 to detect an effect comparable to Neath et al. However, and because we are expecting a reduced effect for mixed lists, we computed a sensitivity analysis. Results from the sensitivity analysis with default parameters revealed that 80 participants would allow us to detect a smaller effect $d = 0.41$.

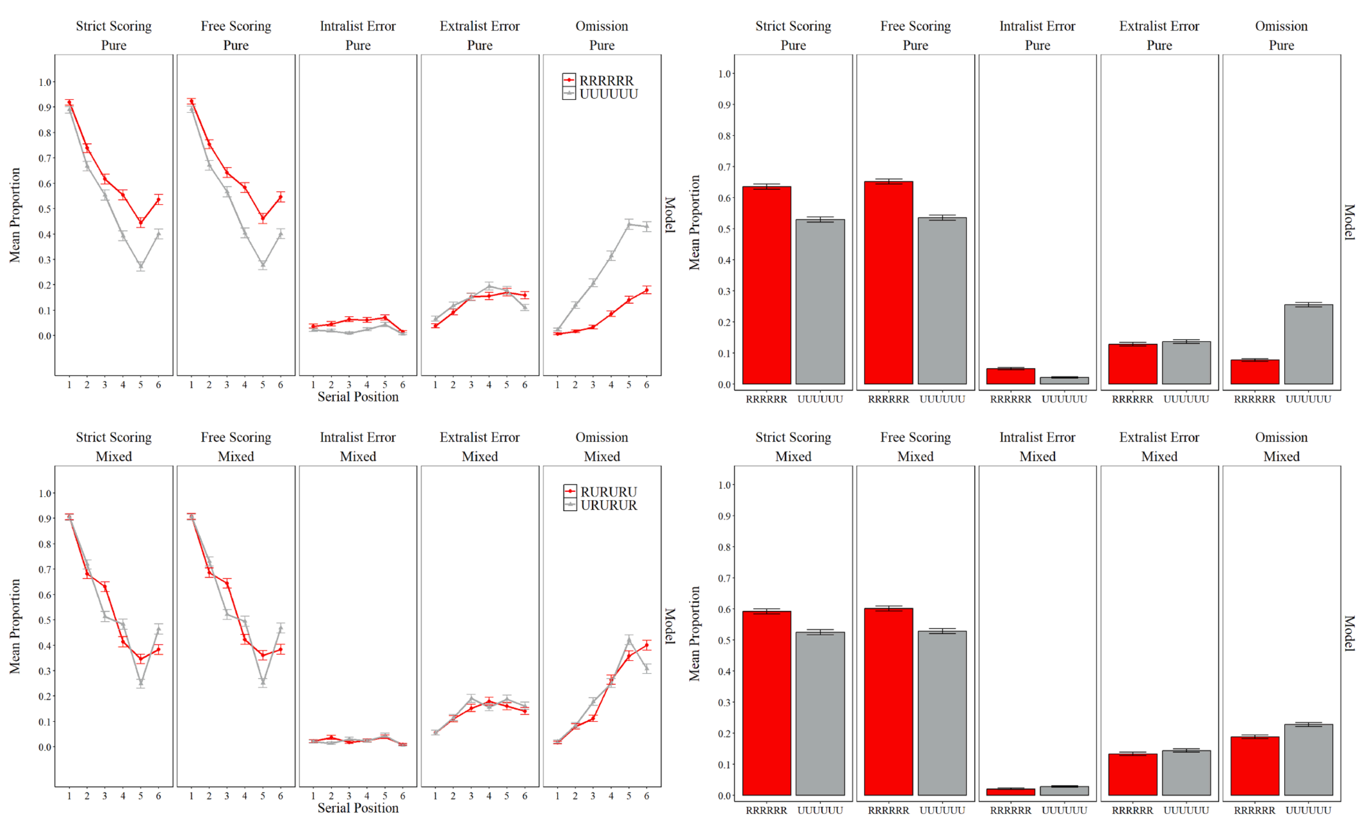

Fig. 9. Model simulations of Experiment 1. Although we present these simulations as predictions, the predictions are shown from simulations using parameters selected after running the experiment. Error bars correspond to the 95% credible intervals.

Eighty participants took part in this experiment. All participants were recruited via Prolific (https://www.prolific.co/) and received £9.00 per hour (pro-rated) for their participation. To participate in this experiment, participants had to (a) be a native speaker of English, (b) have American nationality, (c) have normal or corrected-to-normal vision, (d) have no cognitive impairment or dementia, (e) have no language-related disorders, (f) be between 18 and 30 years of age, and (g) have an approval rating of at least 90 % on prior submissions at Prolific. All inclusion criteria were self-reported by the participants except for the approval rating which is computed by Prolific. Participants in the study had a mean age of 24.79 years (SD = 3.50, range 18–30) and 28 self-identified as female, 47 self-identified as male, and 5 preferred to not specify their gender.

Materials

The stimuli were drawn from Neath et al.'s (2022) Experiment 1 in which semantic relatedness was operationalized by category membership. More exactly, 24 lists of six semantically related words were borrowed from Neath et al. The stimuli were then rearranged to create 24 lists of semantically unrelated words and 48 lists of semantically mixed words: 24 RURURU lists and 24 URURUR lists. The stimulus lists are presented in Appendix A. Across participants, the words were presented equally often in each of the four possible list compositions (i.e., RRRRRR, UUUUUU, RURURU, URURUR) and a word was never repeated for any given participant. The order of the words within a list was fixed but the order of the lists was randomized for each participant.

Design

The experiment was a 2 (pure lists vs. mixed lists) × 2 (related vs. unrelated) × 6 (serial position 1 to 6) repeated-measures factorial design. There were 24 experimental trials (i.e., 6 RRRRRR lists, 6 UUUUUU lists, 6 RURURU lists, and 6 URURUR lists) each corresponding to a list composition. The order of list composition was individually randomized for each participant. Participants completed two practice trials before beginning the experiment.

Ethic

These and subsequent methods were approved by the School of Psychology Ethics Committee at Cardiff University, and all participants provided informed consent before the experimental session.

Procedure

The experiment was programmed with PsyToolKit (Stoet, 2010, 2017) and took participants approximately 20 min to complete. The progression of the experiment was controlled by the participant who initiated each trial by pressing the space bar within 60 s after finishing the preceding trial. If the participant did not initiate the next trial before the 60 s window had passed, the next trial was presented to ensure the procedure would be completed within the expected time. Once the trial was initiated, the six to-be-remembered words were presented sequentially at a rate of one word every second (1000 ms on 0 ms off) at the centre of the computer screen. The words were presented on a black background in white uppercase 30 points Times New Roman font. Immediately, after the presentation of the last word, a recall cue ("Type the first word") was presented in red at the top of the computer screen. Participants were instructed to type the words in their presentation order by pressing the return key after each word. Once participants registered their response by pressing the return key, the typed word disappeared, and the message was updated, "Type the second word". This was repeated until all responses were typed. If participants were unable to remember a given word, they were instructed to type the word "Skip". Participants were not allowed to backtrack to change a response once they registered an answer.

Data analysis

The experimental data are available on the OSF page associated with this project. In this study, a strict spelling criterion was adopted. Thus, recalled words were only scored as correct if they were spelled correctly. Analyses in this study were conducted on performance scored with a strict serial recall criterion (serial scoring) in which a word had to be recalled in its presented serial position to be scored correct and with a free recall criterion (free scoring) in which recall of a word was scored correct whether it was recalled in the correct or a different serial position.

For statistical analyses, we conduced both Frequentist and Bayes factors (BF) analyses via the statistical software R (R Core Team, 2022). Frequentist analyses are presented as descriptive information and BF analyses are presented to guide our inferences. The BF analyses were computed with the "BayesFactor" R package with the default parameters (Version 0.9.12–4.2; see Morey & Rouder, 2018; Rouder et al., 2009) and the frequentist analyses were computed with the R package "ez" (Version 4.4–0; Lawrence, 2016). For BF analyses, we report the strength of evidence in favour of the alternative hypothesis indicated by BF_{10} or the corresponding strength of evidence in favour of the null hypothesis indicated by BF_{01} ($BF_{01} = 1/BF_{10}$).

Results

In all experiments, analyses of variance were conducted separately for the strict serial recall scoring and the free recall scoring criteria. The results of Experiment 1 are displayed in Fig. 10 for all scoring procedures for transparency, but we focus on strict serial recall scoring and the free recall scoring criteria as we had no prior hypotheses for the other scoring.

Strict Scoring. Overall, as predicted, participants' performance was superior for semantically related words (M = 0.70, SD = 0.15) relative to semantically unrelated words (M = 0.64, SD = 0.18). Participants' overall performance was also nearly identical between mixed lists (M = 0.67, SD = 0.17) and pure lists (M = 0.67, SD = 0.16).

A 2 (related vs. unrelated) x 2 (pure vs. mixed) x 6 (serial position 1 to 6) ANOVA was conducted to confirm these trends. The results from the analysis of variance revealed a main effect of semantic relatedness, $F(1,79) = 31.00$, $\eta_p^2 = .28$, $BF_{10} > 1000$, and a main effect of serial position, $F(5,395) = 157.30$, $\eta_p^2 = .67$, $BF_{10} > 1000$, but no main effect of list composition, $F < 1$, $\eta_p^2 = .00$, $BF_{01} = 42.08$. The results from the analyses revealed that there was a credible interaction between semantic relatedness and list composition, $F(1, 79) = 9.59$, $\eta_p^2 = .11$, $BF_{10} = 30.77$. This interaction reflects the larger semantic relatedness effect for pure lists compared to mixed lists.

The latter interaction was further investigated via separate paired-sample t-tests. For pure lists, participants recalled related words (M = 0.71, SD = 0.16) better than unrelated words (M = 0.63, SD = 0.19), $t(79) = 4.97$, *Cohen's d* = .56, $BF_{10} > 1000$. The same pattern was observed for mixed list albeit reduced in magnitude. More exactly, for mixed list, participants recalled related words (M = 0.68, SD = 0.17) slightly but reliably better than unrelated words (M = 0.66, SD = 0.18), $t(79) = 2.93$, *Cohen's d* = .33, $BF_{10} = 6.33$.

Returning to the main analysis, there was no credible interaction between semantic relatedness and serial position, $F(5, 395) = 3.05$, $\eta_p^2 = .04$, $BF_{01} = 86.96$, no interaction between list composition and serial position, $F(5, 395) = 1.78$, $\eta_p^2 = .02$, $BF_{01} > 1000$, and no three-way interactions, $F(5, 395) = 3.45$, $\eta_p^2 = .04$, $BF_{01} = 23.13$.

Free Scoring. With free recall scoring, participants recalled semantically related words (M = 0.79, SD = 0.11) better than semantically unrelated words (M = 0.71, SD = 0.15). Participants' memory

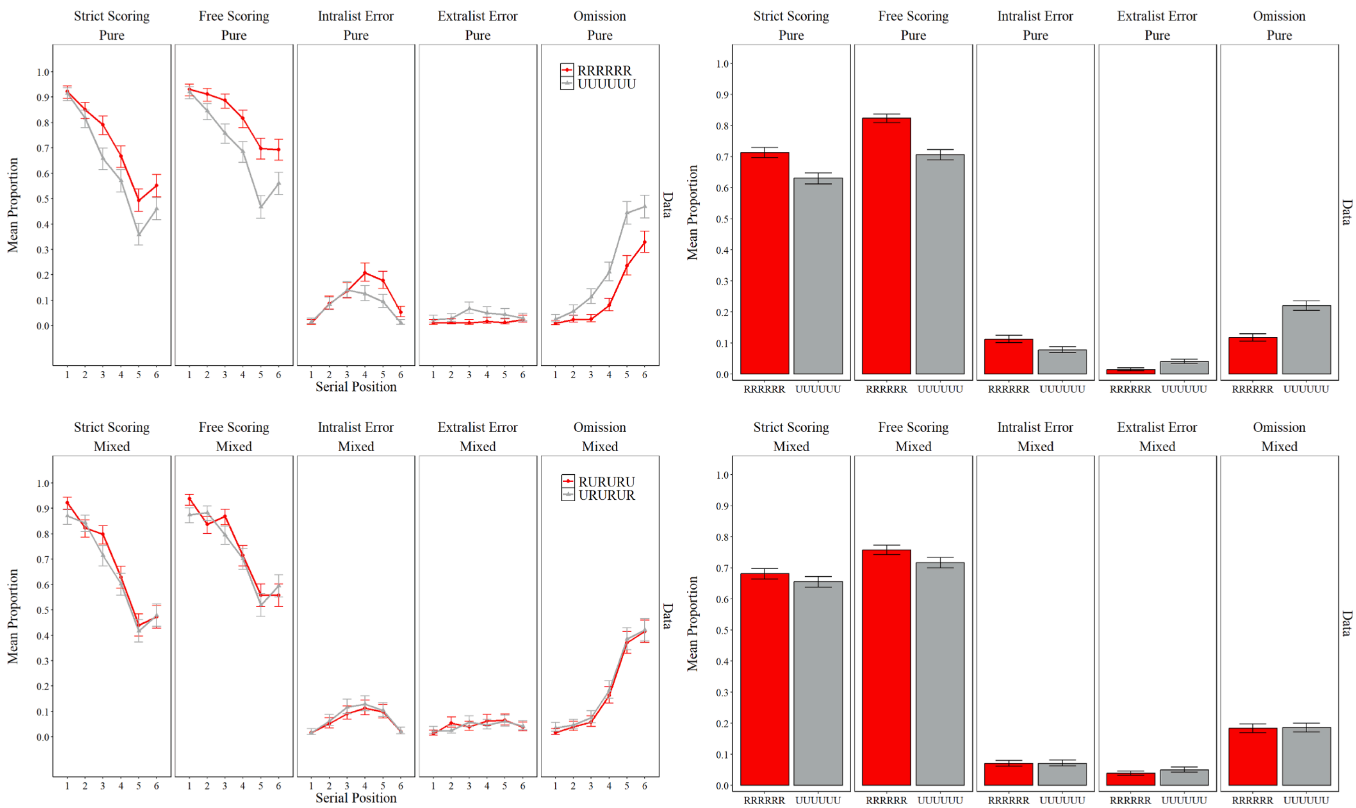

Fig. 10. Mean proportion of response as a function of list composition (pure lists, mixed lists), semantic relatedness (related, unrelated), serial positions (1 to 6) and scoring procedures (strict scoring, free scoring, intralist error, extralist error, omission) in Experiment 1. Error bars correspond to the 95% credible intervals.

performance was also better with pure lists (M = 0.76, SD = 0.12) than mixed lists (M = 0.74, SD = 0.14).

The results from the analysis of variance confirmed the presence of all main effects. More exactly, there was a main effect of semantic relatedness, $F(1,79) = 65.08$, $\eta_p^2 = .45$, $BF_{10} > 1000$, a main effect of serial position, $F(5,395) = 97.01$, $\eta_p^2 = .55$, $BF_{10} > 1000$, and a main effect of list composition, $F(1,79) = 12.07$, $\eta_p^2 = .13$, $BF_{10} = 27.69$. As observed with the strict recall scoring, the results with the free recall scoring revealed the presence of an interaction between semantic relatedness and list composition, $F(1,79) = 24.29$, $\eta_p^2 = .23$, $BF_{10} > 1000$. This interaction reflects the larger semantic relatedness effect for pure lists relative to mixed lists as shown in Fig. 10.

Once again, we further analysed this interaction. Echoing the results with strict serial scoring, better recall of related words over unrelated words was larger in pure compared to mixed lists. More exactly, for pure lists, participants' performance was superior for related words (M = 0.82, SD = 0.10) relative to unrelated words (M = 0.71, SD = 0.16), $t(79) = 7.97$, *Cohen's d* = .89, $BF_{10} > 1000$. Again, the same pattern was observed for mixed lists, albeit reduced in magnitude. For mixed lists, participants recalled related words (M = 0.76, SD = 0.13) better than unrelated words (M = 0.72, SD = 0.16), $t(79) = 4.20$, *Cohen's d* = .47, $BF_{10} = 276.77$.

Returning back to the main analyses, there was also a three-way interaction, $F(5, 395) = 6.30$, $\eta_p^2 = .07$, $BF_{10} = 64.32$, that reflects the increasing larger semantic relatedness effect as one moves from early to late serial positions for pure but not mixed lists. There was no interaction between semantic relatedness and serial position, $F(5, 395) = 4.03$, $\eta_p^2 = .05$, $BF_{01} = 8.06$, or between list composition and serial position, $F(5, 395) = 1.94$, $\eta_p^2 = .02$, $BF_{01} > 1000$.

Modelling Assessment. In this section, we briefly highlight the experimental results in relation to the model predictions. As mentioned, model simulations of Experiment 1 are presented in Fig. 9. The overall model fit across all scoring methods was relatively good ($R^2 = 0.91$). As shown, the model tracks several key features in the empirical data, including some unexpected and more exploratory aspects. For example, for omissions, the model accurately tracks the higher omission rate in unrelated than in related lists with a pure list design, and its important reduction in mixed lists. The model captures, to some extent, the relatively similar number of extralist errors between related and unrelated lists in mixed lists, although it slightly underpredicts the number of extralist errors in unrelated pure lists.

The model tracks relatively well, albeit to a lower magnitude, the higher number of intralist errors for related lists compared to unrelated lists in pure lists and their reduction in mixed lists. This aligns well with recent evidence suggesting that mixed lists might constrain order errors (Kowialiewski et al., 2024). Returning to the main point of interest, the model tracks the noisy but visible sawtooth recall pattern for RURURU and URURUR lists in early serial positions but overpredicts the persistence of the sawtooth recall functions at later serial positions; this misprediction is particularly true when performance is measured by a strict scoring method. More critically, the model predicts the semantic relatedness effect: better recall of related (R) over unrelated (U) items in all cases, with the size of the advantage being slightly greater in pure lists compared to mixed lists.

Discussion

A beneficial effect of semantic relatedness was observed in pure lists that was greatly reduced in mixed lists. The serial position functions for mixed-list conditions show some signs of a sawtooth pattern, though that holds more strongly at early than late serial positions, unsurprising given the increase in variance and thus precision of the means from serial positions 1 through 6. This pattern of results echoes the results of Saint-Aubin et al. (2014) in which the influence of semantically related pairs was reduced when they were separated by an unrelated item. We also show that the model tracks many of the critical features in our participants' serial recall performance for both pure and mixed lists; acknowledging that the model's behaviour is much more stereotyped from simulated participant to participant than our experimental participants' behaviour was consistent from human participant to participant (e.g., the model predicts a stronger and more reliable sawtooth serial position functions in the mixed list conditions). In Experiment 2, we carry our analysis forward to investigate serial recall of blocked instead of alternating mixed lists.

Demonstration 5: Serial recall for pure and blocked mixed lists of related and unrelated words

Serial recall in Experiment 1 provided a reasonable match to predictions from our model. We now turn to additional predictions to evaluate if our model tracks the details of people's serial recall performance in blocked mixed lists. Here like Kowialiewski et al. (2021, 2022) we used one block of semantically related words and one block of unrelated words (e.g., **HOUR**, **MINUTE**, **DAY**, JADE, GENERAL, JOURNAL). However, as our interest was to examine the influence of list organization on the semantic relatedness, we also added a pure condition with all related or all unrelated words. This is an important test for the model. As seen in Experiment 1, the sawtooth patterns for alternating lists, where recall is better for related than unrelated items at each serial position was attenuated. These results were consistent with the finding of Saint-Aubin et al. (2014) in which a pair of semantically related words were better remembered when presented in adjacent serial positions. The remaining question is whether the model can predict a more robust benefit when semantically related words are presented in immediately adjacent serial positions and whether the empirical results will support this prediction.

As in Experiment 1, we conducted simulations for serial recall of pure related (RRRRRR) and pure unrelated (UUUUUU) word lists. However, in difference to Experiment 1, we tested serial recall for blocked instead of alternating mixed lists. In an RRRUUU blocked mixed list condition, lists were composed of three related and three unrelated words (as in Experiment 1) but with the three related words appearing in the first three serial positions whereas the three unrelated words appeared in the last three serial positions (e.g., **HOUR**, **MINUTE**, **DAY**, JADE, GENERAL, JOURNAL). In a UUURRR blocked mixed list condition, lists were composed of three related and three unrelated words, with the three related words appearing in the last three serial positions instead (e.g., JADE, GENERAL, JOURNAL, **HOUR**, **MINUTE**, **DAY**). To bring some precision to our predictive exercise, we generated specific word lists for all four test conditions: the same 24 pure related and 24 pure unrelated word lists from Experiment 1, 24 blocked mixed lists with related words appearing in the first three serial positions, and 24 blocked mixed lists with related words appearing in the last three serial positions. For completeness, the word lists in all six test conditions are presented in Appendix B. Although the same 144 words occur once in each of the six list conditions, the list-wise arrangement of words changes between conditions.

The simulation parameters were identical to the previous demonstration and are presented in Table 1. To simulate performance, we conducted simulations for 100 simulations of 24 serial recall trials for each of the RRRRRR, UUUUUU, RRRUUU, and UUURRR lists structure, which is equivalent to 400 fictional participants.

Simulation results are presented in Fig. 11. Again, the key predictions of interest are on strict serial scoring (i.e., words had to be recalled in their presented position to be correct) and free scoring (i.e., words were scored as correct if they were recalled in either their presented position or any other), but for transparency we reported all of them. Like in our simulations for Experiment 1, the model makes two main predictions for empirical evaluation. First, the model predicts a similar memorial advantage for R over U words in both the pure and mixed list conditions. Second, and at the more fine-grained level of

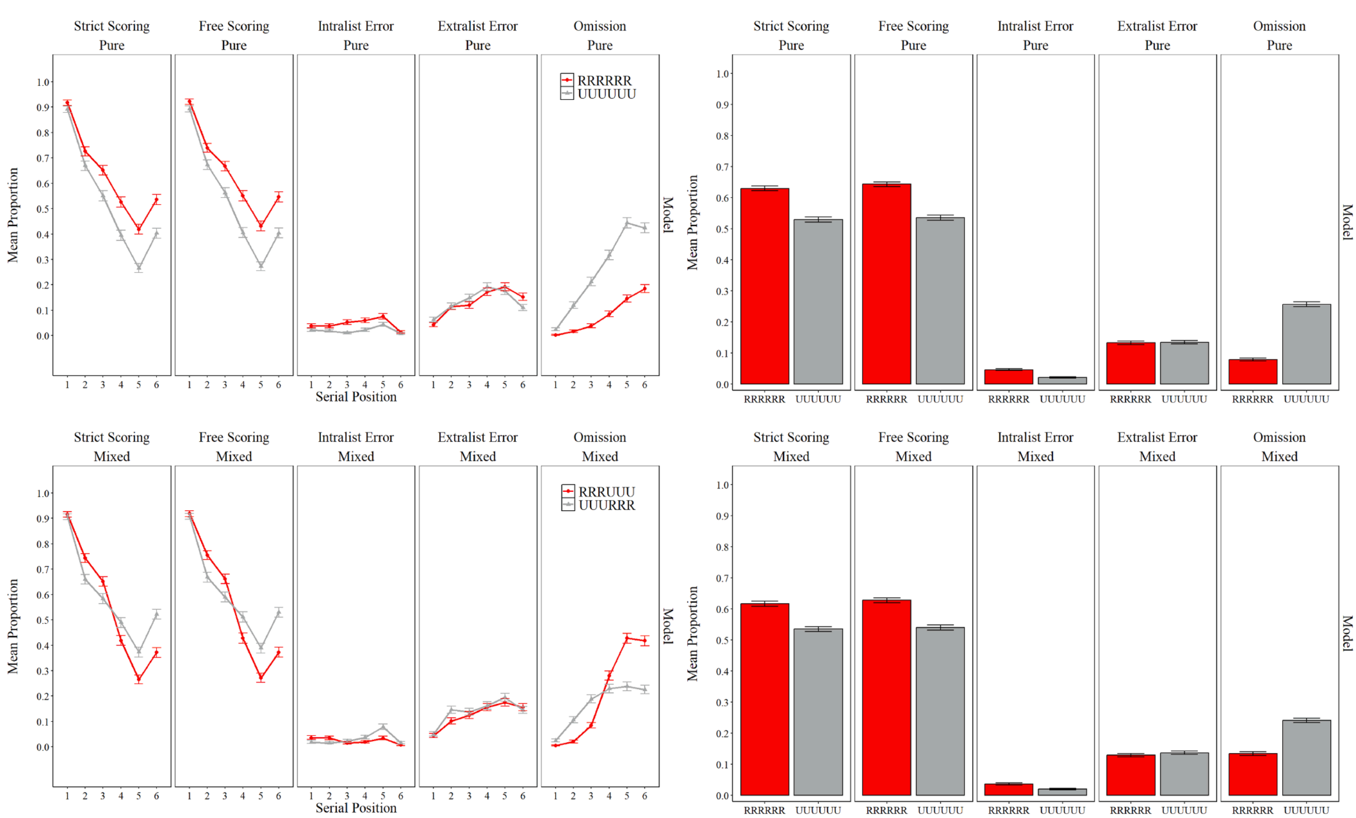

Fig. 11. Model simulations of Experiment 2. Although we present these simulations as predictions, the predictions are shown from simulations using parameters selected after running the experiment. Error bars correspond to the 95% credible intervals.

analysis, the model predicts better recall of R over U words at almost all serial positions with a crossover interaction between the third and fourth serial positions where R words shift to U words in RRRUUU lists and where U words shift to R words in UUURRR lists (i.e., by analogy, the blocked list equivalent of the sawtooth pattern in mixed lists). We now turn to an experimental test of the model predictions.

Experiment 2

Experiment 2 was identical to Experiment 1, except that related and unrelated words in mixed lists were presented in blocked (RRRUUU and UUURRR) rather than alternating (RURURU and URURUR) serial position order. This strategy was previously used by Brooks and Watkins (1990) in a span task, where they found better recall for lists beginning with related words (see also, Kowialiewski et al., 2021, 2022). Unfortunately, they did not include the necessary pure lists as controls, nor did they report serial position curves, which are key to testing the model.

Method

Participants

Eighty participants who did not take part in the previous experiment were recruited via Prolific with the same inclusion criteria as in Experiment 1. In Experiment 2, participants' mean age was 24.83 years (SD = 2.98, range 18–30). A total of 24 participants self-identified as female, 54 as male, and 2 preferred not to specify their gender.

Materials, design, procedure, and data analysis

The materials, design, procedure, and data analysis procedure were identical to Experiment 1 except for the list composition. As shown in Appendix B, the mixed lists were rearranged to ensure that the semantically related words were adjacent to each other (RRRUUU or UUURRR).

Results

The results of Experiment 2 across all scoring procedures are presented in Fig. 12.

Strict Scoring. Like in Experiment 1, and according to a strict serial recall scoring criterion, participants were better at recalling semantically related words ($M = 0.70$, $SD = 0.16$) relative to semantically unrelated words ($M = 0.65$, $SD = 0.19$). Overall recall performance was equivalent for mixed lists ($M = 0.67$, $SD = 0.18$) and pure lists ($M = 0.68$, $SD = 0.18$).

A 2 (related vs. unrelated) x 2 (pure vs. mixed) x 6 (serial position) ANOVA was conducted to confirm the descriptive results. The statistical analyses revealed a main effect of semantic relatedness, $F(1,79) = 27.12$, $\eta_p^2 = .27$, $BF_{10} > 1000$, a main effect of serial position, $F(5,395) = 119.46$, $\eta_p^2 = .60$, $BF_{10} > 1000$, and a three-way interaction between semantic relatedness, list composition, and serial position, $F(5,395) = 8.82$, $\eta_p^2 = .10$, $BF_{10} = 631.72$. The three-way interaction reflects the increasing size of the semantic relatedness from serial position 1 to 6 in pure but not in mixed lists. The results also revealed no main effect of list composition, $F < 1$, $\eta_p^2 = .01$, $BF_{01} = 29.07$. Furthermore, unlike in Experiment 1, there was no interaction between semantic relatedness and list composition, $F < 1$, $\eta_p^2 = .01$, $BF_{01} = 29.41$. The absence of an interaction supports the equivalent semantic relatedness effect in pure and mixed lists. The results from the analyses also provide evidence against the interactions between semantic relatedness and serial position, $F < 1$, $\eta_p^2 = .01$, $BF_{01} > 1000$, and between list composition and serial position, $F(5,395) = 4.02$, $\eta_p^2 = .05$, $BF_{01} = 105.82$.

Free Scoring. As shown in Fig. 12, the results based on free scoring echo those based on strict serial scoring. Participants' performance was superior for semantically related words ($M = 0.80$, $SD = 0.12$) compared to unrelated words ($M = 0.71$, $SD = 0.16$) and performance was once again nearly identical for recall of words in pure lists ($M = 0.76$, $SD =$

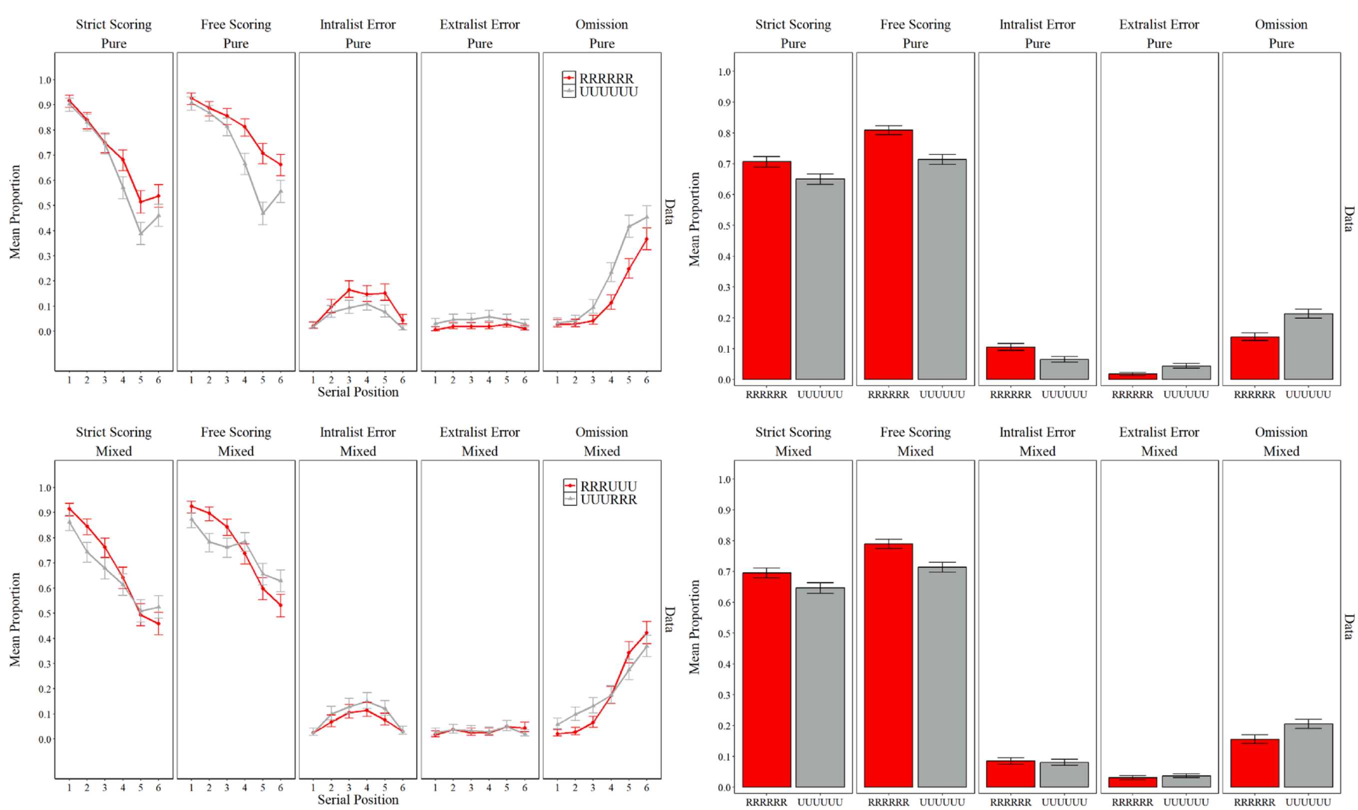

Fig. 12. Mean proportion of response as a function of list composition (pure lists, mixed lists), semantic relatedness (related, unrelated), serial positions (1 to 6) and scoring procedures (strict scoring, free scoring, intralist error, extralist error, omission) in Experiment 2. Error bars correspond to the 95% credible intervals.

0.14) and mixed lists ($M = 0.75$, $SD = 0.14$).

The results from the statistical analyses were in line with those with strict serial recall scoring. More exactly, there was a main effect of semantic relatedness, $F(1,79) = 69.54$, $\eta_p^2 = .47$, $BF_{10} > 1000$, a main effect of serial position, $F(5,395) = 81.82$, $\eta_p^2 = .51$, $BF_{10} > 1000$, and a three-way interaction between semantic relatedness, list composition, and serial position, $F(5,395) = 8.95$, $\eta_p^2 = .10$, $BF_{10} > 1000$. Again, there was no main effect of list composition, $F(1,79) = 1.23$, $\eta_p^2 = .02$, $BF_{01} = 19.74$, no interaction between semantic relatedness and list composition, $F(1,79) = 1.62$, $\eta_p^2 = .02$, $BF_{01} = 13.26$, anecdotal evidence in favour of an interaction between semantic relatedness and serial position, $F(5,395) = 5.98$, $\eta_p^2 = .07$, $BF_{10} = 1.38$, and no interaction between list composition and serial position, $F(5,395) = 3.56$, $\eta_p^2 = .04$, $BF_{01} = 53.13$.

Modelling Assessment. In this section, we briefly highlight the experimental results in relation to the model predictions presented in Fig. 11. Overall, the model had a good fit to the data across all scoring procedures ($R^2 = 0.90$). As shown, the model captured the main features of our empirical data with some minor differences. Before describing the predictions of primary interest, we briefly summarize exploratory predictions in relation to the results for transparency.

The model successfully captured the interactions for omissions as a function of RRRUUU versus UUURRR lists, showing more omissions for R than U as a function of serial position. Additionally, the model accurately predicted fewer omissions in related lists compared to unrelated lists in pure lists. At the overall level, when isolating R and U, the model predicted fewer intralist errors for related than unrelated lists in both pure and mixed lists; however, this was not observed in the data for mixed lists. At the serial position level, without isolating R and U, the model predicted, albeit with a very small magnitude, more intralist errors for unrelated than related items in mixed lists and the opposite in pure lists. This pattern echoes the empirical results and suggests that list organization can constrain order errors (e.g., Poirier et al., 2015; Kowialiewski et al., 2024). For extralist errors, the model predicted relatively similar or slightly higher levels of extralist errors for both pure and mixed lists, whereas the data showed slightly more extralist errors in related pure lists.

Regarding the prediction of interest, the model predicts that the benefit for related over unrelated lists was about equal for pure and mixed lists. The model predicts better recall of R over U items at almost all serial positions, with a reduced difference at the first serial position. This pattern is highlighted by the crossover interaction of the serial position functions for RRRUUU and UUURRR lists at or near the transition from serial positions 3 to 4. While the crossover is clearly visible in the free scoring data, it is strongly suggested by the difference in slopes of the serial position functions for the RRRUUU and UUURRR lists, although not immediately crossing over at the gap between serial positions 3 and 4 in the strict scoring data.

Discussion

Unlike Experiment 1, participants were better at recalling lists of semantically related words in both pure and mixed lists and the effect was not attenuated. In effect, unlike Experiment 1, there was no interaction between semantic condition and list composition, suggesting that the semantic relatedness effect in mixed lists was comparable when related items were presented in adjacent positions, an outcome consistent with Saint-Aubin et al. (2014) and Kowialiewski et al. (2021, 2022). However, unlike Kowialiewski et al. (2021, 2022), when carefully examining across both mixed and pure lists, we found little proactive effect of semantic relatedness. Specifically, in their study, there was an advantage for unrelated lists that preceded related lists relative to unrelated lists, but this was not clear in both the experiment and the simulations. This finding was not of particular theoretical interest, and further work might be needed before concluding that this is an important feature of memory. More importantly, for mixed lists, there was a clearer interaction between the blocked mixed list conditions and serial positions, as expected, especially in the free scoring analysis. Although the crossover point was not precisely at the shift from serial position 3 to 4, the slopes of the mixed list serial position functions provide good evidence for the predicted interaction. In summary, the key details of participants' performance were predicted by our model, and we conclude that it captures the important details of serial recall performance for pure and mixed lists of semantically related and unrelated words.

Demonstration 6: Presentation rate and semantic relatedness in serial recall

Our two previous experiments, combined with our simulations, provide good evidence that the models can predict many important features regarding the influence of list organization on how semantic information interacts with list structure. With our better understanding of list organization, we now focus on pure lists for further investigation into the influence of semantic relatedness. In this demonstration, we examine if the model can predict the influence of task difficulty on the semantic relatedness effect by manipulating presentation rate. As highlighted in the introduction, increasing task difficulty should increase the magnitude of the semantic relatedness effect (e.g., Neale & Tehan, 2007; Poirier & Saint-Aubin, 1995; Saint-Aubin et al., 2005). Our simulations of Saint-Aubin et al. (2005) showed that the model can postdict this pattern by reducing the learning base rate. However, we could not fully evaluate the influence of task difficulty due to some missing information in the materials and the data. Here, we address this issue with additional simulations and experimental investigations. Inspired by Coltheart and Langdon's (1998) manipulation with phonological similarity, we manipulated presentation rate, presenting words either at a rate of one every 114 ms in one group or one every 500 ms in another group. The influence of presentation rate on task difficulty is well-known (see e.g., Guitard & Cowan, 2023; Dauphinee et al., 2024), but the exact effects on semantic relatedness are not well documented.

As in previous experiments, we conducted simulations for serial recall of pure related (RRRRRR) and pure unrelated (UUUUUU) word lists using the same 24 pure related and 24 pure unrelated word lists from Experiments 1 and 2 (see Appendix C). We simulated performance by conducting 100 simulations of 24 serial recall trials for each of the RRRRRR and UUUUUU list structures, equivalent to 400 fictional participants. To simulate the influence of task difficulty, similar to our approach for simulating Saint-Aubin et al. (2005), we set a lower learning base rate (L = 0.125) for participants in the fast presentation group (114 ms) compared to the slower presentation group (L = 0.165). The other simulation parameters are available in Table 1.

Simulation results are shown in Fig. 13 for each scoring procedure. Our main interest was the magnitude of the difference, so we focus on the strict and free scoring procedures. The model predicts two key outcomes: first, better performance for the 500 ms presentation rate compared to the 114 ms rate; second, as expected based on prior experimental demonstrations, a larger difference between related and unrelated words at the faster presentation rate compared to the slower rate, indicating a two-way interaction between semantic relatedness and presentation rate. We now turn to an experimental test of the model predictions.

Experiment 3

Experiment 3 was similar to our previous experiments, except that participants studied only related and unrelated words in pure lists (RRRRRR vs. UUUUUU). Additionally, one group of participants studied the lists at a rate of one word every 114 ms, while another group studied the lists at a rate of one word every 500 ms (see also, Coltheart &

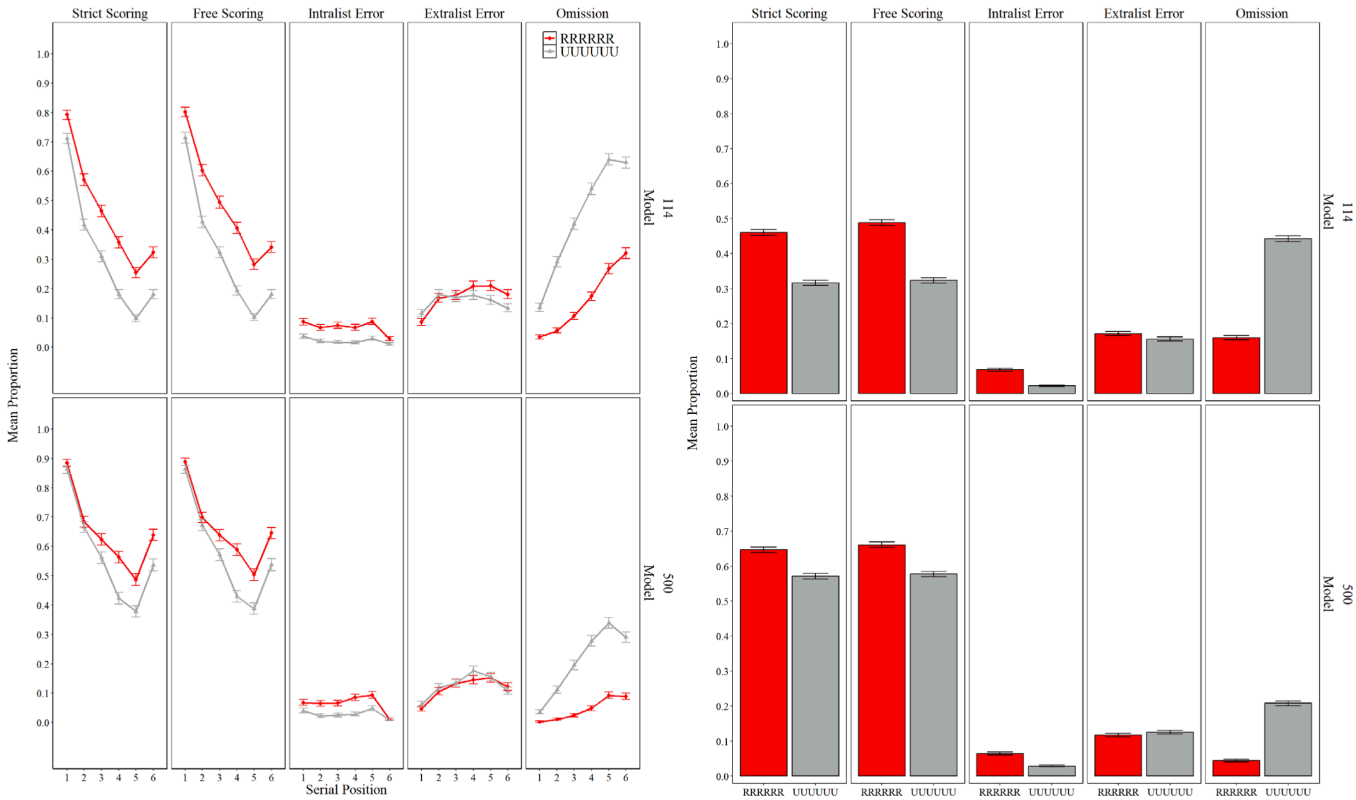

Fig. 13. Model simulations of Experiment 3. Although we present these simulations as predictions, the predictions are shown from simulations using parameters selected after running the experiment. Error bars correspond to the 95% credible intervals.

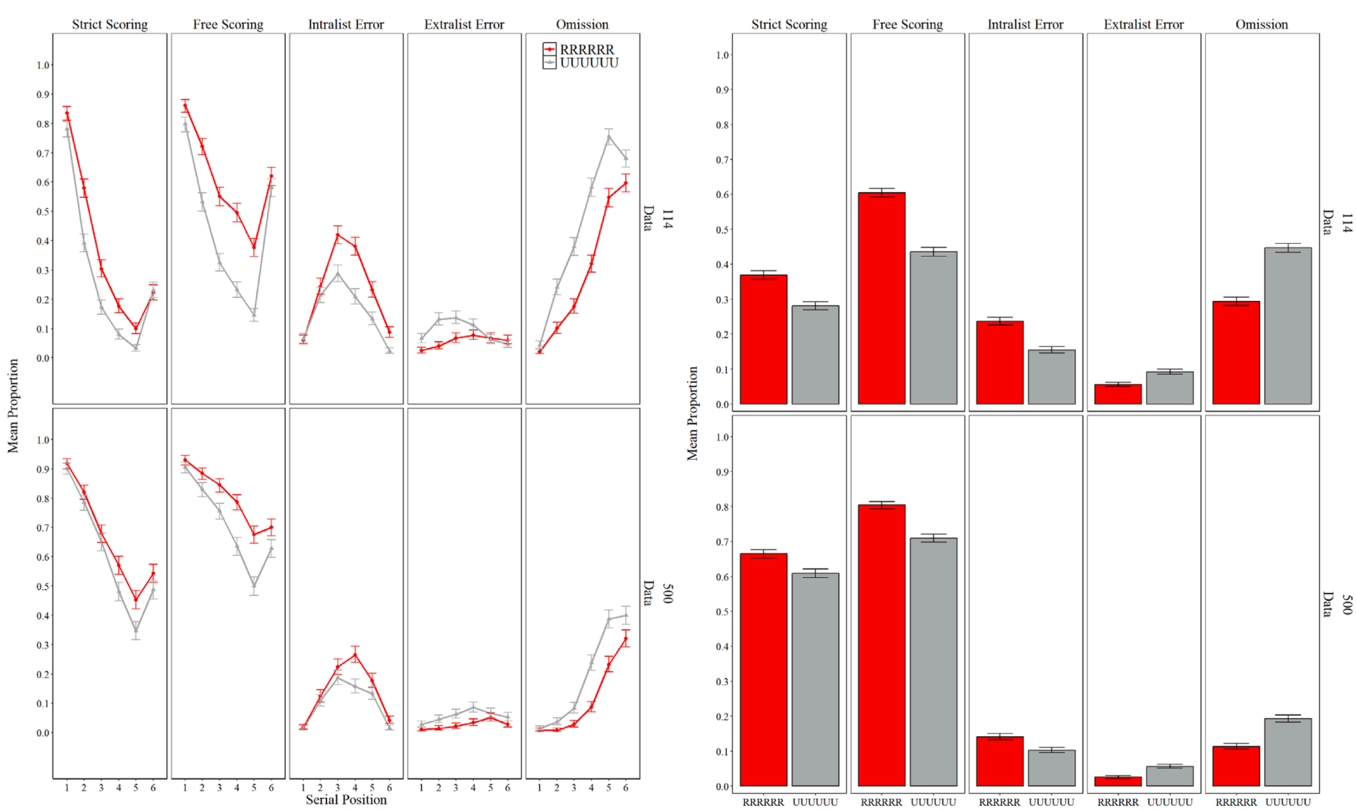

Fig. 14. Mean proportion of response as a function of list composition (pure lists, mixed lists), semantic relatedness (related, unrelated), serial positions (1 to 6) and scoring procedures (strict scoring, free scoring, intralist error, extralist error, omission) in Experiment 3. Error bars correspond to the 95% credible intervals.

Langdon, 1998).

Method

Participants

A total of 160 participants, who had not taken part in the previous experiments, were recruited via Prolific. Participants were divided equally into two presentation rate groups: 80 participants for the 114 ms rate and 80 participants for the 500 ms rate. The same inclusion criteria as in previous experiments were applied. In Experiment 3, the participants' mean age was 24.03 years (SD = 3.01, range 18–31). Of these, 73 participants self-identified as female, 79 as male, and 8 preferred not to specify their gender.

Materials, design, procedure, and data analysis

The materials, design, procedure, and data analysis were similar to those in Experiment 1 and Experiment 2, except that the presentation rate (114 ms, 500 ms) was manipulated instead of list composition. Participants studied 12 lists of related words and 12 lists of unrelated words (see Appendix C). As in previous experiments, each participant saw each word once, either in the unrelated or related conditions, with the conditions counterbalanced across participants.

Results

The results of Experiment 3 are presented in Fig. 14 for all scoring procedures. As shown in Fig. 14, performance was superior at the slower presentation rate (500 ms), and the difference between semantic related and unrelated lists was much larger at the faster presentation rate (114 ms).

We conducted a 2 (related vs. unrelated) x 2 (114 ms vs. 500 ms) x 6 (serial position) ANOVA to evaluate the descriptive results presented in Fig. 14. As in our previous experiments, our main focus was on strict scoring and free scoring. However, data and analyses for all scoring procedure can be found on the OSF page associated with the manuscript.

Strict Scoring. Of particular theoretical interest, the statistical analyses revealed a main effect of semantic relatedness, $F(1,158) = 60.89$, $\eta_p^2 = .28$, $BF_{10} > 1000$, a main effect of presentation rate, $F(1,158) = 60.89$, $\eta_p^2 = .28$, $BF_{10} > 1000$, and an interaction between semantic relatedness and presentation rate, $F(1,158) = 6.039$, $\eta_p^2 = .04$, $BF_{10} = 3.54$. These results confirm a larger difference between related and unrelated words at the faster presentation rate compared to the slower presentation rate. Of less theoretical interest, there was also a main effect of serial position, $F(5,790) = 195.46$, $\eta_p^2 = .\ 55$, $BF_{10} > 1000$, an interaction between serial position and presentation rate, $F(5,790) = 24.35$, $\eta_p^2 = .\ 13$, $BF_{10} > 1000$, an interaction between serial position and semantic relatedness, $F(5,790) = 17.57$, $\eta_p^2 = .\ 10$, $BF_{10} = 148.79$, and a three-way interaction, $F(5,790) = 13.15$, $\eta_p^2 = .\ 07$, $BF_{10} > 1000$.

Free Scoring. The results from the free scoring were comparable to those with strict scoring. Again there was a main effect of semantic relatedness, $F(1,158) = 229.49$, $\eta_p^2 = .\ 59$, $BF_{10} > 1000$, a main effect of presentation rate, $F(1,158) = 118.20$, $\eta_p^2 = .43$, $BF_{10} > 1000$, and an interaction between semantic relatedness and presentation rate, $F(1,158) = 30.70$, $\eta_p^2 = .16$, $BF_{10} > 1000$. There was also a main effect of serial position, $F(5,790) = 147.45$, $\eta_p^2 = .\ 48$, $BF_{10} > 1000$, an interaction between serial position and presentation rate, $F(5,790) = 40.05$, $\eta_p^2 = ..202$, $BF_{10} > 1000$, an interaction between serial position and semantic relatedness, $F(5,790) = 18.67$, $\eta_p^2 = .\ 11$, $BF_{10} > 1000$, and a three-way interaction, $F(5,790) = 5.98$, $\eta_p^2 = .\ 04$, $BF_{10} = 37.05$.

Modelling Assessment. Here we examined the experimental results in relation to the model predictions presented in Fig. 13. As can be seen by comparing Figs. 13 and 14, the model had a relatively good fit to the data across all scoring procedures ($R^2 = 0.83$), with some minor discrepancies. Before describing the predictions of primary interest, we briefly summarize other exploratory predictions in relation to the results for transparency.

For omissions, although the model overpredicted the difference between related and unrelated words, it captured the higher amount of omissions for the faster presentation rate compared to the slower rate. The model also captured the higher number of intralist errors for related words compared to unrelated words for both presentation rates. However, it did not accurately capture the higher amount of intralist errors for the faster presentation rate, which could likely be accounted for by adjusting the parameter d, which was kept constant in these simulations. The model predicted relatively no difference between related and unrelated words for extralist errors. However, in the data, there were slightly more extralist errors in related lists compared to unrelated lists. The model's performance for free scoring and strict scoring was relatively similar, while in the data, free scoring was much higher. This discrepancy might reflect parameter choices rather than important differences and is not of particular theoretical interest to the current study.

Regarding the prediction of interest, the model accurately predicted the benefit for related over unrelated lists. More importantly, the model captured the larger magnitude of the semantic relatedness effect at the fast presentation rate, although with a slightly smaller magnitude than observed in the experimental data. Overall, this provides clear evidence that the model can track this key feature with only small adjustments to the learning base rate.

Discussion

In Experiment 3, we demonstrated that presentation rate modulated the influence of semantic information on verbal short-term memory as expected based on our model. Specifically, in line with previous findings, a faster presentation rate was associated with lower overall memory performance (e.g., Guitard & Cowan, 2023; Dauphinee et al., 2024). More importantly, consistent with prior research on task difficulty, the magnitude of the semantic relatedness effect was larger at a faster presentation rate when there was more opportunity to benefit from the redintegration process (Neale & Tehan, 2007; Poirier & Saint-Aubin, 1995; Saint-Aubin et al., 2005).

Demonstration 7: Revisiting the detrimental effect of semantic relatedness on order information in immediate serial recall

The previous experiment showed that the model effectively captured the key features of how task difficulty influences the semantic relatedness in verbal short-term memory performance. Here, we aimed to revisit whether semantic information also influences order information. As reviewed in the introduction, a major contention is that semantic information affects item information but not order information. However, notable exceptions, such as the studies by Saint-Aubin et al. (2005) and Tse (2009, Tse et al., 2011), have found detrimental effects on order information.

In the second demonstration, we showed that by using lists nearly identical to those of Saint-Aubin et al. (2005), the model also predicted a detrimental effect of semantic relatedness on order information. Because of their theoretical question, Saint-Aubin et al. (2005) used a large number of participants, but they also had to implement an unusual between-participants manipulation of semantic relatedness and fixed-order lists, which have been shown to lead to atypical results in some cases (Guitard, Neath, & Saint-Aubin, 2023). Tse (2009, Tse et al., 2011) addressed some of these issues, but it remains unclear if the phenomenon is limited to particular stimuli or specific methodological choices, as has been the case with the syllable word length effect, where words with fewer syllables are better recalled (Guitard et al., 2018). For these reasons, we decided to conduct an additional demonstration of this important question.

In this study, the simulation procedures were similar to our approach for the study of Poirier et al. (2015), where we simulated each participant's specific random list. After completing the experiment with 160 participants, we simulated their exact lists in the same order (see Table 1 for the simulation parameters). Therefore, we conducted simulations for the serial recall of pure related (RRRRRRR) and pure unrelated (UUUUUUU) word lists, using the 26 pure related and 26 pure unrelated word lists that each participant studied. We simulated performance by conducting 4 simulations of each participant's specific list, equivalent to 640 fictional participants.

Simulation results are shown in Figs. 15, and 16, accompanied by the experimental data. Our main interest was to examine the influence of semantic information on order information. Like in previous experiments, we report the strict and free scoring procedures and add conditional order errors as in Saint-Aubin et al. (2005). As shown, the model predicts better memory for related words than unrelated words, as previously observed, and a higher number of order errors for related words compared to unrelated words. We now turn to an experimental test of the model predictions.

Experiment 4

In Experiment 4, we re-examined the influence of semantic information on order information using a new set of stimuli and a large number of participants.

Method

Participants

In this study, to ensure sufficient power to detect a credible difference between semantic related and unrelated words for order information, we doubled our sample size from 80 participants to 160 participants. A non-directional paired-sample sensitivity analysis conducted with G*Power 3.1.9.7 (Faul et al., 2007) and default parameters revealed that 160 participants would allow us to detect a small effect size of $d = 0.29$.

Therefore, a total of 160 participants who had not taken part in the previous experiments were recruited via Prolific, using the same inclusion criteria as in the previous experiments. In Experiment 4, the participants' mean age was 26.03 years (SD = 3.22, range 18–31). Of these, 75 participants self-identified as female, 84 as male, and 1 preferred not to specify their gender.

Materials, design, procedure, and data analysis

The materials, design, procedure, and data analysis were similar to those in previous experiments, with the following changes. A new set of 52 semantically related words taken from Chubala et al. (2019) was used (see Appendix D). Each participant completed 26 related lists of 7 words and 26 lists of unrelated words. The list length was increased to reduce the overall level of performance, as it is not possible to observe intralist errors if performance is near perfection, which is more likely for related than unrelated words with shorter list lengths. As in Neath et al. (2022), the order of the words within a list for each participant was randomized, and each participant saw each word only once in either the related or unrelated conditions to prevent any compensatory strategy. For unrelated words, as done by Neath et al., the words were randomly chosen from different categories. The other details were similar to our previous experiments.

Results

The results of Experiment 4 are presented in Figs. 15 and 16. As shown in Fig. 15, performance was superior for related words compared to unrelated words as in previous experiments. However, more importantly, Fig. 16 highlights the detrimental effects on order information as

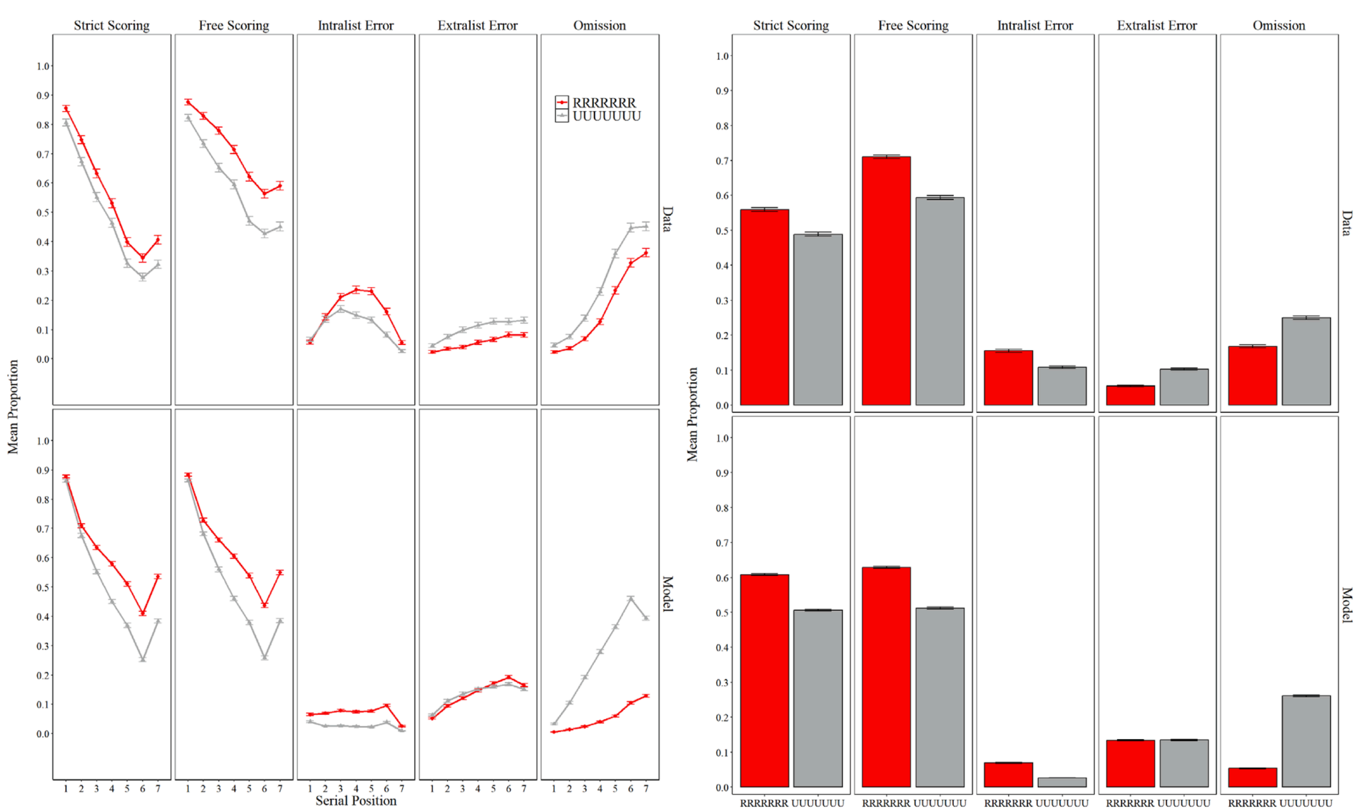

Fig. 15. The model's serial recall performance (identified by model) for related and unrelated lists for Experiment 4 (identified by data). The left panels show performance across serial positions according to each scoring procedures (strict scoring, free scoring, intralist error, extralist error, omission), while right panels display performance collapsed across serial positions for related and unrelated lists. Error bars correspond to the 95% credible intervals.

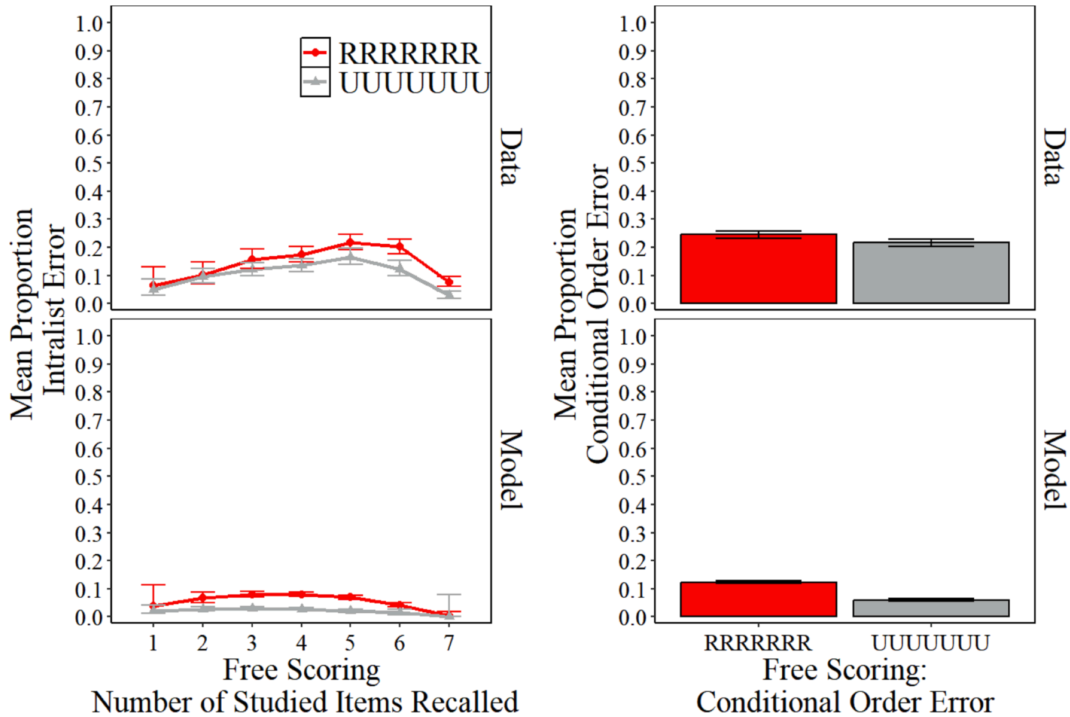

Fig. 16. The model's serial recall performance (identified by model) for related and unrelated lists for Experiment 4 (identified by data). The left panel show the number of intralist error as function of the number of studied items recalled based on the free scoring procedure, while right panels display condition order performance for related and unrelated lists.

a function of each item recalled and the overall conditional order error, indicating a small detrimental effect on order information.

We conducted a 2 (related vs. unrelated) x 7 (serial position) ANOVA to evaluate the descriptive results presented in Fig. 15 and a simple Bayesian paired-sample *t*-test for conditional order error presented in Fig. 16. For consistency with previous experiments and to ensure that we observed the standard beneficial effect, we report analyses for both strict scoring and free scoring. Data and analyses for all scoring procedures can be found on the OSF page associated with the manuscript.

Strict Scoring. As in previous experiments, there was a main effect of semantic relatedness, $F(1,159) = 153.25$, $\eta_p^2 = .49$, $BF_{10} > 1000$, serial position, $F(6,954) = 348.54$, $\eta_p^2 = .69$, $BF_{10} > 1000$, but no interaction between semantic relatedness and serial position, $F(6,954) = 1.85$, $\eta_p^2 = .01$, $BF_{01} > 1000$.

Free Scoring. The results of with free scoring echo the results with strict scoring, there was a main effect of semantic relatedness, $F(1,159) = 291.59$, $\eta_p^2 = .65$, $BF_{10} > 1000$, serial position, $F(6,954) = 174.68$, $\eta_p^2 = .52$, $BF_{10} > 1000$. However, this time there was interaction between semantic relatedness and serial position, $F(6,954) = 11.59$, $\eta_p^2 = .07$, $BF_{10} > 1000$. The latter interaction is not of theoretical interest to the current study and reflects the slightly larger benefit of semantic related information for items at the end of the lists.

Conditional Order. In this section, we investigated the influence of semantic information on order information. To control for the different opportunities for intralist errors, the number of order errors was divided by the number of items recalled in any position. This was done by dividing the number of intralist errors for each list by the number of items recalled in any position (Murdock, 1976; Saint-Aubin & Poirier, 1999b). Similar to our previous simulations with Saint-Aubin et al. (2005), when the number of items recalled was 0, the number of conditional order errors was also 0. The results are presented in Fig. 16.

We also present the distribution of intralist errors as a function of the number of studied items recalled based on the free scoring procedure. As shown in Fig. 16, participants made more conditional order errors in related lists (M = 0.25, SD = 0.16) than in unrelated lists (M = 0.22, SD = 0.18). The difference was small, but these results were confirmed by a simple non-directional Bayesian paired-sample *t*-test, $BF_{10} = 122.90$, Cohen's d = 0.31.

Modelling Assessment. In this section, we briefly review the model predictions in relation to the experimental results. As shown in Fig. 15, the model tracks almost all key features of memory performance across all scoring procedures. It captures the better performance for related lists compared to unrelated lists for both strict and free scoring. However, while the performance between strict and free scoring was relatively similar in the model, the actual data showed superior performance with free scoring. The model, like the data, produces more intralist errors for both related and unrelated lists and fewer omissions for related lists compared to unrelated lists. The model does not capture the higher magnitude of extralist errors for unrelated lists compared to related lists. Despite these minor discrepancies, the fit of the model was relatively good, with an $R^2 = 0.90$.

Discussion

One of the dominant conceptions of the influence of semantic information on verbal short-term memory performance is that semantic information primarily affects item information with minimal influence on order information. Here, contrary to this dominant view, but consistent with the results of Saint-Aubin et al. (2005) and Tse (2009, Tse et al., 2011), we have shown with a large number of participants that semantic information does affect order information, although the effect was relatively small. Although further work is needed to investigate this with different stimuli, it appears that semantic information does indeed affect order information. This has been a point of contention in the literature compared to other forms of similarity. For example, the detrimental effect of semantic relatedness is smaller than what has been observed with other shallow features (visual, orthographic, and phonological). However, the difference might be more apparent than real and could be related to the nature of feature space. As recently suggested by Caplan (2023) and Caplan and Guitard (2024a,b), orthographic and phonological information exist in a dense feature space due to the limited number of features needed to represent letters and phonemes which increase confusion, while semantic information exists in a large, sparse vector space with reduce confusion. Directly testing this idea is beyond the scope of this manuscript but would be of high importance in subsequent work. Overall, we encourage researchers to further evaluate the influence of semantic information on verbal short-term memory. Researchers should use different stimulus sets and include a larger number of participants to capture the small effect. In summary, semantic information affects memory for items and has a small detrimental effect on order information.

Demonstration 8: semantic relatedness and false memory

Across our previous demonstrations, we have shown that our model successfully captures many key aspects of the influence of semantic information on veridical memory performance of the studied lists. However, a unique feature of our model is its ability to examine what happens beyond the studied lists. As shown in Demonstration 3 with Poirier et al. (2015), the model offers a unique opportunity to investigate memory at the item level. By incorporating a lexicon based on a well-established model of semantics, it also captures the relationship between studied and unstudied words.

Therefore, to further highlight the value of our approach, this final demonstration explores how semantic information affects false memory, specifically extralist errors (e.g., recalling information that was not studied). Here, we aim to determine whether the model can predict the influence of semantic information on the likelihood of recalling a semantic critical lure—a word that was not studied but is semantically related to the studied words.

As briefly highlighted in the introduction, there is a rich history of studying false memory in free recall and recognition tasks (see Chang & Brainerd, 2021, for a review), but relatively limited research in serial recall (e.g., Tehan, 2010). One key signature that has received recent computational interest is the increased likelihood of producing a specific critical lure when the number of related studied words associated with the critical lure also increases (Spens & Burgess, 2024), a finding initially demonstrated by Robinson and Roediger (1997) in recognition and free recall. Inspired by their work, we aimed to conduct a simple demonstration to evaluate if the model can also track this important phenomenon in serial recall.

We conducted simulations for serial recall of 11 lists of 10 words each, with each list varying in the number of words associated with the critical non-presented lure "SLEEP." Specifically, we systematically constructed lists based on Robinson and Roediger's material, in which there were 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 related words associated with the critical lure (see Appendix E).

We simulated performance by conducting 500 simulations of the 11 lists (see Table 1 for the simulation parameters). Additionally, we took the opportunity to simulate the results with LSA and fastText, a more recent DSM that also incorporates subword information (Bojanowski et al., 2017; Petilli et al., 2024). Evaluating the model with representations from different DSMs is relatively simple and allows us to show the flexibility of the model. Simulation results are shown in Fig. 17, accompanied by the empirical results. The results from our simulation are relatively straightforward: as the number of related words increases, the likelihood of falsely recalling the word "SLEEP" (the critical lure) also increases—a finding captured by both LSA and fastText. We now turn to an experimental test of the model predictions.

Experiment 5

In Experiment 5, inspired by Robinson and Roediger (1997), we

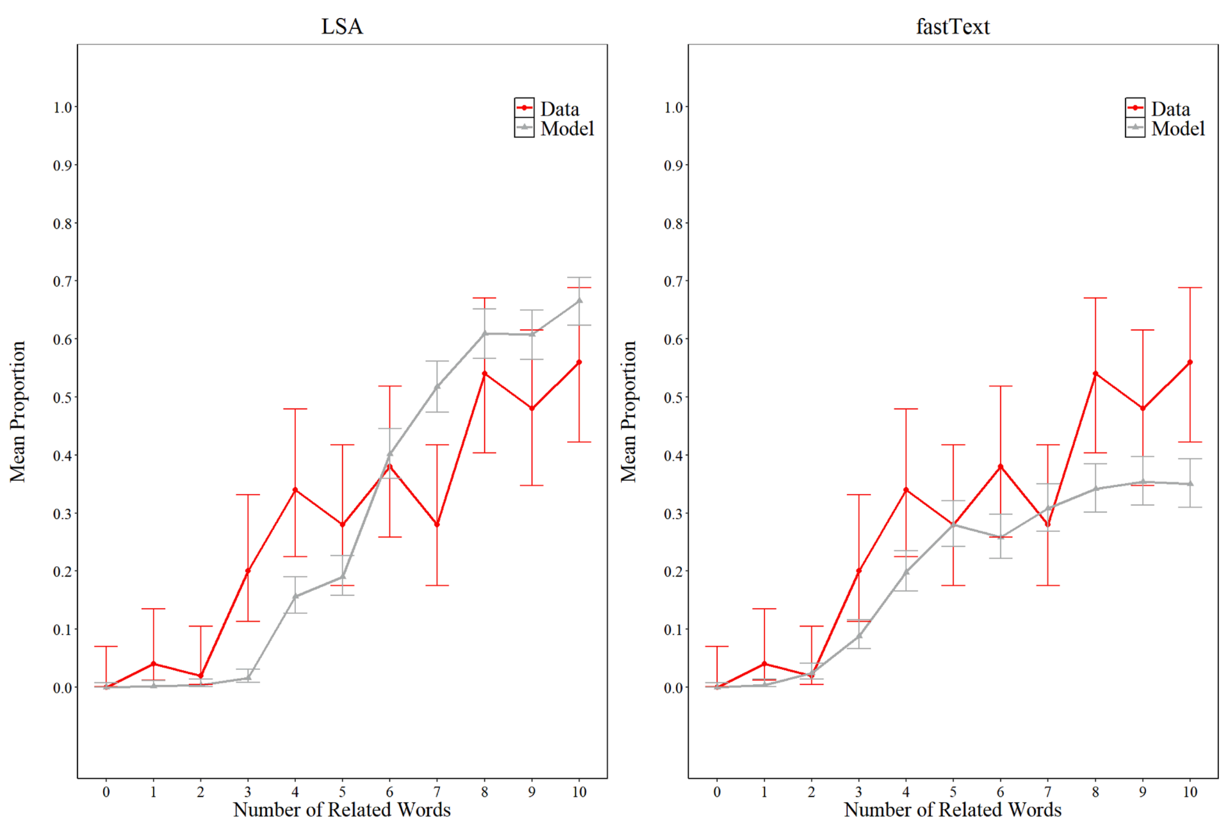

Fig. 17. The mean proportion of participants recalling the critical word "SLEEP" in Experiment 5 is shown as a function of the number of related words in the studied lists. The data are presented in red and simulations in grey. The simulation results using LSA are presented in the left panel, and those using fastText are presented in the right panel. Error bars correspond to the 95% credible intervals.

examined the likelihood of producing a specific critical lure based on the number of studied associates. We systematically manipulated the number of related words associated with the critical lure in the lists, with participants receiving lists containing 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 related words.

Method

Participants

A total of 550 participants, with 50 participants per condition (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 related words), who had not taken part in the previous experiments, were recruited via Prolific using the same inclusion criteria. The participants' mean age was 25.53 years (SD = 3.29, range 18–31). Of these, 199 participants self-identified as female, 340 as male, and 11 preferred not to specify their gender. The total of 550 participants was used to ensure stable estimates for comparison with our simulations.

Materials, Design, Procedure, and data analysis

The materials, design, procedure, and data analysis were similar to the previous experiments, with the following changes. Each participant was randomly allocated to one of the 11 conditions. All participants completed a single trial to minimize strategy contamination between conditions. They studied 10 words presented sequentially at a rate of one word per second and had to recall them in their presentation order, as in the previous experiments. The experiment lasted approximately 2 min.

The stimuli were taken from Robinson and Roediger (1997), based on Roediger and McDermott (1995). From these stimuli, we created 11 lists varying in the number of related words associated with the critical lure "SLEEP," which has been shown to produce reliable false memories. We chose to focus on one critical lure for a systematic exploration, but we believe the results should be stable across various critical lures and materials, as shown in the original demonstration. Like in Experiments 1, 2, and 3, the order of the words within each list was fixed (see Appendix E).

For the analyses, we focused on the proportion of participants that recalled the critical lure as a function of condition.

Results

The results of Experiment 5 are presented in Fig. 17, along with the simulation results using both LSA and fastText. As shown, when the number of words associated with the critical lure increases in the list, the probability of recalling that critical word also increases.

Although the results are clear, we conducted a one-way ANOVA with condition as the sole factor to confirm the descriptive results. As expected, there was a main effect of condition, $F(10,539) = 12.09$, $\eta_p^2 = .18$, $\mathrm{BF}_{10} > 1000$.

Modelling Assessment. As illustrated in Fig. 17, the model had a good fit to the data for both the simulation with LSA ($R^2 = 0.81$) and the simulation with fastText ($R^2 = 0.87$). The model captured the key feature of the data: an increase in the probability of recalling the critical lure as a function of the number of related studied words. LSA slightly overpredicted with larger number of related words (e.g., 7) and slightly underpredicted for lower number of related words (e.g., 3, 4, 5), while fastText slightly underpredicted with larger number of related words (e.g., 8 and 10). However, the goal of this demonstration was not to decide which semantic model is better, but to show that the model can integrate different semantic models of memory and predict the false recall.

Discussion

In the false memory literature, it is well-established that increasing the number of related words associated with a non-presented word increases the likelihood of participants falsely recalling that word (Robinson & Roediger, 1997). In our model, because we have a lexicon, and memory is reconstructive, if all information is similar to a critical lure, comparison between the echo and the critical lure in lexicon will be more similar, increasing the likelihood of it being falsely retrieved above presented words. Although this was a simple demonstration, it shows for the first time to our knowledge that in serial recall, as in recognition and free recall, the same empirical pattern emerges, and our simple model can capture these results.

General discussion

Over the years, numerous empirical findings on the influence of semantic information on verbal short-term memory performance have emerged and painted a complex portrait (e.g., Crowder, 1979; Oberauer et al., 2018). Reconciling these findings under a common model has been challenging. Here, we presented an attempt to capture some of these important findings by proposing a simple computational model of memory, eCFM. This model was designed to bridge the gaps between semantic and episodic memory models, moving toward a more holistic understanding of memory. In the next section, we briefly summarized the simulations and empirical work conducted in this manuscript.

Empirical and computational summary

In this manuscript, we have revisited a large number of phenomena through computational, empirical, or combined investigations. We first revisited the semantic relatedness paradox—a benefit of semantic relatedness in serial recall, but not in serial reconstruction—by simulating the results of Neath et al. (2022). Our simulations captured the overall pattern of results. However, in some instances, the model predicted a small detrimental effect, leading to further exploration at the participant level in our supplementary analysis. This revealed that almost half the participants experienced a detrimental effect of semantic information, aligning well with our model's simulations. Additional simulations revealed that the reduce beneficial influence of semantic information in serial reconstruction, at least in our implementation, was largely due to task characteristics and lexicon size. As the lexicon size increased, the effect became positive, similar to serial recall.

We then re-examined the influence of semantic information on order information by revisiting the study by Saint-Aubin et al. (2005), which showed a detrimental effect of semantic information on order information. Our simulations, using almost identical lists, also predicted this detrimental effect. In Experiment 4, we conducted a study with 160 participants and a new set of stimuli, providing clear evidence of a detrimental effect for related lists compared to unrelated lists. These results warrant further consideration of the influence of semantic information on order information, highlighting the need for further investigation of this important finding. Although the effect is much smaller than that observed with phonological related information (e.g., Roodenrys et al., 2022), this difference might be due to the nature of feature spaces. Orthographic and phonological features likely result in dense representations of verbal information due to their limited number of features (e.g., letters and phonemes), whereas semantic features might create sparse representations of verbal information given their near-infinite number of features to represent information semantically (e.g., Buchanan et al., 2019; McRae et al., 2005). This suggests that shallow features would yield larger detrimental effects, likely due to an interaction between features encoded, features interference, feature space, and feature retrieve, as recently suggested in recognition memory (e.g., Caplan, 2023; Caplan & Guitard, 2024a,b). However, further studies are needed to directly assess feature space interactions with well-matched stimuli and larger participant numbers.

We revisited the pattern of migration errors, where an item is recalled. The model provided a good fit to the results of Poirier et al. (2015) and our investigation in Experiment 4. Overall, it seems clear that semantic information can have a specific impact on how order information is recalled (e.g., see also Kowialiewski et al., 2021, 2024).

In our experimental investigations, we demonstrated how a recall model that integrates a lexicon of vector-based representations for word meanings was able to generate predictions for mixed lists, which we tested in two experimental studies. The influence of semantic relatedness was slightly diminished in mixed alternating lists (URURUR or RURURU) in Experiment 1 compared to pure lists. In contrast, Experiment 2 showed a similar effect with blocked mixed lists (RRRUUU or UUURRR). The smaller sawtooth pattern observed in Experiment 1 aligns with the findings of Saint-Aubin et al. (2014), who reported a reduced semantic relatedness effect with an increased number of words presented between related words. This pattern differs from the word frequency effect, which, unlike the semantic relatedness effect, was eliminated in a mixed list design, as indicated by Hulme et al. (2003) and Morin et al. (2006).

The models also made interesting predictions regarding presentation rate as a proxy for task difficulty. The model and experimental results show that the magnitude of the semantic relatedness effect was larger at a faster presentation rate compared to a slower rate. This result is consistent with previous demonstrations and adds new empirical support for the influence of presentation rate (Neale & Tehan, 2007; Poirier & Saint-Aubin, 1995; Saint-Aubin et al., 2005).

Lastly, our models predicted that increasing the number of related words in a studied list to a critical related word that was not presented would make the latter more likely to be recalled. This simple prediction was successfully tested in Experiment 5. These results provide growing support for this classic finding obtained in free recall and recognition (Robinson & Roediger, 1997), but this time, for the first time to our knowledge, in serial recall. This opens interesting avenues, showing that a model with a lexicon can make precise predictions for false recall, demonstrating the value of embedding a lexicon to predict information that was never presented. As we also demonstrated in our simulations of Poirier et al. (2015) and the supplementary analysis of Experiment 4, the model can be directly assessed at the item level and is affected by different materials in a similar way humans are. This opens exciting precision into memory research, moving beyond average curve fitting to gain new insights into human cognitive processes.

In summary, we have shown or confirmed several key findings: the influence of semantic related information was larger in serial recall across almost all participants and small or slightly negative in serial reconstruction for many participants; semantic information appears to have a small detrimental effect on order information; list organization modulated but did not abolish the influence of semantic information in alternating or blocked lists; presentation rate, similar to previous demonstrations of task difficulty, had a large effect on the size of the semantic relatedness effect, with a larger effect observed with lower memory performance; specific list organization, as shown by Poirier et al. (2015), can affect the pattern of migration errors; and increasing the number of related words associated with a non-presented word affects the likelihood of recalling that critical lure in immediate serial recall.

Overall, our numerous demonstrations and empirical studies corroborated the general predictions of our model and represent a novel contribution to our database on memory performance and our theoretical understanding of the impact of participants' lexicon on verbal short-term memory performance. This work underscores the importance of embedding a lexicon in memory models to predict both veridical and false performance. It is short-sighted to explain performance in short-term memory tasks without considering the influence of long-term memory and unrealistic to have randomly generated vectors that are artificially created without a theoretical justification. In that sense, our account connects tightly to Cowan's (1988, 2019; Cowan et al., 2024) work on embedded memory systems and argues that including a well-defined theoretically driven lexicon is a necessary next step in developing a complete model of cognition.

Theoretical implications

As mentioned in the introduction, the dominant account of semantic information that inspired the integration of semantic and episodic models of memory in this study is the redintegration hypothesis. The redintegration hypothesis posits that memory retrieval is a reconstructive process based on the parallel retrieval of memory traces, significantly influencing how memory researchers conceptualize recall (Schweickert, 1993; Lewandowsky, 1999) and understand the impact of semantic relatedness on memory performance (Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999ab). According to this hypothesis, semantically related words facilitate the redintegration of degraded memory representations during retrieval, thereby aiding recall. Specifically, the redintegration process is thought to be particularly beneficial, or perhaps only beneficial, in situations where item information is critical, such as in serial recall or when the task is very difficult (e.g., Neale & Tehan, 2007), but not in serial reconstruction. This perspective offers a compelling explanation for the positive influence of semantically related words on enhancing serial recall and the limited influence in serial reconstruction. Our model provides a practical and effective example of how this hypothesis can be operationalized and sheds new light on the influence of semantic information, such as the small negative effect in order reconstruction observed for a large number of participants. Therefore, our work has significant implications for this dominant hypothesis, offering a formal implementation that bridges the gap between a dominant theoretical narrative and a comprehensive formal account of the phenomenon.

Interactive Activation Model. Recently, an alternative proposition has been put forward to account for the influence of semantic information, known as the interactive activation model (Kowialiewski et al., 2021; Kowialiewski & Majerus, 2020). In this neural network model, the beneficial influence of semantic relatedness is explained by the principles of spreading activation (e.g., Collins & Loftus, 1975). Specifically, semantically related words, through a process of interactive activation in the network, achieve stronger activation relative to semantically unrelated words, which do not benefit from this iterative activation process. This model represents a significant departure from traditional verbal accounts of memory and provides important insights and novel possibilities for understanding the influence of semantic relatedness. In its current form, despite this significant advancement, it remains unclear whether the model can account for all the findings presented here or how it could generate item-level predictions, such as specific false memories and accurate memory performance, without further modifications.

However, rather than debating which model is the "best," a more constructive approach is to integrate key assumptions from different models and evaluate them. For instance, one could incorporate the iterative spreading activation function for retrieval proposed by Hintzman (1986) into our model (see Equation 5 on p. 422). The overall goal is to advance the field by ensuring that initial good work also functions within a more comprehensive system, allowing us to determine whether the artificial processes created in a controlled environment remain valid. Such integration would likely provide deeper insights and new predictions that were not previously conceivable.

Embedded Computational Framework of Memory. In this manuscript, we present a simple theoretical alternative to previously described accounts. Specifically, we propose one of the few recall models that utilize vector-based representations for words. This vector-based approach has been highly successful in accounting for complex phenomena such as false recognition (e.g., Johns et al., 2012; Reid & Jamieson, 2023), release from proactive interference (e.g., Mewhort et al., 2018), free recall (e.g., Sirotin et al., 2005; Kimball et al., 2007), and other benchmark memory phenomena (see Jamieson et al., 2022, for a review).

By incorporating vector-based lexical representations, our model can account for a wide range of phenomena in both serial recall and serial reconstruction while having a better account of the complexity and richness of language experiences (Johns & Jones, 2010). This simple model provides a crucial mechanistic explanation for the influence of semantic information in verbal short-term memory. Although further work is needed to refine the model, we believe that integrating a lexicon of word representations into a mechanistic computational model of storage and retrieval is a valuable method for advancing and more precisely evaluating our understanding of human memory.

This approach allows us to analyze specific word lists and equips the model to handle serial recall by including a semantic memory component, not just a memory of the study list. This method offers a productive way to understand the scaled-up interactions between knowledge and memory, which have been under-investigated thus far. While we are not the first to advocate for this integration, the work presented here reinforces a growing consensus that the field is ready to build increasingly sophisticated models of memory and cognition (e.g., Chang & Johns, 2023; Chubala et al., 2016; Johns et al., 2012; Kimball et al., 2007; Mewhort et al., 2018; Monaco et al., 2007; Morton & Polyn, 2016; Osth et al., 2020; Osth & Zhang, 2023; Polyn et al., 2009; Reid & Jamieson, 2023; Steyvers, 2000).

Future directions

We have proposed a relatively simple model of memory that combines assumptions from episodic memory models (MINERVA 2: Hintzman, 1986) and semantic models (LSA: Landauer & Dumais, 1997) to account for the influence of semantic information in verbal short-term memory. While we have demonstrated that the model effectively captures the influence of semantic information across many demonstrations, it is clear that we are far from a complete theory of memory. Here, we highlight several important future directions.

Representations. In the current study, we focused primarily on semantic information to represent relationships within participants' lexicons. However, a comprehensive representation of verbal information includes many more features, such as early visual features (e.g., lines and shapes), orthographic features (e.g., letters), and phonological features (e.g., phonemes), among others, to capture the complexity of our verbal experience. Additionally, we assumed that each participant has a static and equivalent lexicon. This is a simplification, as individual experiences shape our lexicons. However, promising ongoing work is currently being undertaken that can refine our representations and make novel dynamic predictions (see e.g., Aujla, 2021; Jamieson et al., 2018; Johns & Jamieson, 2019). We have shown that the model can integrate French and English lexicons; addressing the linguistic complexity of bilingual individuals is a natural progression for enabling precise predictions, but will required further work. Overall, our ongoing work aims to develop a comprehensive representation including orthographic, phonological, and semantic information (see e.g., Guitard et al., 2024; Reid et al., 2024), while also capturing the dynamic reconstructive nature of individual linguistic experiences (e.g., Jamieson et al., 2018).

Integration. The eCFM exemplifies the benefits of embedding a lexicon into a memory model, which could be extended to a wide range of models (e.g., Brown et al., 2007; Nairne, 1988; Murdock, 1974; Saint-Aubin et al., 2021). Adding such a comprehensive lexicon increases computational complexity, but the tradeoff for enhanced precision and knowledge is undeniable. While our eCFM can be based on different combinations of memory and language models, a notable advantage of using MINERVA 2 (Hintzman, 1986) is its successful extension across various paradigms, including frequency judgement, recognition, categorization, cued recall, implicit learning, associative learning, and heuristic decision making (see Jamieson et al., 2022 for review). Although more work is needed to fully evaluate the integration across paradigms, we believe this is an important future direction for bridging memory paradigms with lexicon-based models.

We strongly believe that to move toward a general theory of memory, we need, as suggested by Hebb (1949), to break down silos within our field and integrate mechanisms to achieve a complete and precise understanding of human memory across phenomena and tasks (e.g., recognition, reconstruction, recall, cued recall, learning, etc.). Humans do not need a new brain for every task, and neither should our memory models.

Conclusion

Our overall goal was to account for the influence of semantic information on short-term verbal performance. To do so, we presented a computational model that integrates advances in human memory theory and computational linguistics, demonstrating its ability to capture a wide range of phenomena at the overall level while also providing unique item-level assessments. We also showed that our simple model can generate predictions for new studies and found empirical support for the model's general predictions. Overall, this study suggests that integrating an existing vector-based approach for word knowledge into an existing process model for memory can offer a fruitful and insightful tool for understanding human memory performance.

CRediT authorship contribution statement

Dominic Guitard: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Jean Saint-Aubin:** Writing – review & editing, Funding acquisition, Conceptualization. **J. Nick Reid:** Writing – review & editing, Methodology. **Randall K. Jamieson:** Writing – review & editing, Software, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The stimuli, the data, the codes, and the R markdowns are available on the Open Science Framework project (OSF: https://osf.io/zxbct/).

Acknowledgements

While working on this manuscript, Dominic Guitard was supported by an Experimental Psychology Society Small Grant. During the preparation of the manuscript, J. Nick Reid, Jean Saint-Aubin, and Randall K. Jamieson were supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). We would like to acknowledge Ian Neath for his support with data collection for Experiment 4.

Appendix A

Stimuli lists used in Experiment 1

Related (RRRRRR) Lists					
Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	IRON	COPPER	LEAD	BRASS	BRONZE
HOUR	SECOND	MINUTE	YEAR	DAY	MONTH
UNCLE	AUNT	COUSIN	MOTHER	FATHER	SISTER
MAGAZINE	BOOK	NEWSPAPER	JOURNAL	NOVEL	TEXTBOOK
SERGEANT	GENERAL	CAPTAIN	COLONEL	ADMIRAL	PRIVATE
DIAMOND	RUBY	EMERALD	SAPPHIRE	PEARL	JADE
Set 02					
COTTON	SILK	WOOL	SATIN	LINEN	FLANNEL
BLUE	RED	GREEN	YELLOW	PURPLE	BROWN
LEG	ARM	FINGER	ELBOW	ANKLE	KNEE
APPLE	BANANA	GRAPE	PEAR	PEACH	PLUM
GUN	PISTOL	SWORD	RIFLE	GRENADE	MISSILE
HOUSE	APARTMENT	MANSION	TENT	CABIN	HUT
Set 03					
TORNADO	HURRICANE	LIGHTNING	EARTHQUAKE	TYPHOON	BLIZZARD
SHIRT	SOCKS	HAT	JACKET	SWEATER	JEANS
WINDOW	DOOR	FLOOR	WALL	ROOF	STAIRS
DRUMS	GUITAR	FLUTE	PIANO	TRUMPET	VIOLIN
EAGLE	ROBIN	HAWK	CROW	SEAGULL	PIGEON
CAR	BUS	TRUCK	TRAIN	VAN	TAXI
Set 04					
CARROT	PEAS	CORN	SPINACH	BEANS	CABBAGE
FLY	ANT	BEE	MOSQUITO	BUTTERFLY	WASP
SALMON	TROUT	CATFISH	TUNA	SHARK	COD
FRANCE	MEXICO	GERMANY	SPAIN	ITALY	BRAZIL
RAP	CLASSICAL	JAZZ	ROCK	PUNK	FOLK
JUICE	COFFEE	WATER	TEA	LEMONADE	SODA

Unrelated (UUUUUU) Lists					
Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	HOUR	UNCLE	MAGAZINE	SERGEANT	DIAMOND
IRON	SECOND	AUNT	BOOK	GENERAL	RUBY
COPPER	MINUTE	COUSIN	NEWSPAPER	CAPTAIN	EMERALD
LEAD	YEAR	MOTHER	JOURNAL	COLONEL	SAPPHIRE
BRASS	DAY	FATHER	NOVEL	ADMIRAL	PEARL
BRONZE	MONTH	SISTER	TEXTBOOK	PRIVATE	JADE
Set 02					
COTTON	BLUE	LEG	APPLE	GUN	HOUSE
SILK	RED	ARM	BANANA	PISTOL	APARTMENT
WOOL	GREEN	FINGER	GRAPE	SWORD	MANSION
SATIN	YELLOW	ELBOW	PEAR	RIFLE	TENT
LINEN	PURPLE	ANKLE	PEACH	GRENADE	CABIN
FLANNEL	BROWN	KNEE	PLUM	MISSILE	HUT
Set 03					
TORNADO	SHIRT	WINDOW	DRUMS	EAGLE	CAR
HURRICANE	SOCKS	DOOR	GUITAR	ROBIN	BUS
LIGHTNING	HAT	FLOOR	FLUTE	HAWK	TRUCK
EARTHQUAKE	JACKET	WALL	PIANO	CROW	TRAIN
TYPHOON	SWEATER	ROOF	TRUMPET	SEAGULL	VAN
BLIZZARD	JEANS	STAIRS	VIOLIN	PIGEON	TAXI
Set 04					
CARROT	FLY	SALMON	FRANCE	RAP	JUICE
PEAS	ANT	TROUT	MEXICO	CLASSICAL	COFFEE
CORN	BEE	CATFISH	GERMANY	JAZZ	WATER
SPINACH	MOSQUITO	TUNA	SPAIN	ROCK	TEA
BEANS	BUTTERFLY	SHARK	ITALY	PUNK	LEMONADE
CABBAGE	WASP	COD	BRAZIL	FOLK	SODA

Mixed (RURURU) Lists					
Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	BOOK	COPPER	SAPPHIRE	BRASS	PRIVATE
HOUR	GENERAL	MINUTE	JOURNAL	DAY	JADE
UNCLE	RUBY	COUSIN	COLONEL	FATHER	TEXTBOOK
MAGAZINE	IRON	NEWSPAPER	MOTHER	NOVEL	MONTH
SERGEANT	SECOND	CAPTAIN	LEAD	ADMIRAL	SISTER
DIAMOND	AUNT	EMERALD	YEAR	PEARL	BRONZE
Set 02					
COTTON	BANANA	WOOL	TENT	LINEN	MISSILE
BLUE	PISTOL	GREEN	PEAR	PURPLE	HUT
LEG	APARTMENT	FINGER	RIFLE	ANKLE	PLUM
APPLE	SILK	GRAPE	ELBOW	PEACH	BROWN
GUN	RED	SWORD	SATIN	GRENADE	KNEE
HOUSE	ARM	MANSION	YELLOW	CABIN	FLANNEL
Set 03					
TORNADO	GUITAR	LIGHTNING	TRAIN	TYPHOON	PIGEON
SHIRT	ROBIN	HAT	PIANO	SWEATER	TAXI
WINDOW	BUS	FLOOR	CROW	ROOF	VIOLIN
DRUMS	HURRICANE	FLUTE	WALL	TRUMPET	JEANS
EAGLE	SOCKS	HAWK	EARTHQUAKE	SEAGULL	STAIRS
CAR	DOOR	TRUCK	JACKET	VAN	BLIZZARD
Set 04					
CARROT	MEXICO	CORN	TEA	BEANS	FOLK
FLY	CLASSICAL	BEE	SPAIN	BUTTERFLY	SODA
SALMON	COFFEE	CATFISH	ROCK	SHARK	BRAZIL
FRANCE	PEAS	GERMANY	TUNA	ITALY	WASP
RAP	ANT	JAZZ	SPINACH	PUNK	COD
JUICE	TROUT	WATER	MOSQUITO	LEMONADE	CABBAGE

Mixed (URURUR) Lists					
Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
MAGAZINE	IRON	EMERALD	LEAD	ADMIRAL	BRONZE
SERGEANT	SECOND	NEWSPAPER	YEAR	PEARL	MONTH
DIAMOND	AUNT	CAPTAIN	MOTHER	NOVEL	SISTER
STEEL	BOOK	COUSIN	JOURNAL	DAY	TEXTBOOK
HOUR	GENERAL	COPPER	COLONEL	FATHER	PRIVATE
UNCLE	RUBY	MINUTE	SAPPHIRE	BRASS	JADE
Set 02					
APPLE	SILK	MANSION	SATIN	GRENADE	FLANNEL
GUN	RED	GRAPE	YELLOW	CABIN	BROWN
HOUSE	ARM	SWORD	ELBOW	PEACH	KNEE
COTTON	BANANA	FINGER	PEAR	PURPLE	PLUM
BLUE	PISTOL	WOOL	RIFLE	ANKLE	MISSILE
LEG	APARTMENT	GREEN	TENT	LINEN	HUT
Set 03					
DRUMS	HURRICANE	TRUCK	EARTHQUAKE	SEAGULL	BLIZZARD
EAGLE	SOCKS	FLUTE	JACKET	VAN	JEANS
CAR	DOOR	HAWK	WALL	TRUMPET	STAIRS
TORNADO	GUITAR	FLOOR	PIANO	SWEATER	VIOLIN
SHIRT	ROBIN	LIGHTNING	CROW	ROOF	PIGEON
WINDOW	BUS	HAT	TRAIN	TYPHOON	TAXI
Set 04					
FRANCE	PEAS	WATER	SPINACH	PUNK	CABBAGE
RAP	ANT	GERMANY	MOSQUITO	LEMONADE	WASP
JUICE	TROUT	JAZZ	TUNA	ITALY	COD
CARROT	MEXICO	CATFISH	SPAIN	BUTTERFLY	BRAZIL
FLY	CLASSICAL	CORN	ROCK	SHARK	FOLK
SALMON	COFFEE	BEE	TEA	BEANS	SODA

Appendix B

Stimuli lists used in Experiment 2.

Related (RRRRRR) Lists Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	COPPER	BRASS	IRON	LEAD	BRONZE
HOUR	MINUTE	DAY	SECOND	YEAR	MONTH
UNCLE	COUSIN	FATHER	AUNT	MOTHER	SISTER
MAGAZINE	NEWSPAPER	NOVEL	BOOK	JOURNAL	TEXTBOOK
SERGEANT	CAPTAIN	ADMIRAL	GENERAL	COLONEL	PRIVATE
DIAMOND	EMERALD	PEARL	RUBY	SAPPHIRE	JADE
Set 02					
COTTON	WOOL	LINEN	SILK	SATIN	FLANNEL
BLUE	GREEN	PURPLE	RED	YELLOW	BROWN
LEG	FINGER	ANKLE	ARM	ELBOW	KNEE
APPLE	GRAPE	PEACH	BANANA	PEAR	PLUM
GUN	SWORD	GRENADE	PISTOL	RIFLE	MISSILE
HOUSE	MANSION	CABIN	APARTMENT	TENT	HUT
Set 03					
TORNADO	LIGHTNING	TYPHOON	HURRICANE	EARTHQUAKE	BLIZZARD
SHIRT	HAT	SWEATER	SOCKS	JACKET	JEANS
WINDOW	FLOOR	ROOF	DOOR	WALL	STAIRS
DRUMS	FLUTE	TRUMPET	GUITAR	PIANO	VIOLIN
EAGLE	HAWK	SEAGULL	ROBIN	CROW	PIGEON
CAR	TRUCK	VAN	BUS	TRAIN	TAXI
Set 04					
CARROT	CORN	BEANS	PEAS	SPINACH	CABBAGE
FLY	BEE	BUTTERFLY	ANT	MOSQUITO	WASP
SALMON	CATFISH	SHARK	TROUT	TUNA	COD
FRANCE	GERMANY	ITALY	MEXICO	SPAIN	BRAZIL
RAP	JAZZ	PUNK	CLASSICAL	ROCK	FOLK
JUICE	WATER	LEMONADE	COFFEE	TEA	SODA

Unrelated (UUUUUU) Lists Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	HOUR	UNCLE	MAGAZINE	SERGEANT	DIAMOND
IRON	SECOND	AUNT	BOOK	GENERAL	RUBY
COPPER	MINUTE	COUSIN	NEWSPAPER	CAPTAIN	EMERALD
LEAD	YEAR	MOTHER	JOURNAL	COLONEL	SAPPHIRE
BRASS	DAY	FATHER	NOVEL	ADMIRAL	PEARL
BRONZE	MONTH	SISTER	TEXTBOOK	PRIVATE	JADE
Set 02					
COTTON	BLUE	LEG	APPLE	GUN	HOUSE
SILK	RED	ARM	BANANA	PISTOL	APARTMENT
WOOL	GREEN	FINGER	GRAPE	SWORD	MANSION
SATIN	YELLOW	ELBOW	PEAR	RIFLE	TENT
LINEN	PURPLE	ANKLE	PEACH	GRENADE	CABIN
FLANNEL	BROWN	KNEE	PLUM	MISSILE	HUT
Set 03					
TORNADO	SHIRT	WINDOW	DRUMS	EAGLE	CAR
HURRICANE	SOCKS	DOOR	GUITAR	ROBIN	BUS
LIGHTNING	HAT	FLOOR	FLUTE	HAWK	TRUCK
EARTHQUAKE	JACKET	WALL	PIANO	CROW	TRAIN
TYPHOON	SWEATER	ROOF	TRUMPET	SEAGULL	VAN
BLIZZARD	JEANS	STAIRS	VIOLIN	PIGEON	TAXI
Set 04					
CARROT	FLY	SALMON	FRANCE	RAP	JUICE
PEAS	ANT	TROUT	MEXICO	CLASSICAL	COFFEE
CORN	BEE	CATFISH	GERMANY	JAZZ	WATER
SPINACH	MOSQUITO	TUNA	SPAIN	ROCK	TEA
BEANS	BUTTERFLY	SHARK	ITALY	PUNK	LEMONADE
CABBAGE	WASP	COD	BRAZIL	FOLK	SODA

Mixed (RRRUUU) Lists Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	COPPER	BRASS	SAPPHIRE	BOOK	PRIVATE
HOUR	MINUTE	DAY	JOURNAL	GENERAL	JADE
UNCLE	COUSIN	FATHER	COLONEL	RUBY	TEXTBOOK
MAGAZINE	NEWSPAPER	NOVEL	MOTHER	IRON	MONTH
SERGEANT	CAPTAIN	ADMIRAL	LEAD	SECOND	SISTER
DIAMOND	EMERALD	PEARL	YEAR	AUNT	BRONZE
Set 02					
COTTON	WOOL	LINEN	TENT	BANANA	MISSILE
BLUE	GREEN	PURPLE	PEAR	PISTOL	HUT
LEG	FINGER	ANKLE	RIFLE	APARTMENT	PLUM
APPLE	GRAPE	PEACH	ELBOW	SILK	BROWN
GUN	SWORD	GRENADE	SATIN	RED	KNEE
HOUSE	MANSION	CABIN	YELLOW	ARM	FLANNEL
Set 03					
TORNADO	LIGHTNING	TYPHOON	TRAIN	GUITAR	PIGEON
SHIRT	HAT	SWEATER	PIANO	ROBIN	TAXI
WINDOW	FLOOR	ROOF	CROW	BUS	VIOLIN
DRUMS	FLUTE	TRUMPET	WALL	HURRICANE	JEANS
EAGLE	HAWK	SEAGULL	EARTHQUAKE	SOCKS	STAIRS
CAR	TRUCK	VAN	JACKET	DOOR	BLIZZARD
Set 04					
CARROT	CORN	BEANS	TEA	MEXICO	FOLK
FLY	BEE	BUTTERFLY	SPAIN	CLASSICAL	SODA
SALMON	CATFISH	SHARK	ROCK	COFFEE	BRAZIL
FRANCE	GERMANY	ITALY	TUNA	PEAS	WASP
RAP	JAZZ	PUNK	SPINACH	ANT	COD
JUICE	WATER	LEMONADE	MOSQUITO	TROUT	CABBAGE

Mixed (UUURRR) Lists Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
MAGAZINE	EMERALD	ADMIRAL	IRON	LEAD	BRONZE
SERGEANT	NEWSPAPER	PEARL	SECOND	YEAR	MONTH
DIAMOND	CAPTAIN	NOVEL	AUNT	MOTHER	SISTER
STEEL	COUSIN	DAY	BOOK	JOURNAL	TEXTBOOK
HOUR	COPPER	FATHER	GENERAL	COLONEL	PRIVATE
UNCLE	MINUTE	BRASS	RUBY	SAPPHIRE	JADE
Set 02					
APPLE	MANSION	GRENADE	SILK	SATIN	FLANNEL
GUN	GRAPE	CABIN	RED	YELLOW	BROWN
HOUSE	SWORD	PEACH	ARM	ELBOW	KNEE
COTTON	FINGER	PURPLE	BANANA	PEAR	PLUM
BLUE	WOOL	ANKLE	PISTOL	RIFLE	MISSILE
LEG	GREEN	LINEN	APARTMENT	TENT	HUT
Set 03					
DRUMS	TRUCK	SEAGULL	HURRICANE	EARTHQUAKE	BLIZZARD
EAGLE	FLUTE	VAN	SOCKS	JACKET	JEANS
CAR	HAWK	TRUMPET	DOOR	WALL	STAIRS
TORNADO	FLOOR	SWEATER	GUITAR	PIANO	VIOLIN
SHIRT	LIGHTNING	ROOF	ROBIN	CROW	PIGEON
WINDOW	HAT	TYPHOON	BUS	TRAIN	TAXI
Set 04					
FRANCE	WATER	PUNK	PEAS	SPINACH	CABBAGE
RAP	GERMANY	LEMONADE	ANT	MOSQUITO	WASP
JUICE	JAZZ	ITALY	TROUT	TUNA	COD
CARROT	CATFISH	BUTTERFLY	MEXICO	SPAIN	BRAZIL
FLY	CORN	SHARK	CLASSICAL	ROCK	FOLK
SALMON	BEE	BEANS	COFFEE	TEA	SODA

Appendix C

Stimuli lists used in Experiment 3

Related (RRRRRR) Lists Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	COPPER	BRASS	IRON	LEAD	BRONZE
HOUR	MINUTE	DAY	SECOND	YEAR	MONTH
UNCLE	COUSIN	FATHER	AUNT	MOTHER	SISTER
MAGAZINE	NEWSPAPER	NOVEL	BOOK	JOURNAL	TEXTBOOK
SERGEANT	CAPTAIN	ADMIRAL	GENERAL	COLONEL	PRIVATE
DIAMOND	EMERALD	PEARL	RUBY	SAPPHIRE	JADE
COTTON	WOOL	LINEN	SILK	SATIN	FLANNEL
BLUE	GREEN	PURPLE	RED	YELLOW	BROWN
LEG	FINGER	ANKLE	ARM	ELBOW	KNEE
APPLE	GRAPE	PEACH	BANANA	PEAR	PLUM
GUN	SWORD	GRENADE	PISTOL	RIFLE	MISSILE
HOUSE	MANSION	CABIN	APARTMENT	TENT	HUT
Set 02					
TORNADO	LIGHTNING	TYPHOON	HURRICANE	EARTHQUAKE	BLIZZARD
SHIRT	HAT	SWEATER	SOCKS	JACKET	JEANS
WINDOW	FLOOR	ROOF	DOOR	WALL	STAIRS
DRUMS	FLUTE	TRUMPET	GUITAR	PIANO	VIOLIN
EAGLE	HAWK	SEAGULL	ROBIN	CROW	PIGEON
CAR	TRUCK	VAN	BUS	TRAIN	TAXI
CARROT	CORN	BEANS	PEAS	SPINACH	CABBAGE
FLY	BEE	BUTTERFLY	ANT	MOSQUITO	WASP
SALMON	CATFISH	SHARK	TROUT	TUNA	COD
FRANCE	GERMANY	ITALY	MEXICO	SPAIN	BRAZIL
RAP	JAZZ	PUNK	CLASSICAL	ROCK	FOLK
JUICE	WATER	LEMONADE	COFFEE	TEA	SODA

Unrelated (UUUUUU) Lists Position 01	Position 02	Position 03	Position 04	Position 05	Position 06
Set 01					
STEEL	HOUR	UNCLE	MAGAZINE	SERGEANT	DIAMOND
IRON	SECOND	AUNT	BOOK	GENERAL	RUBY
COPPER	MINUTE	COUSIN	NEWSPAPER	CAPTAIN	EMERALD
LEAD	YEAR	MOTHER	JOURNAL	COLONEL	SAPPHIRE
BRASS	DAY	FATHER	NOVEL	ADMIRAL	PEARL
BRONZE	MONTH	SISTER	TEXTBOOK	PRIVATE	JADE
COTTON	BLUE	LEG	APPLE	GUN	HOUSE
SILK	RED	ARM	BANANA	PISTOL	APARTMENT
WOOL	GREEN	FINGER	GRAPE	SWORD	MANSION
SATIN	YELLOW	ELBOW	PEAR	RIFLE	TENT
LINEN	PURPLE	ANKLE	PEACH	GRENADE	CABIN
FLANNEL	BROWN	KNEE	PLUM	MISSILE	HUT
Set 02					
TORNADO	SHIRT	WINDOW	DRUMS	EAGLE	CAR
HURRICANE	SOCKS	DOOR	GUITAR	ROBIN	BUS
LIGHTNING	HAT	FLOOR	FLUTE	HAWK	TRUCK
EARTHQUAKE	JACKET	WALL	PIANO	CROW	TRAIN
TYPHOON	SWEATER	ROOF	TRUMPET	SEAGULL	VAN
BLIZZARD	JEANS	STAIRS	VIOLIN	PIGEON	TAXI
CARROT	FLY	SALMON	FRANCE	RAP	JUICE
PEAS	ANT	TROUT	MEXICO	CLASSICAL	COFFEE
CORN	BEE	CATFISH	GERMANY	JAZZ	WATER
SPINACH	MOSQUITO	TUNA	SPAIN	ROCK	TEA
BEANS	BUTTERFLY	SHARK	ITALY	PUNK	LEMONADE
CABBAGE	WASP	COD	BRAZIL	FOLK	SODA

Appendix D

Stimuli used in Experiment 4

Related (RRRRRRR)						
Set 01						
OVEN	MICROWAVE	FRIDGE	MIXER	TOASTER	BLENDER	STOVE
RAP	CLASSICAL	JAZZ	ROCK	PUNK	FOLK	BLUES
TRIANGLE	SQUARE	CIRCLE	CUBE	HEXAGON	RECTANGLE	CYLINDER
OAK	PINE	MAPLE	REDWOOD	SPRUCE	FIR	ELM
MILE	METER	INCH	KILOMETER	CENTIMETER	YARD	MILLIMETER
CARROT	PEAS	CORN	SPINACH	BEANS	CABBAGE	TURNIP
COW	PIG	HORSE	GOAT	SHEEP	DONKEY	BULL
ROSE	DAISY	TULIP	LILY	DAFFODIL	ORCHID	DANDELION
APPLE	ORANGE	BANANA	GRAPE	PEAR	PEACH	PLUM
CHEMISTRY	PHYSICS	BIOLOGY	ZOOLOGY	BOTANY	GEOLOGY	ASTRONOMY
BULLDOG	TERRIER	POODLE	BEAGLE	SPANIEL	DOBERMAN	GREYHOUND
STEEL	IRON	COPPER	TIN	LEAD	BRASS	BRONZE
LOVE	SADNESS	HAPPINESS	FEAR	SORROW	ANGER	JOY
EYE	EAR	NOSE	LIPS	TONGUE	CHIN	CHEEK
SALT	PEPPER	GARLIC	OREGANO	BASIL	MUSTARD	PAPRIKA
SALMON	TROUT	CATFISH	TUNA	SHARK	COD	HALIBUT
SHOE	BOOTS	SANDALS	SLIPPERS	LOAFERS	SNEAKERS	CLOGS
CHAIR	TABLE	COUCH	BED	SOFA	RECLINER	OTTOMAN
MEASLES	MUMPS	FLU	POLIO	MALARIA	HEPATITIS	CANCER
YACHT	CANOE	SUBMARINE	FERRY	SPEEDBOAT	BATTLESHIP	BARGE
NOUN	VERB	ADJECTIVE	ADVERB	PRONOUN	PREPOSITION	PARTICIPLE
CAR	BUS	TRUCK	TRAIN	VAN	TAXI	MOTORCYCLE
RIVER	OCEAN	LAKE	CREEK	STREAM	WATERFALL	LAGOON
COTTON	SILK	WOOL	SATIN	DENIM	LINEN	FLANNEL
KING	QUEEN	PRINCE	PRINCESS	DUKE	COUNT	EARL
LEG	ARM	FINGER	TOE	ELBOW	ANKLE	KNEE

Related (RRRRRRR)						
Set 02						
KNIFE	FORK	SPOON	SPATULA	WHISK	LADLE	TONGS
HOUR	SECOND	MINUTE	YEAR	DAY	MONTH	WEEK
MILK	JUICE	COFFEE	WATER	TEA	LEMONADE	SODA
MOZART	BEETHOVEN	BACH	BRAHMS	VERDI	WAGNER	HANDEL
FOOTBALL	BASKETBALL	BASEBALL	TENNIS	HOCKEY	GOLF	RUGBY
PRIEST	POPE	BISHOP	NUN	CARDINAL	RABBI	PASTOR
TORNADO	HURRICANE	LIGHTNING	EARTHQUAKE	TYPHOON	BLIZZARD	MONSOON
DOCTOR	TEACHER	LAWYER	NURSE	DENTIST	SECRETARY	BANKER
BLUE	RED	GREEN	YELLOW	PURPLE	BROWN	PINK
HAMMER	SCREWDRIVER	CHISEL	PLIERS	NAILS	DRILL	SAW
SERGEANT	GENERAL	CAPTAIN	COLONEL	MAJOR	ADMIRAL	PRIVATE
OXYGEN	HYDROGEN	HELIUM	NITROGEN	CHLORINE	ARGON	RADON
HEART	LIVER	KIDNEY	BRAIN	LUNGS	STOMACH	INTESTINE
DRUMS	GUITAR	FLUTE	PIANO	TRUMPET	VIOLIN	BANJO
DIAMOND	RUBY	EMERALD	SAPPHIRE	PEARL	OPAL	JADE
BEER	VODKA	WINE	RUM	GIN	SCOTCH	MARTINI
SHIRT	SOCKS	HAT	JACKET	SWEATER	JEANS	COAT
LION	TIGER	ELEPHANT	GIRAFFE	ZEBRA	CHEETAH	WILDEBEEST
UNCLE	AUNT	COUSIN	MOTHER	FATHER	SISTER	BROTHER
BREAKFAST	LUNCH	DINNER	SUPPER	BARBECUE	BRUNCH	BANQUET
EAGLE	ROBIN	HAWK	CROW	SEAGULL	PIGEON	SPARROW
MAGAZINE	BOOK	NEWSPAPER	JOURNAL	NOVEL	TEXTBOOK	LETTER
GUN	PISTOL	SWORD	RIFLE	BAYONET	GRENADE	MISSILE
WINDOW	DOOR	FLOOR	WALL	ROOF	STAIRS	CELLAR
TRUNK	BRANCH	LIMB	BARK	LEAF	ROOT	SAP
CHURCH	TEMPLE	SYNAGOGUE	MOSQUE	CATHEDRAL	CHAPEL	MONASTERY

Appendix E

Stimuli lists used in Experiment 5

Position 01	Position 02	Position 03	Position 04	Position 05	Position 06	Position 07	Position 08	Position 09	Position 10	Critical Lure	Number of Related words
BED	**NAP**	**SNORE**	**AWAKE**	**SNOOZE**	**TIRED**	**WAKE**	**REST**	**BLANKET**	**DREAM**	**SLEEP**	10
BED	**NAP**	**SNORE**	**AWAKE**	**SNOOZE**	**TIRED**	**WAKE**	**REST**	**BLANKET**	GLASS	**SLEEP**	9
BED	**NAP**	**SNORE**	**AWAKE**	**SNOOZE**	**TIRED**	**WAKE**	**REST**	WASTE	GLASS	**SLEEP**	8
BED	**NAP**	**SNORE**	**AWAKE**	**SNOOZE**	**TIRED**	**WAKE**	CANDY	WASTE	GLASS	**SLEEP**	7
BED	**NAP**	**SNORE**	**AWAKE**	**SNOOZE**	**TIRED**	SHARP	CANDY	WASTE	GLASS	**SLEEP**	6
BED	**NAP**	**SNORE**	**AWAKE**	**SNOOZE**	STOP	SHARP	CANDY	WASTE	GLASS	**SLEEP**	5
BED	**NAP**	**SNORE**	**AWAKE**	BUMPY	STOP	SHARP	CANDY	WASTE	GLASS	**SLEEP**	4
BED	**NAP**	**SNORE**	PIN	BUMPY	STOP	SHARP	CANDY	WASTE	GLASS	**SLEEP**	3
BED	**NAP**	VALLEY	PIN	BUMPY	STOP	SHARP	CANDY	WASTE	GLASS	**SLEEP**	2
BED	HILL	VALLEY	PIN	BUMPY	STOP	SHARP	CANDY	WASTE	GLASS	**SLEEP**	1
FEAR	HILL	VALLEY	PIN	BUMPY	STOP	SHARP	CANDY	WASTE	GLASS	**SLEEP**	0

Appendix F. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2024.104573.

References

Aujla, H. (2021). Language experience predicts semantic priming of lexical decision. *Canadian Journal of Experimental Psychology, 75*(3), 235–244. https://doi.org/10.1037/cep0000255

Berry, M. W., & Browne, M. (1999). Understanding search engines - mathematical modeling and text retrieval. *Software*.

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition, 37*(5), 689–713. https://doi.org/10.3758/MC.37.5.689

Bireta, T. J., Guitard, D., Neath, I., & Surprenant, A. M. (2021). Valence does not affect serial recall. *Canadian Journal of Experimental Psychology, 75*, 35–47. https://doi.org/10.1037/cep0000239.

Bireta, T. J., Guitard, D., Neath, I., & Surprenant, A. M. (2023). Valence and concreteness in item recognition: Evidence against the affective embodiment account. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02442-8.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. arXiv. https://arxiv.org/abs/1607.04606.

Brooks, J. O., III, & Watkins, M. J. (1990). Further evidence of the intricacy of memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 1134–1141. https://doi.org/10.1037/0278-7393.16.6.1134

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods, 44*, 991–997. https://doi.org/10.3758/s13428-012-0190-4

Buchanan, E. M., Valentine, K. D., & Maxwell, N. P. (2019). English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods, 51*(4), 1849–1863. https://doi.org/10.3758/s13428-019-01243-z

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 551–581. https://doi.org/10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*(4), 627–652. https://doi.org/10.1016/j.jml.2006.08.005

Caplan, J. B. (2023). Sparse attentional subsetting of item features and list-composition effects on recognition memory. *Journal of Mathematical Psychology, 116*, Article 102802. https://doi.org/10.1016/j.jmp.2023.102802

Caplan, J. B., & Guitard, D. (2024a). A feature-space theory of the production effect in recognition. *Experimental Psychology, 71*(1), 64–82. https://doi.org/10.1027/1618-3169/a000611

Caplan, J. B., & Guitard, D. (2024b). Stimulus duration and recognition memory: An attentional subsetting account. *Journal of Memory and Language.*

Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language, 119*, Article 104210. https://doi.org/10.1016/j.jml.2020.104210

Chang, M., & Johns, B. (2023). Integrating distributed semantic models with an instance memory model to explain false recognition. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society* (pp. 2042-2049). Retrieved from https://escholarship.org/uc/item/2s14p686.

Chubala, C. M., Johns, B. T., Jamieson, R. K., & Mewhort, D. J. K. (2016). Applying an exemplar model to an implicit rule-learning task: Implicit learning of semantic structure. *Quaterly Journal of Experimental Psychology, 69*(6), 1049–1055. https://doi.org/10.1080/17470218.2015.1130068

Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory & Cognition, 26*(2), 330–342. https://doi.org/10.3758/BF03201144

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology, 55*(4), 429–432. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin, 145*(8), 822–847. https://doi.org/10.1037/bul0000199

Cowan, N., Bao, C., Bishop-Chrzanowski, B. M., Costa, A. N., Greene, N. R., Guitard, D., Li, C., Musich, M. L., & Ünal, Z. E. (2024). The relation between attention and memory. *Annual Review of Psychology, 75*, 183–214. https://doi.org/10.1146/annurev-psych-040723-012736

Crowder, R. G. (1979). Similarity and order in memory. In G. Bower (Ed.), *Psychology of learning and motivation* (vol. 13, pp. 319–353). New York: Academic Press.

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review, 111*(3), 800–807. https://doi.org/10.1037/0033-295X.111.3.800

Criss, A. H., & Shiffrin, R. M. (2005). List Discrimination in Associative Recognition and Implications for Representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1199–1212. https://doi.org/10.1037/0278-7393.31.6.1199

Dauphinee, I., Roy, M., Guitard, D., Yearsley, J. M., Poirier, M., & Saint-Aubin, J. (2024). Give me enough time to rehearse: Presentation rate modulates the production effect. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02437-5

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*, 987–1006. https://doi.org/10.3758/s13428-018-1115-7

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22. https://doi.org/10.1037/h0046671

Farrell, S., & Lewandowsky, S. (2002). An endongeous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*(1), 59–79. https://doi.org/10.3758/BF03196257

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Firth, J. R. (1957). Applications of general linguistics. *Transactions of the Philological Society, 56*, 1–14. https://doi.org/10.1111/j.1467-968X.1957.tb00568.x

Franklin, D. R. J., & Mewhort, D. J. K. (2002). An Analysis of Immediate Memory: The Free-Recall Task. In N. J. Dimopoulos, & K. F. Li (Eds.), *High Performance Computing*

Systems and Applications (vol 657). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-0849-6_30.

Franklin, D. R. J., & Mewhort, D. J. K. (2015). Memory as a hologram: An analysis of learning and recall. *Canadian Journal of Experimental Psychology, 69*(1), 115–135. https://doi.org/10.1037/cep0000035

Guérard, K., & Saint-Aubin, J. (2012). Assessing the effect of lexical variables in backward recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 312–324. https://doi.org/10.1037/a0025481

Guitard, D., & Cowan, N. (2020). Do we use visual codes when information is not presented visually? *Memory & Cognition, 48*, 1522–1536. https://doi.org/10.3758/s13421-020-01054-0

Guitard, D., Saint-Aubin, J., & Cowan, N. (2021). Asymmetrical interference between item and order information in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(2), 243–263. https://doi.org/10.1037/xlm0000956

Guitard, D., Saint-Aubin, J., & Cowan, N. (2022). Tradeoffs between item and order information in short-term memory. *Journal of Memory and Language, 122*. https://doi.org/10.1016/j.jml.2021.104300

Guitard, D., Saint-Aubin, J., & Neath I. (2023). Additional evidence that valence does not affect serial recall. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218221126635.

Guitard, D., & Cowan, N. (2023). The tradeoff between item and order information in short -term memory does not depend on encoding time. *Journal of Experimental Psychology: Human Perception and Performance, 49*(1), 51–70. https://doi.org/10.1037/xhp0001074

Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word length, set size, and lexical factors: Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(11), 1824–1844. https://doi.org/10.1037/xlm0000551

Guitard, D., Miller, L. M., Neath, I., & Roodenrys, S. (2019). Does contextual diversity affect serial recall? *Journal of Cognitive Psychology, 31*(4), 379–396. https://doi.org/10.1080/20445911.2019.1626401

Guitard, D., Saint-Aubin, J., Reid, J. N., & Jamieson, R. K. (2024). An Embedded Computational Framework of Memory: The Critical Role of Representations in Veridical and False Recall Predictions.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behav Res, 47*, 930–944. https://doi.org/10.3758/s13428-014-0529-0

Harris, Z. S. (1954). Distributional structure. *WORD, 10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hollis, G., & Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin & Review, 23*(6), 1744–1756. https://doi.org/10.3758/s13423-016-1053-2

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Howard, M. W., Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46* (3), 269-299. Doi: 10.1006/jmps.2001.1388.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1217–1232. https://doi.org/10.1037/0278-7393.23.5.1217

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language, 49*(4), 500–518. https://doi.org/10.1016/S0749-596X(03)00096-2

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*(2), 339–373. https://doi.org/10.1037/a0034221

Ishiguro, S., & Saito, S. (2021). The detrimental effect of semantic similarity in short-term memory tasks: A meta-regression approach. *Psychonomic Bulletin & Review, 28*(2), 384–408. https://doi.org/10.3758/s13423-020-01815-7

Ishiguro, S., & Saito, S. (2024). The semantic similarity effect on short-term memory: Null effects of affectively defined semantic similarity. *Journal of Cognition, 7(1): 24*, 1–13. https://doi.org/10.5334/joc.349

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2018). An instance theory of semantics. *Computational Brain & Behavior, 1*, 119–136. https://doi.org/10.1007/s42113-018-0008-2

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology, 1*, 173–184. https://doi.org/10.1038/s44159-022-00025-3

Jones, M. N. (2019). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience, 34*(10), 1338–1346. https://doi.org/10.1080/23273798.2018.1431679

Johns, B. T., & Jamieson, R. K. (2019). The influence of time and place on lexical behavior: A distributional analysis. *Behavior Research Methods, 51*, 2483. https://doi.org/10.3758/s13428-019-01289-z

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review, 17*, 662–672. https://doi.org/10.3758/PBR.17.5.662

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology, 65*(4), 486–518. https://doi.org/10.1016/j.cogpsych.2012.07.002

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*, 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE, 5*, e8622.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review, 114*, 954–993. https://doi.org/10.1037/0033-295X.114.4.954

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7*(2), 257–266. https://doi.org/10.3758/BF03212981

Kowialiewski, B., & Majerus, S. (2020). The varying nature of semantic effects in working memory. *Cognition, 202*. https://doi.org/10.1016/j.cognition.2020.104278

Kowialiewski, B., Lemaire, B., & Portrat, S. (2021). How does semantic knowledge impact working memory maintenance? Computational and behavioral investigations. *Journal of Memory and Language, 117*. https://doi.org/10.1016/j.jml.2020.104208

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2022). The semantic relatedness effect in serial recall: Deconfounding encoding and recall order. *Journal of Memory and Language, 127*, 104377. https://doi.org/10.1016/j.jml.2022.104377

Kowialiewski, B., Lemaire, B., & Portrat, S. (2022). Between-item similarity frees up working memory resources through compression: A domain-general property. *Journal of Experimental Psychology: General, 151*(11), 2641–2665. https://doi.org/10.1037/xge0001235

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2023). Verbal working memory encodes phonological and semantic information differently. *Cognition, 233*. https://doi.org/10.1016/j.cognition.2022.105364

Kowialiewski, B., Majerus, S., & Oberauer, K. (2024). Does semantic similarity affect immediate memory for order? Usually not, but sometimes it does. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*(1), 68–88. https://doi.org/10.1037/xlm0001279

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Landry, É. R., Guitard, D., & Saint-Aubin, J. (2022). Arousal affects short-term serial recall. *Canadian Journal of Experimental Psychology, 76*(2), 99–110. https://doi.org/10.1037/cep0000272

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0, .

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics, 4*, 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: A dynamic network model. *International Journal of Psychology, 34*(5–6), 434–446. https://doi.org/10.1080/002075999399792

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review, 96*(1), 25–57. https://doi.org/10.1037/0033-295X.96.1.25

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition, 44*, 590–607. https://doi.org/10.3758/s13421-015-0580-9

Logan, G. D., & Cox, G. E. (2023). Serial order depends on item-dependent and item-independent contexts. *Psychological Review, 130*(6), 1672–1687. https://doi.org/10.1037/rev0000422

Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 35–55). Lawrence Erlbaum Associates Publishers.

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.04.016

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547–559. https://doi.org/10.3758/BF03192726

Mewhort, D. J. K., Shabahang, K. D., & Franklin, D. R. J. (2018). Release from PI: An analysis and a model. *Psychonomic Bulletin & Review, 25*, 932–950. https://doi.org/10.3758/s13423-017-1327-3

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality (pp. 3111-3119). In C. J. C. Burges, L, Bottou, M., Welling, Z., Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, 26.

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science, 34*, 1388–1429. https://doi.org/10.1111/j.1551-6709.2010.01106.x

Morin, C., Poirier, M., Fortin, C., & Hulme, C. (2006). Word frequency and the mixed-list paradox in immediate and delayed serial recall. *Psychonomic Bulletin & Review, 13*(4), 724–729. https://doi.org/10.3758/BF03193987

Monaco, J. D., Abbott, L. F., & Kahana, M. J. (2007). Lexico-semantic structure and the word-frequency effect in recognition memory. *Learning & Memory, 14*, 204–213.

Morey, R. D. & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor.

Morton, N. W., & Polyn, S. M. (2016). A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language, 86*, 119–140. https://doi.org/10.1016/j.jml.2015.10.002

Murdock, B. B. (1974). Human memory: Theory and data. Lawrence Erlbaum.

Murdock, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General, 105*(2), 191–216. https://doi.org/10.1037/0096-3445.105.2.191

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*(6), 609–626. https://doi.org/10.1037/0033-295X.89.6.609

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100*(2), 183–203. https://doi.org/10.1037/0033-295X.100.2.183

Murdock, B. B., Jr., & Vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology, 74*, 137–143. https://doi.org/10.1037/h0024507

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16*, 343–352. https://doi.org/10.3758/BF03197045

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251–269. https://doi.org/10.3758/BF03213879

Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition, 35*(8), 1940–1953. https://doi.org/10.3758/BF03192927

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition, 25*, 256–263. https://doi.org/10.3758/BF03201116

Neath, I., Saint-Aubin, J., & Surprenant, A. M. (2022). Semantic relatedness effects in serial recall but not in serial reconstruction of order. *Experimental Psychology, 69*(4), 196–209. https://doi.org/10.1027/1618-3169/a000557

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018a). Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. *Journal of Experimental Psychology: General, 147*(3), 328–353. https://doi.org/10.1037/xge0000369

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018b). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods, 50*(2), 530–556. https://doi.org/10.3758/s13428-017-0884-8

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*(9), 885–958. https://doi.org/10.1037/bul0000153

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review, 18*(1), 10–45. https://doi.org/10.3758/s13423-010-0020-6

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Osth, A. F., & Hurlstone, M. J. (2023). Do item-dependent context representations underlie serial order in cognition? Commentary on Logan (2021). *Psychological Review, 130*(2), 513–545. https://doi.org/10.1037/rev0000352

Osth, A. F., Shabahang, K. D., Mewhort, D. J. K., & Heathcote, A. (2020). Global semantic similarity effects in recognition memory: Insights from BEAGLE representations and the diffusion decision model. *Journal of Memory and Language, 111*, Article 104071. https://doi.org/10.1016/j.jml.2019.104071

Osth, A. F., & Zhang, L. (2023). Integrating word-form representations with global similarity computation in recognition memory. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02402-2

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

Petilli, M. A., Marelli, M., Mazzoni, G., Marchetti, M., Rinaldi, L., & Gatti, D. (2024). From vector spaces to DRM lists: False Memory Generator, a software for automated generation of lists of stimuli inducing false memories. *Behavior Research Methods, 56*(4), 3779–3793. https://doi.org/10.3758/s13428-024-02425-0

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology, 48A*, 384–404. https://doi.org/10.1080/14640749508401396

Poirier, M., Saint-Aubin, J., Musselwhite, K., Mohanadas, T., & Mahammed, G. (2007). Visual similarity effects on short-term memory for order: The case of verbally labeled pictorial stimuli. *Memory & Cognition, 35*(4), 711–723. https://doi.org/10.3758/BF03193309

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43*(3), 489–499. https://doi.org/10.3758/s13421-014-0470-6

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*, 129–156. https://doi.org/10.1037/a0014420

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), Psychology of Learning and Motivation (Vol. 14, pp. 207-262). Academic Press. Doi: 10.1016/S0079-7421(08)60162-0.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93

Reid, J. N., & Jamieson, R. K. (2023). True and false recognition in MINERVA 2: Extension to sentences and metaphors. *Journal of Memory and Language, 129*, Article 104397.

Reid, J. N., & Katz, A. N. (2018). Vector space applications in metaphor comprehension. *Metaphor and Symbol, 33*(4), 280–294. https://doi.org/10.1080/10926488.2018.1549840

Reid, J. N., Yang, H., & Jamieson, R. K. (2023). A computational account of item-based directed forgetting for nonwords: Incorporating orthographic representations in MINERVA 2. *Memory & Cognition, 51*, 1785–1806. https://doi.org/10.3758/s13421-023-01433-3

Reid, N., Guitard, D., & Jamieson, R. (2024). MINERVA OPS: A computational framework for the representation and recognition of orthographic, phonological, and semantic associates.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*(3), 231–237. https://doi.org/10.1111/j.1467-9280.1997.tb00417.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roodenrys, S., Guitard, D., Miller, L. M., Barron, J., & Saint-Aubin, J. (2022). Phonological similarity in the serial recall task hinders item recall, not just order. *British Journal of Psychology., 113*(4), 1100–1120. https://doi.org/10.1111/bjop.12575

Roodenrys, S., Hulme, C., Alban, J., et al. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22*, 695–701. https://doi.org/10.3758/BF03209254

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Rubenstein, H., & Goodenough, J. B. (1965). Contextual Correlates of Synonymy. *Communications of the ACM, 627–633*. https://doi.org/10.1145/365628.365657

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*(1), 63–77. https://doi.org/10.1037/h0031185

Saint-Aubin, J., Guérard, K., Chamberland, C., & Malenfant, A. (2014). Delineating the contribution of long-term associations to immediate recall. *Memory, 22*, 360–373. https://doi.org/10.1080/09658211.2013.794242

Saint-Aubin, J., & Poirier, M. (1999a). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52*, 367–394. https://doi.org/10.1080/713755814

Saint-Aubin, J., & Poirier, M. (1999b). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34*, 347–352. https://doi.org/10.1080/002075999399675

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials. *Psychonomic Bulletin & Review, 12*, 171–177. https://doi.org/10.3758/BF03196364

Saint-Aubin, J., Yearsley, J., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language., 118*, 1–26. https://doi.org/10.1016/j.jml.2021.104219

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*(2), 168–175. https://doi.org/10.3758/BF03202729

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review, 12*(5), 787–805. https://doi.org/10.3758/BF03196773

Sonier, R.-P., Guitard, D., Melanson, E., Jamieson, R., & Saint-Aubin, J. (2024). Can semantic similarity be better represented by valence, arousal, and dominance?

Spens, E., & Burgess, N. (2024). A generative model of memory construction and consolidation. *Nature Human Behaviour, 8*, 526–543. https://doi.org/10.1038/s41562-023-01799-z

Steyvers, M. (2000). *Modeling semantic and orthographic similarity effects on memory for individual words (Ph.D.)*. Indiana University.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2005). Word Association Spaces for Predicting Semantic Similarity Effects in Episodic Memory. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 237–249). American Psychological Association. https://doi.org/10.1037/10895-018.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolKit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Tehan, G. (2010). Associative relatedness enhances recall and produces false memories in immediate serial recall. *Canadian Journal of Experimental Psychology, 64*(4), 266–272. https://doi.org/10.1037/a0021375

Tse, C.-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory, 17*, 874–891. https://doi.org/10.1080/09658210903376250

Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *Quarterly Journal of Experimental Psychology, 64*(12), 2425–2437. https://doi.org/10.1080/17470218.2011.604787