Article:
Mak, Matthew H C, Ball, Lewis V, O'Hagan, Alice et al. (2025) Involvement of episodic memory in language comprehension: Naturalistic comprehension pushes unrelated words closer in semantic space for at least 12 h. Cognition. 106086. ISSN: 0010-0277

https://doi.org/10.1016/j.cognition.2025.106086

Cognition 258 (2025) 106086

Contents lists available at ScienceDirect

Cognition

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

Involvement of episodic memory in language comprehension: Naturalistic comprehension pushes unrelated words closer in semantic space for at least 12 h

Matthew H.C. Mak [a,*], Lewis V. Ball [b], Alice O'Hagan [c], Catherine R. Walsh [d], M. Gareth Gaskell [b]

[a] Department of Psychology, University of Warwick, Coventry, UK
[b] Department of Psychology, University of York, York, UK
[c] Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[d] Department of Psychology, University of California, Los Angeles, CA, USA

ARTICLE INFO

Keywords:
Episodic memory
Language comprehension
Relatedness
Lexical processing
Semantic space

ABSTRACT

Recent experience with a word significantly influences its subsequent interpretation. For instance, encountering *bank* in a river-related context biases future interpretations toward 'side of a river' (vs. 'financial bank'). To explain this effect, the *episodic context* account posits that episodic memory helps bind word meanings in the language input, creating a temporary, context-specific representation that can bias subsequent lexical interpretation. This account predicts that even *unrelated* words would be linked together in episodic memory, potentially altering their interpretation. In Experiments 1–3, participants read unrelated word pairs (e.g., *sword—microwave*, *privacy—export*) embedded in meaningful sentences, then completed a speeded relatedness judgement task after delays of 5 min, 20 min, or 12 h (including sleep). Results showed that sentence exposure increased the likelihood of the unrelated pairs being judged as related—a robust effect observed across all delay intervals. Experiment 4 showed that this exposure effect was abolished when words in a target pair were read in separate sentences, suggesting that the exposure effect may be dependent on lexical co-occurrence. Experiment 5, also with a 12-h delay (including sleep), additionally used an innovative word arrangement task to assess word relatedness without presenting the target pairs simultaneously or successively. In line with relatedness judgement, sentence exposure pushed the unrelated words closer in semantic space. Overall, our findings suggest that a context-specific representation, supported by episodic memory, is generated during language comprehension, and in turn, these representations can influence lexical interpretation for at least 12 h and across different linguistic circumstances. We argue that these representations endow the mental lexicon with the efficiency to deal with word burstiness and the dynamic nature of language.

1. Introduction

Language comprehension necessitates retrieving the meanings of individual words, while constructing a coherent representation that integrates their meanings. In this article, we present empirical evidence from four behavioral experiments, arguing that episodic memory contributes to this coherent representation, which may, in turn, bias future lexical processing.

We set the stage by first considering the processing of homonyms. The homonymic word, *bank*, has at least two distinct meanings, one of which has a high frequency (financial institute) while another is less common (side of a river), and typically, language users prefer the high-frequency meaning. For example, eye-tracking data have shown that when *bank* is read in a neutral context where both meanings are plausible (e.g., "The man knew that the bank…"), readers are biased to retrieve the high-frequency meaning (e.g., Rayner & Duffy, 1986). Similarly, in associate production, where participants are prompted to provide the first word that comes to mind upon hearing or seeing *bank* presented in isolation, participants tend to give associates related to a financial institute (e.g., *money*) rather than those pertaining to rivers (Gilbert & Rodd, 2022; Twilley et al., 1994). While individuals typically favour the high-frequency meaning, this preference is not set in stone and can be influenced by recent linguistic experience. In studies by Rodd et al., 2013, Rodd et al., 2016), participants were first exposed to

* Corresponding author at: Department of Psychology, University of Warwick, Coventry CV4 7ES, UK.
E-mail address: matthew.mak@warwick.ac.uk (M.H.C. Mak).

https://doi.org/10.1016/j.cognition.2025.106086
Received 30 September 2024; Received in revised form 21 January 2025; Accepted 12 February 2025
Available online 20 February 2025

sentences containing homonyms that primed interpretation toward their low-frequency meanings (e.g., "The seal came up onto the bank of the river"). In a subsequent associate production task that took place 20 min later, participants were more inclined to provide associates related to the low-frequency meanings (e.g., "river"), compared to when these meanings were not primed. This finding, referred to as *word-meaning priming*, has been replicated numerous times (see Rodd, 2020 for a review), across modalities (Gilbert et al., 2018), age groups (Rodd et al., 2016), using eye-tracking (Parker et al., 2023), after both short (e.g., 20–40 min; Rodd et al., 2016) and more substantial delays (e.g., 24 h; Gaskell et al., 2019).

Word-meaning priming has recently been extended to non-homonyms, which by dictionary definition, only have one meaning (e.g., *bathtub*). In Curtis et al. (2022), non-homonymic target words were paired with a probe word referring to a specific aspect of the targets' meaning. For example, the target word, *bathtub*, was paired with the probe, *slip*. Using this probe, a prime sentence was constructed to bias the interpretation of the non-homonym target toward the probe (e.g., "The old man fell while getting out of the bathtub"). Importantly, the probe word was never included in the sentence, thus never directly associated with the target. Following exposure to these sentences, participants completed two tasks shortly after (within 10–20 min): (a) a speeded relatedness judgement task, where participants determined if a target-probe pair (e.g., *bathtub-slip*) is semantically related, and (b) an associate production task, where participants provided the first word that came to mind upon seeing a target word in isolation. Across three experiments, both tasks showed that experiencing a non-homonymic target in a specific context biased participants to subsequently interpret the word in a way that was consistent with the context (i.e., consistent with the probe), providing compelling evidence for word-meaning priming extending to *unambiguous* words.

To explain word-meaning priming, Gaskell et al. (2019) proposed the episodic context account (see also Curtis et al., 2022). This theoretical framework hypothesises that during language comprehension, episodic memory—supported by the hippocampus—contributes to the generation of a new and temporary context-specific representation that binds together the words and concepts in the sentence, extracting the core conceptual information of the linguistic episode. This episodic representation may provide an additional source of information—on top of the established word knowledge in long-term memory—to guide subsequent lexical interpretation. This may bias language users toward the prior context-specific interpretation, providing a possible mechanism for which word-meaning priming in (non-)homonyms arises. A key tenet of the episodic context account, therefore, is that language exposure results in the formation of a new context-specific representation in episodic memory.

In previous word-meaning priming experiments, the target words were always primed toward a probe that was at least moderately related to the target in the first place. For example, in Curtis et al. (2022), *bathtub* was primed toward *slip; predator* toward *shark*; *museum* toward *painting,* etc. This raises the question of whether the role of episodic memory in language comprehension is as robust when the target and probe do not share any pre-existing relationships. This is important as the episodic context account argues that during language comprehension, episodic memory would bind the meanings of *any* content words, even unrelated words, together, influencing their subsequent processing. This prediction is in line with findings from previous studies, reviewed below.

In Prior and Bentin (2003), participants first read meaningful sentences for comprehension (e.g., "Ravit slapped the bee away, and continued eating her apple"). Afterwards, they completed an explicit learning phase, where they encoded 72 word pairs, with some being previously read together in sentence reading (e.g., *bee—apple*), and some being control items not previously read together (e.g., *car—dress*). Finally, participants completed a cued recall task where they recalled the second word in a pair (e.g., what word paired with *bee*?). The authors found that pairs that were read in sentence reading were better recalled than the control pairs, suggesting that language comprehension may have led to the formation of a temporary associative link between the unrelated words. Potentially, then, this may be attributed to episodic memory binding those words together during comprehension.

Similar findings have also been reported in a series of experiments by Ratcliff & McKoon, 1978, Ratcliff & McKoon, 1981a, Ratcliff & McKoon, 1981b; see also McKoon & Ratcliff, 1979, 1986). In these studies, young adults were asked to first read a sequence of sentences: "The youth stole a car. The car sideswiped a pole. The pole smashed a hydrant." Up to 20 min later, participants completed an old/new recognition task, where they decided if *youth* was read in the sentences. The researchers found that response time was faster if the preceding trial showed another word from the sentence (e.g., *car*) than if the preceding trial showed a control word that was not read in sentence exposure. These findings are in line with and predicted by the episodic context account, such that language comprehension bound the unrelated words together in episodic memory, leading to the priming effect observed.

Notably, the above studies tested participants shortly after (<20 min) sentence exposure. We know from a vast literature on paired-associate learning that both related and unrelated word pairs that were *explicitly* encoded together (e.g., *palace—computer*) can endure over long delays (e.g., 12–24 h; Abel et al., 2019; Mak et al., 2024; Payne et al., 2012). Less clear is whether word bindings established during naturalistic language comprehension can survive similarly long periods. This is important, because the episodic context account argues that any episodic bindings between words should survive beyond 20 min and possibly up to 24 h, especially when the delay interval contains a period of sleep (e.g., Ball et al., 2025; Gaskell et al., 2019; Mak, Curtis, et al., 2023; Mak & Nation, 2024). If these bindings between unrelated words can indeed persist over such long intervals, it would provide evidence supporting the robustness of the episodic context account in naturalistic comprehension and provide insights into how the mental lexicon acquires its structure. We, therefore, tested in the current paper whether experiencing unrelated words in a meaningful sentence can influence their subsequent processing across intervals of 5 min, 20 min, and 12 h (including sleep).

In addition to using longer delays, the current experiments also aim to fill a key gap in the literature: As far as we can see, most, if not all, relevant studies (e.g., Graf & Schacter, 1985; McKoon & Ratcliff, 1986; Moss et al., 1995; Prior & Bentin, 2003) relied on concrete and highly imageable words (e.g., *apple*, *grass, hydrant*). It remains unclear whether the effects observed in prior studies generalize to abstract words. The episodic context account suggests that during language comprehension, a context-specific episodic representation is formed, independent of a word's concreteness or imageability. This theory, therefore, explicitly predicts that such representations can affect the processing of all content words (Mak, Curtis, et al., 2023). We tested this possibility by using both concrete and abstract unrelated words in our experiments. Another reason why this is worth investigating is that abstract and concrete words differ in their semantic versatility (Reggin et al., 2021), with abstract words likely having fuzzier semantic boundaries. Potentially then, the effect of recent sentence exposure may have differing effects.

1.1. The present experiments

We conducted five experiments to test whether bindings between unrelated words, established during naturalistic reading, can influence subsequent lexical processing and survive delays up to 12 h. Experiments 1 to 3 each contained a reading phase and a surprise test phase,. (See Fig. 2.)

In the reading phase, participants read both related and unrelated word pairs (e.g., *sword – microwave*) embedded in naturalistic sentences (e.g., "The first thing you notice as you walk into the kitchen is a large **sword** resting on the **microwave**"). Participants read one sentence per trial with the instruction to read for comprehension. No attention was

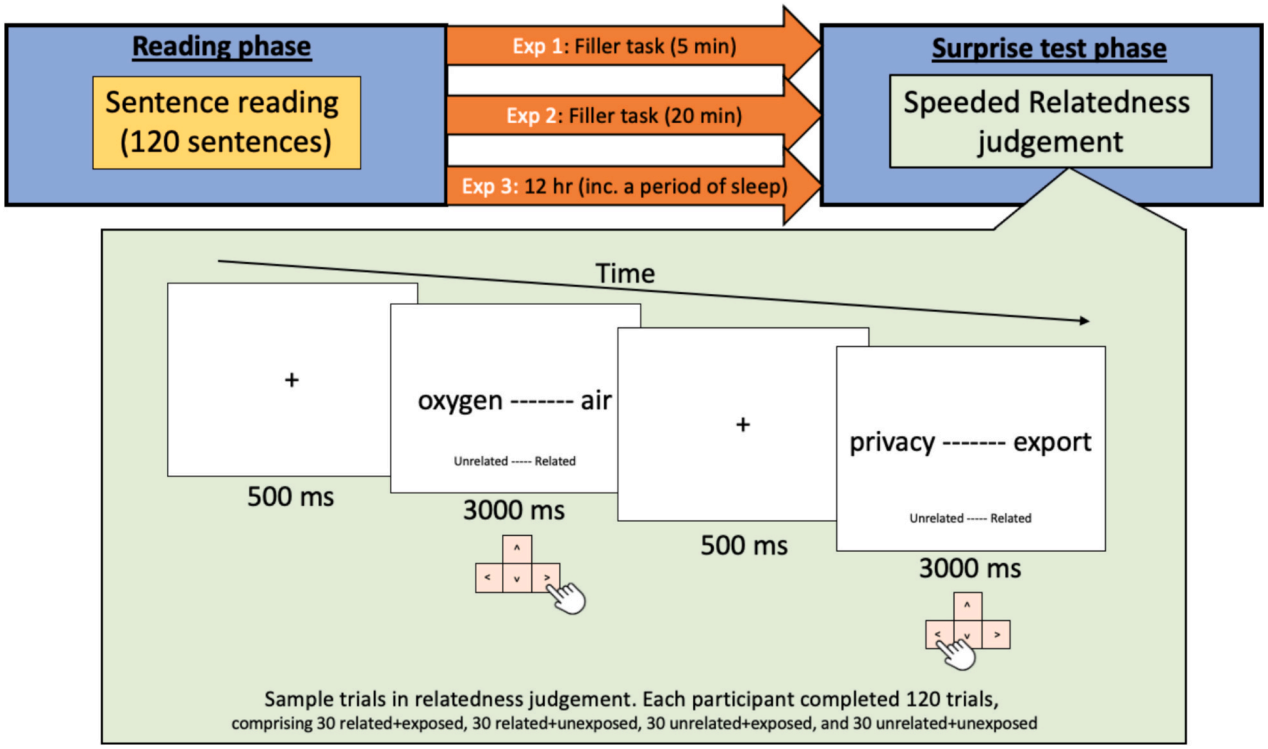

Fig. 1. Experimental procedure of Experiments 1 to 3, with sample trials in relatedness judgement.

drawn to the target words. Each target pair was read twice in the reading phase, each time embedded in a slightly different sentence. The decision to present each pair twice, instead of once, was motivated by our suspicion that a single exposure might be insufficient to establish a strong enough association between unrelated words to elicit an observable effect. Although our experiments were primarily interested in the unrelated pairs, inclusion of the related pairs allowed us to assess the robustness of episodic memory in binding word meanings across different levels of pre-existing associations.

In the subsequent test phase, participants completed a speeded relatedness judgement task, where they decided as quickly as possible if two words in a pair are related in meaning (e.g., *sword – microwave*). We reasoned that if two unrelated words are bound together in episodic memory, participants would be more inclined to judge these words as related and/or make faster judgments.

Experiments 1 to 3 were almost identical, differing only in the delay interval between study and test: Experiment 1 had an interval of 5 min, Experiment 2 an interval of 20 min, and Experiment 3 an interval of 12 h (including a period of sleep). To foreshadow the findings, Experiments 1 to 3 consistently showed a robust exposure effect such that experiencing unrelated words in the same sentence increased their likelihood of being judged as related later (vs. unrelated control pairs that were absent in the reading phase). Experiment 4, also with a 12-h interval containing sleep, tested whether this exposure effect was due to unrelated words being read together in the same sentence or simply to them being present in the reading phase. Finally, Experiment 5 used the same delay interval as Experiments 3 and 4 but incorporated an innovative outcome measure that can estimate the degree of relatedness between two words *without* showing participants the two words together (Walsh & Rissman, 2023).

2. Experiments 1 to 2

2.1. Methods

Experiments 1 and 2 each consisted of two sub-experiments: A and B. Experiments 1 A and 2 A used concrete word pairs only (e.g., *sword – microwave*), while Experiments 1B and 2B used abstract word pairs only (e.g., *privacy – export*). This series of studies began with Experiment 1 A, which was conducted as the third author's (AO) undergraduate final-year dissertation and was pre-registered ahead of data collection (https://aspredicted.org/2SL_VJT). Note that the pre-registered statistical approach was ANOVA, because it was deemed the most appropriate statistical technique among those taught to undergraduates in our department. However, in today's psycholinguistics literature, mixed-effect modelling is the default as it can accommodate by-subject and by-item variance in one model (Jaeger, 2008; Winter, 2013). We, therefore, used mixed-effect modelling in all our analyses and pre-registered this before conducting Experiment 1B (https://aspredicted.org/au7se.pdf). Experiments 2 (and 3) followed the same pre-registered protocol.

2.1.1. Design and pre-registered hypothesis

Each sub-experiment had a 2 (Exposure: Exposed vs. Unexposed) x 2 (Relatedness: Related vs. Unrelated) within-participant design. The pre-registered dependent variables were 1) whether a pair was judged related or not, 2) reaction time (RT) regardless of whether a trial was correct or not.

We hypothesised that experiencing words that are not typically related in a meaningful sentence will subsequently increase their perceived level of relatedness. As a result, we predicted that pairs in the "exposed+unrelated" condition would have a greater likelihood of being judged as related and/or be responded to faster than pairs in the "unexposed+unrelated" condition.

2.2. Participants

The pre-registered inclusion criteria were 1) aged 18 or above, 2) have no known history of any developmental disorders, and 3) native speakers of English. The pre-registered exclusion criteria were 1) failing >20 % of the attention checks in the reading phase and 2) missing >10 % of the trials in relatedness judgement. Characteristics of the

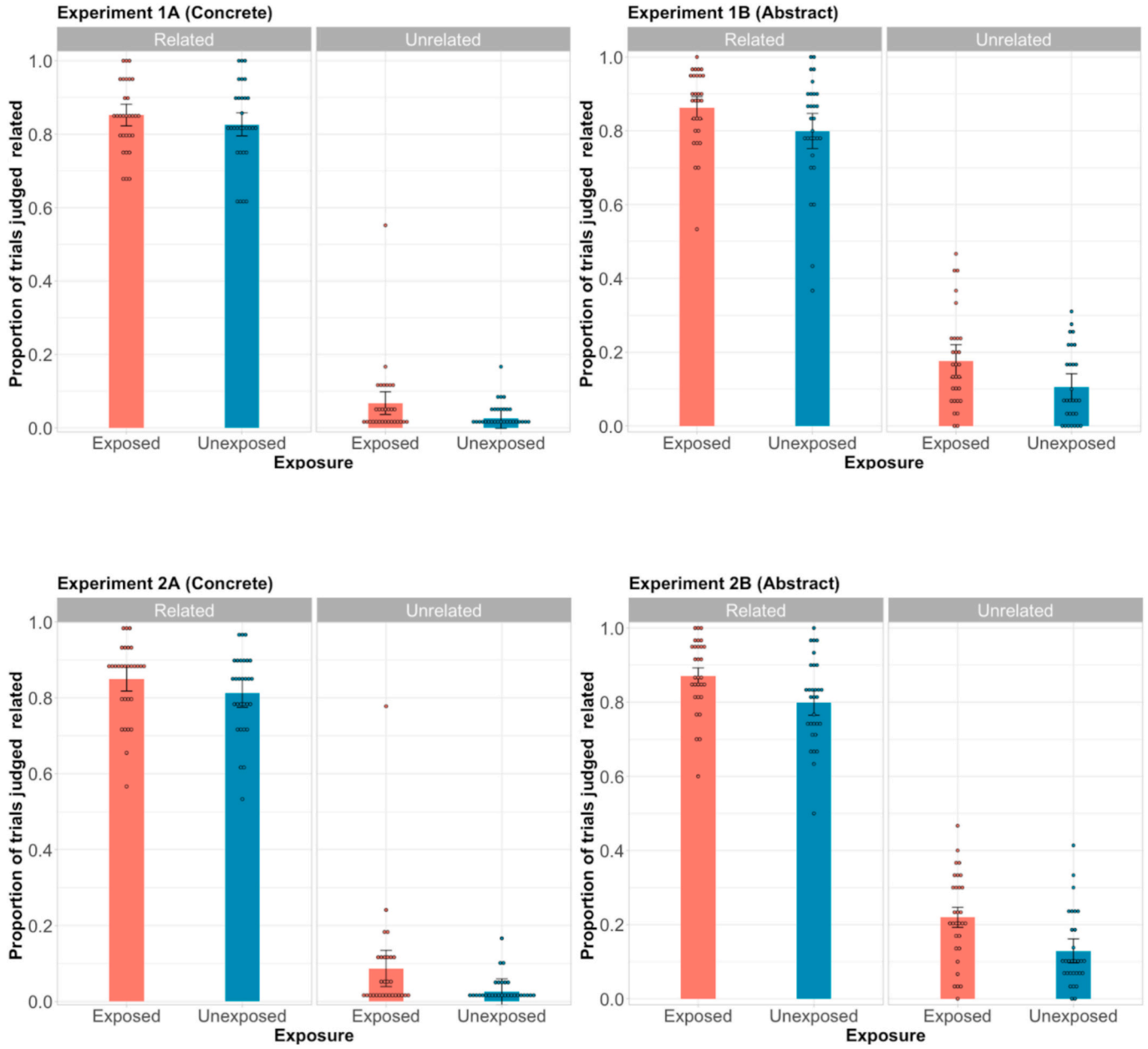

Fig. 2. Proportion of trials receiving a related judgement across Exposure and Relatedness conditions in Experiments 1A, 1B, 2A, and 2B. Each dot represents the mean of an individual participant. Error bars represent 95 % within-subject confidence intervals.

participants in Experiments 1 and 2 are summarised in Table 1.

Table 1
Participants characteristics.

	Experiment 1 A	Experiment 1B	Experiment 2 A	Experiment 2B
Concrete or Abstract words	Concrete	Abstract	Concrete	Abstract
N before exclusion	33	30	30	30
N after exclusion	30	29	29	30
Mean Age (*SD*)	28 (4.47)	31 (5.57)	19 (1.13)	19.2 (0.79)
Gender (F:M: Other)	23:6:1	15:14:0	26:2:1	26:4:0
Recruitment method	Word of mouths	Prolific Academic	Undergraduate participant pool	

2.3. Materials

2.3.1. Concrete word pairs and sentences (Exp 1A & 2A)

Potential word pairs/sentences were identified from reputable fictional and non-fictional sources (e.g., the *Guardian* newspaper). A total of 120 word pairs (hence 240 words) were chosen. These words

contain an average of 6.1 letters ($SD = 2.0$), are high-frequency[1] ($M_{\log \text{freq}} = 2.58$/million; $SD_{\log \text{freq}} = 0.66$/million; van Heuven et al., 2014), concrete nouns ($M_{\text{concreteness}} = 4.66$; $SD_{\text{concreteness}} = 0.51$; Brysbaert et al., 2014). Appendix A shows the full list of word pairs.

Words in sixty of the pairs were related in meaning (e.g., *dentist – teeth*) while words in the other 60 pairs were unrelated in meaning (e.g., *sword – microwave*).[2] A Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) confirmed that the related pairs ($Mdn = 0.474$) had significantly higher cosine values than the unrelated pairs ($Mdn = 0.132$) ($z = -7.086$, $p < .001$).[3]

Each word pair was associated with two sentences, one of which was taken and edited from the source from which it was originally identified, while the other was a paraphrase of the former (see OSF for the full list of sentences). In each sentence, the two target words were separated by at least one but no more than five words (e.g., "The first thing you notice as you walk into the kitchen is a large **sword** resting on the **microwave**"; (Note that the target words were not highlighted through boldface type in the sentences presented to participants). The mean number of words in each sentence was 19.59 ($SD = 8.90$).

2.3.2. Abstract words pairs and sentences (Exp 1B & 2B)

The selection procedure was identical to that for the concrete pairs/sentences. A total of 120 word pairs (hence 240 words) were chosen (see appendix a for the full list). These words contained an average of 6.9 letters (SD = 1.96), were high-frequency ($M_{\log \text{freq}} = 2.54$/million; $SD_{\log \text{freq}} = 0.67$/million), abstract nouns ($M_{\text{concreteness}} = 2.60$; $SD_{\text{concreteness}} = 0.68$). Half of the pairs were related (e.g., *argument – Quarrel*) while the other half were unrelated (e.g., *privacy – Export*). LSA confirmed that the related pairs ($Mdn = 0.453$) had significantly higher cosine values than the unrelated pairs ($Mdn = 0.177$) ($z = -6.161$, $p < .001$)

As per the concrete sentences, there were two sentences for each word pair, one of which was taken from the source from which it was originally identified, while the other was a paraphrase of the former (see OSF for the full list of sentences). In each sentence, the two target words were separated by at least one but no more than five words (e.g., "**Privacy** is paramount for the **export** of personal and confidential goods."). The mean number of words in a sentence was 15.62 ($SD = 5.33$).

2.3.3. Procedure

Experiments 1 to 3 were programmed using Gorilla Experiment Builder (https://gorilla.sc/), following the established procedures from our previous online experiments (e.g., Curtis et al., 2022; Mak et al., 2021a; Mak & Twitchell, 2020). All participants completed the experiment unsupervised, using a desktop computer or laptop, and at a location of their own choosing. They were instructed to complete the study in a reasonably quiet environment where they would not be disturbed for the duration of the study.

2.3.4. Reading phase

In a self-paced reading task, participants read a total of 120 sentences (60 target pairs x 2 sentences). The instructions were: "In this task, you will read some sentences. After you have read a sentence, press the Next button to move to the next one. Following half of the sentences will be a comprehension question, so make sure that you read all of the sentences carefully". Participants were not told to memorise the sentences or that a test on the sentences would follow.

The 120 sentences were divided into two blocks, and each target word pair was read once in each block. Trial and block order was randomised. A trial began with a fixation point at the centre of the screen for 500 ms, followed by the presentation of a sentence. To prevent participants from rushing through the trials, they could only click the Next button 1000 ms after the sentence was presented. Half of the sentences were followed by a simple comprehension question in the format of multiple-choice; these served as attention checks.

2.3.5. Filler tasks

In Experiments 1 A and 1B, participants completed a 5-min digit span task between the reading and test phase. In Experiments 2 A and 2B, participants completed an extended digit span task and watched a 10-min animation that had no verbal dialogue ("Shaun the Sheep"; as in Ball et al., 2025; Curtis et al., 2022). The latter fillers took a total of 17 to 20 min in total to complete.

2.3.6. Test phase

Each participant made speeded relatedness judgments to a total of 120 word pairs [60 exposed (30 related +30 unrelated) + 60 unexposed (30 related +30 unrelated)]. Each trial began with a 500 ms fixation cross, followed by the presentation of a word pair for 3 s, during which the participants indicated whether they thought the two words were related (by pressing the right arrow key on their keyboard) or unrelated (by pressing the left arrow key). Their response and reaction time were recorded. The next trial began as soon as a response was recorded or after 3 s had elapsed. No feedback was provided. Before the experimental trials began, participants completed three practice trials, which provided feedback.

2.4. Results

Following our pre-registered exclusion criteria, trials with a reaction time of ≤ 200 ms were excluded, as well as trials where participants gave no responses (i.e., missed trials).

2.4.1. Judgement

The proportion of trials receiving a related judgement is summarised across Exposure and Relatedness conditions in Fig. 1.

A generalised linear mixed-effect model was fitted to the judgement data from each sub-experiment in R (Version 4.2.2; R Core Teams, 2022). The dependent variable was binary: whether a pair received a 'related' judgement or not (1 vs. 0), so the link function was set to logit. The fixed effects were Exposure (Exposed vs. Unexposed), Relatedness (Related vs. Unrelated), and an Exposure by Relatedness interaction. We used sum contrasts (contr.sum) to code Exposure (Exposed: +1 vs. Unexposed: −1) and Relatedness (Related: +1 vs. Unrelated: −1). The random-effect structure was determined by the 'buildmer' R package (Voeten, 2023), which automatically found the maximal model that was capable of converging using backward elimination (with the 'bobyqa' optimizer). This means that model selection started from the maximal model (as justified by the research design), which is (1 + Exposure * Relatedness | Participant.ID) + (1 + Exposure | Word.Pair). The left-hand side of Table 2 summarises the model outputs.

In Experiments 1B, 2 A, and 2B, there was a main effect of Exposure ($ps < 0.005$) but no interaction effects. This means that both related and unrelated pairs were more likely to be judged as related in the exposed (vs. unexposed) condition, in line with our pre-registered prediction. In Experiment 1 A, there was a main effect of Exposure ($p = .001$), but this was qualified by a significant interaction ($p = .002$). We therefore tested

[1] An anonymous reviewer noted that since high-frequency words tend to appear in varying contexts (i.e., high in contextual diversity), these words may facilitate the formation of arbitrary associations with other words, partially contributing to any Exposure effect in our experiments. This is possible, but given that our item set exclusively comprised high-frequency words, this possibility remains an area for future investigation. Readers interested in exploring the influence of word frequency on the formation of arbitrary associations are encouraged to consult the paired-associate learning experiments conducted by Criss et al. (2011) and Mak and Twitchell (2020).

[2] Two related pairs (e.g., *bee – eagle*) in Experiment 1 A were judged as 'unrelated' by >95 % participants even in the unexposed control condition, so we modified them in Experiment 2 A (e.g., *bee – fly*).

[3] The higher the cosine value between two words, the closer they are in semantic space (i.e., more related). Our LSA was based on the pre-trained LSA semantic space (EN_100k_lsa) provided by Günther et al. (2015).

Table 2
Outputs from confirmatory mixed-effect models examining the effects of Exposure (Exposed vs. Unexposed) and Relatedness (Related vs. Unrelated) in relatedness judgement, Experiments 1 to 4.

DV	Whether a pair was judged as 'related'				RT			
	B	SE	z	p	B	SE	t	p
Experiment 1A (Concrete; 5-min delay)								
Random effect	(1 + Relatedness | Participant.ID)				(1 + Exposure+Relatedness | Participant.ID)			
Exposure	0.314	0.071	4.417	<0.001*	−0.008	0.005	−1.526	0.127
Relatedness	2.665	0.115	23.212	<0.001*	−0.013	0.007	−1.870	0.0614
Exposure x Relatedness	−0.216	0.071	−3.03	0.002*	−0.008	0.004	−2.317	0.0205*
Experiment 1B (Abstract; 5-min delay)								
Random effect	(1 + Exposure*Relatedness | Participant.ID) + (1 + Exposure | Pair)				(1 | Participant.ID)			
Exposure	0.388	0.076	5.107	<0.001*	−0.010	0.004	−2.356	0.0185*
Relatedness	2.603	0.202	12.834	<0.001*	−0.060	0.004	−13.974	<0.001*
Exposure x Relatedness	−0.021	0.102	−0.204	0.838	−0.009	0.004	−2.061	0.0393*
Experiment 2 A (Concrete; 20-min delay)								
Random effect	(1 + Exposure*Relatedness | Participant.ID)				(1 + Exposure+Relatedness | Participant.ID)			
Exposure	0.356	0.126	2.838	0.005*	−0.009	0.006	−1.531	0.126
Relatedness	2.674	0.154	17.349	<0.001*	−0.035	0.006	−5.469	<0.001*
Exposure x Relatedness	−0.21	0.123	−1.712	0.087	−0.018	0.004	−4.154	<0.001*
Experiment 2B (Abstract; 20-min delay)								
Random effect	(1 + Relatedness | Participant.ID) + (1 + Exposure | Pair)				(1 | Participant.ID)			
Exposure	0.402	0.059	6.812	<0.001*	−0.012	0.004	−2.909	0.004*
Relatedness	2.344	0.172	13.649	<0.001*	−0.080	0.004	−19.509	<0.001*
Exposure x Relatedness	−0.091	0.067	−1.363	0.173	−0.009	0.004	−2.308	0.021*
Experiment 3 (Abstract; 12-h delay)								
Random effect	(1 + Relatedness | Participant.ID) + (1 | Pair)				(1 | Participant.ID)			
Exposure	0.243	0.051	12.238	<0.001*	−0.014	0.004	−3.411	<0.001*
Relatedness	2.177	0.178	12.238	<0.001*	−0.055	0.004	−13.860	<0.001*
Exposure x Relatedness	−0.007	0.051	−0.145	0.885	−0.011	0.004	−2.637	0.008*
Experiment 4 (Abstract; Separate; 12-h delay)								
Random effect	(1 + Relatedness | Participant.ID) + (1 | Pair)				(1 + Category + Exposure | Participant.ID) + (1 | Pair)			
Exposure	0.019	0.051	0.380	0.704	−0.005	0.003	−1.64	0.10
Relatedness	2.07	0.178	11.64	<0.001*	−0.070	0.010	−6.78	<0.001
Exposure x Relatedness	0.046	0.051	0.914	0.360	−0.004	0.003	−1.53	0.127
Experiment 5 (Abstract; Swapped; 12-h delay)								
Random effect	(1 + Exposure*Relatedness | Participant.ID) + (1 + Exposure | Pair)				(1|ID)			
Exposure	0.551	0.065	8.523	<0.001*	−0.032	0.003	−10.726	<0.001*
Relatedness	2.555	0.165	15.497	<0.001*	−0.087	0.003	−29.268	<0.001*
Exposure x Relatedness	0.035	0.070	0.497	0.619	−0.022	0.003	−7.324	<0.001*

Note. *$p < .05$. The beta and standard error of the RT analyses have been multiplied by 1000 to improve interpretability. This includes the values reported in the text.

the simple effects of Exposure within the Related and Unrelated conditions, using the 'emmeans' package (Lenth, 2023) in R. Exposure had a significant effect in the Unrelated ($p < .001$) condition, meaning that the exposed unrelated pairs were more likely to be judged as related than their unexposed counterparts. In contrast, while the Related condition patterned in the same direction (exposed > unexposed), the effect of Exposure was not statistically significant ($p = .137$). The cause of this null effect is unclear and somewhat puzzling, especially given the consistent Exposure effect observed in the Related condition of the other sub-experiments. However, since the focus of this paper is the Unrelated condition, the discrepancy has limited implications for our conclusions.

2.4.2. Reaction Time (RT)

The mean RTs are summarised across the Relatedness and Exposure conditions in Fig. 3. Note that both correct and incorrect trials were included in this analysis. That is, if an unrelated trial was judged as related, its RT was not discarded.[4]

We fitted a linear mixed-effect model to the RT data in each sub-experiment. The dependent variable was RT of each trial. As pre-registered, we applied inverse-transformation to the RT data to give a

[4] An anonymous reviewer suggested exploring the RTs for trials judged Related and Unrelated separately. They hypothesised that participants might exhibit slower RTs when responding Unrelated to the exposed (vs. unexposed) unrelated pairs, potentially due to the conflicting episodic bindings established during the reading phase. We ran exploratory analyses on the RT data in Exp 1, 2, 3, and 5, examining the Related (Yes) and Unrelated (No) trials separately (available as Supplementary Material on OSF). Contrary to the prediction, among the No-trials, there was no Exposure effects in either the Related or Unrelated condition.

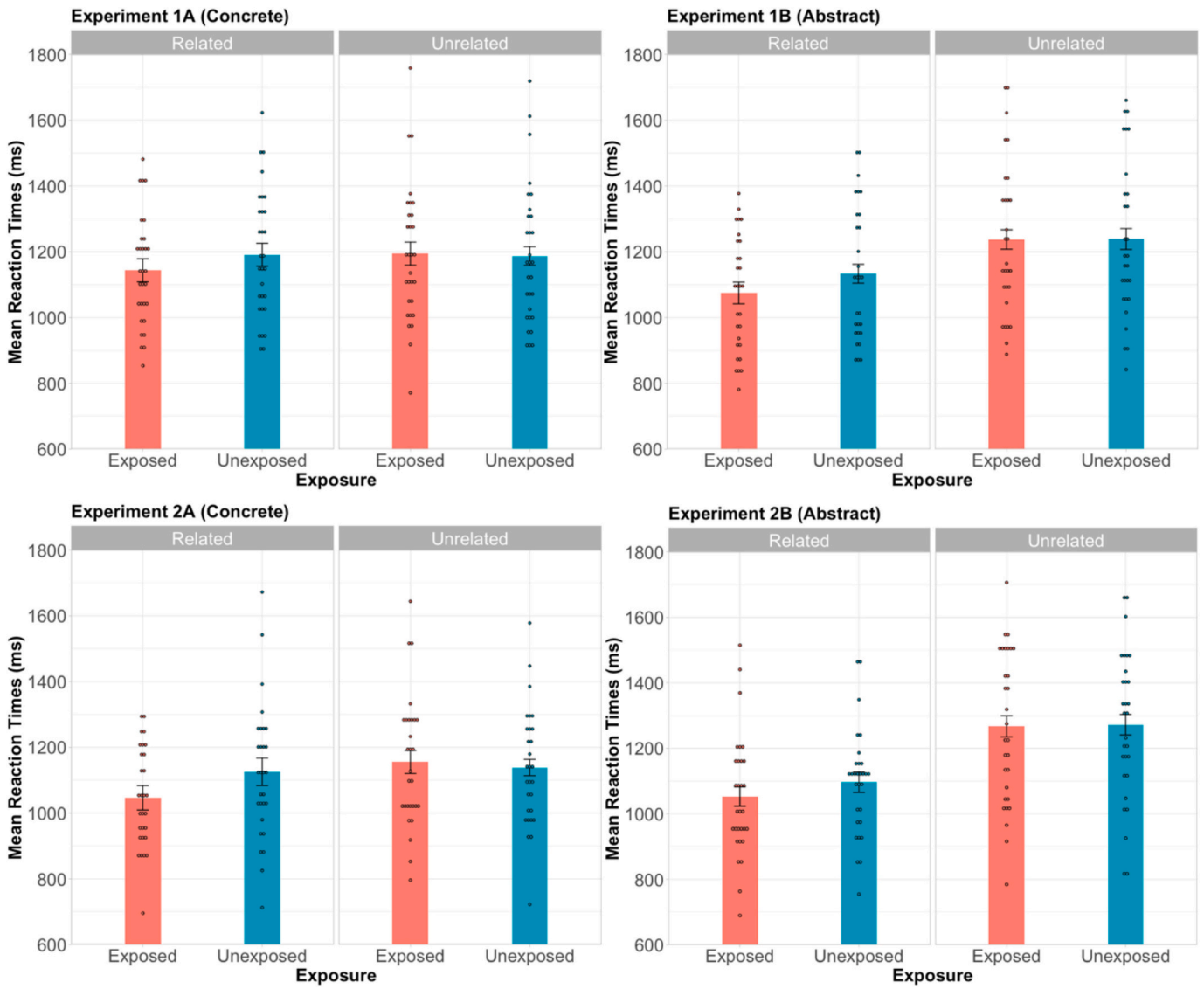

Fig. 3. Mean RTs across Exposure and Relatedness conditions in Experiments 1A, 1B, 2A, and 2B. Each dot represents the mean of an individual participant. Error bars represent 95 % within-subject confidence intervals.

more normal residual distribution. The fixed- and random-effect structures were the same as outlined above. The right-hand side of Table 2 summarises the model outputs.

The Exposure x Relatedness interactions were significant (*ps* < 0.040) in all the sub-experiments, so we followed them up by testing the simple effects of Exposure within the Related and Unrelated conditions, using the 'emmeans' package as pre-registered. Consistently, Exposure had a significant effect in the Related condition (*ps* < 0.01) such that participants made faster judgement to the related pairs in the exposed (vs. unexposed) condition. In contrast, within the unrelated condition, there was no significant difference between the exposed and unexposed pairs (*ps* > 0.80). Therefore, our prediction that sentential exposure may speed up judgement of the exposed (vs. unexposed) unrelated pairs was not supported.

2.5. Exploratory analysis

2.5.1. Concreteness

We explored whether Concreteness (Concrete vs. Abstract) may have moderated the effect of Exposure by combining the data from Experiments 1 A and 1B, as well as Experiments 2 A and 2B. We re-ran the mixed-effect models but added Concreteness as a fixed effect (Details of these analyses are available on OSF). Across experiments and across both the judgement and RT data, Concreteness did not have a main effect or interact with Exposure, suggesting that the effect of Exposure did not significantly differ between concrete and abstract words.

2.5.2. Effect size

We explored the effect size of Exposure (exposed vs. unexposed) in the related and unrelated conditions across experiments in the judgement data. We estimated the effect sizes based on Wilcoxon Signed-Ranked Tests, using the 'rstatix' R package (Kassambara, 2023).

Table 3
Effect size (r) of Exposure in related and unrelated pairs across Experiments 1 to 2 in the judgement data.

	Effect sizes of Exposure (*r*)	
Experiments	Related	Unrelated
Experiment 1A (Concrete)	0.175	0.453
Experiment 1B (Abstract)	0.494	0.526
Experiment 2A (Concrete)	0.347	0.520
Experiment 2B (Abstract)	0.705	0.740

Table 3 summarises the effect sizes (*r*).

The effect size of Exposure ranged from medium to large across Experiments 1 and 2. Consistently, the effect sizes of Exposure were greater.

(1) in the sub-experiments using abstract (vs. concrete) words and.
(2) in the unrelated (vs. related) conditions.

These findings are perhaps not surprising. Regarding the former, abstract words tend to be more semantically versatile than concrete words (Reggin et al., 2021) and their semantic boundary may be more fuzzy, and hence potentially more susceptible to influences from recent exposure. Regarding the latter, related pairs are high in relatedness in the first place, leaving limited room for Exposure to boost their relatedness further (i.e., near-ceiling).

2.6. Discussion

Experiments 1 and 2 conceptually replicated prior findings (e.g., McKoon & Ratcliff, 1986), demonstrating that two (unrelated) words experienced together in a meaningful sentence formed a link and/or are pushed closer together in memory, biasing subsequent lexical processing (up to 20 mins after initial exposure). Our experiments extended prior studies by showing that this exposure effect is not restricted to concrete words but is also observed in abstract words, highlighting its generalisability.

The episodic context account predicts that if the links between unrelated words are supported by episodic memory (as in those explicit links formed during paired-associate learning), these links should be able to bias lexical processing even after a delay of 12 h later (including overnight sleep; Ball et al., 2025; Gaskell et al., 2019; Mak, Curtis, et al., 2023). To test the episodic context account further, we therefore extended the delay interval to 12 h (including sleep) from Experiments 3 onwards. Note that due to budgetary constraints, Experiments 3 to 5 made use of abstract words only. We chose abstract words over concrete words because the effect sizes of Exposure were consistently larger in our abstract words (see Table 3).

3. Experiment 3

This experiment used abstract words only and was identical to Experiments 1B/2B, except the delay interval between the reading and test phase was extended to ~12 h. To control for potential time-of-day effects on language processing (e.g., Mak, O'Hagan, et al., 2023; Natale & Lorenzetti, 1997), all participants completed the reading phase between 8 PM and 10 PM, and the test phase between 8 AM and 10 AM the next day. This means that most, if not all, participants would have had a period of overnight sleep between sessions. This was deliberate, as our previous studies showed that episodic memories of sentences are more likely to survive a sleep period than an equivalent amount of wakefulness (Gaskell et al., 2019; Mak, Curtis, et al., 2023; Mak & Nation, 2024), maximising our likelihood of detecting the desired effects.

4. Methods

4.1. Participants

We ran a power analysis based on the judgement data from Experiment 2B, where Exposure in the unrelated condition had an effect size of $r = 0.74$ (see Table 3). To detect this effect with 80 % power in a within-participant design, a Monte Carlo simulation showed that ~14 participants would be sufficient (assuming alpha = 0.05). We doubled the target sample size to 28 as the effect size of Exposure may be lower after a 12-h delay.

A total of 31 native speakers of English were recruited from Prolific. One participant was excluded from further analysis because they missed over 10 % trials in relatedness judgement. The final sample size was therefore 30 (14 females, 16 males; $M_{age} = 30$; $SD_{age} = 4.09$).

4.2. Results

The judgement and RT data are summarised in Fig. 4 A and C, respectively.

We used the same analysis approach as before, and the model outputs are summarised in Table 2.

In the judgement data, there were main effects of Exposure ($p < .001$) and Relatedness ($p < .001$), but their interaction was not significant ($p = .885$). Therefore, extending the findings from Experiments 1 and 2, sentence exposure increased the likelihood of both related and unrelated pairs receiving a 'related' judgement even after a delay interval of 12 h (Related condition: $M_{exposed} = 0.855$ vs. $M_{unexposed} = 0.804$; Unrelated condition: $M_{exposed} = 0.203$ vs. $M_{unexposed} = 0.154$).

In the RT data, the main effects of Exposure and Relatedness ($ps < 0.001$) were qualified by a significant interaction ($p = .008$). A pairwise comparison showed that Exposure had a significant effect in the Related condition ($z = -4.283$, $p < .001$), such that the exposed (vs. unexposed) pairs were judged faster ($M_{exposed} = 1084$ ms vs. $M_{unexposed} = 1147$ ms). Exposure did not have a significant effect in the Unrelated condition ($z = -0.547$, $p = .585$) ($M_{exposed} = 1249$ ms vs. $M_{unexposed} = 1254$ ms). In sum, these findings mirrored perfectly with those from Experiments 1 and 2.

4.3. Discussion

Experiment 3 extended the delay interval between the reading and test phases to 12 h (including a period of overnight sleep). Despite this longer interval, we replicated the findings from Experiments 1 and 2, demonstrating that unrelated words previously read together in the same sentence were more likely to be judged as related (vs. unexposed control pairs). We suggest that an episodic link was formed between the meanings of the unrelated words during the reading phase and that this link influenced how the words were interpreted even 12 h later. Our findings are in line with and predicted by the episodic context account. However, as suggested by an anonymous reviewer, an alternative explanation is possible. In Experiments 1 to 3, the exposed pairs were each read twice in the reading phase, while the unexposed control pairs were not encountered anywhere in the reading phase. The exposure effects, therefore, could potentially be explained by the exposed words being present in the reading phase, not necessarily a result of them co-occurring in the same sentence (e.g., participants could adopt a strategy whereby overall familiarity from recent experience is used as evidence in favour of relatedness). To address this alternative explanation, we conducted Experiment 4, where words in a target pair are read in separate, rather than the same, sentences.

5. Experiment 4

This experiment was conducted last, in response to a recommendation from peer review, but is presented ahead of Experiment 5 in this article to aid interpretation of the prior results. The experiment tested whether the exposure effects in Experiment 1 to 3 were due to the exposed pairs being read in the same sentence (co-occurrence) or merely to these words being encountered at any point in the reading phase (pre-exposure). To address this question, we replicated Experiment 3 using a new set of sentences. Here, words from each target pair were read in separate sentences during the reading phase, but importantly, the target pairs were presented intact in relatedness judgement, just like earlier experiments. If the exposure effects in Experiments 1 to 3 were driven by sentential co-occurrence, these effects should be abolished here where the words are read in separate sentences. Conversely, if the exposure effects were simply a consequence of the target words being encountered in the reading phase, experiencing them even in separate sentences

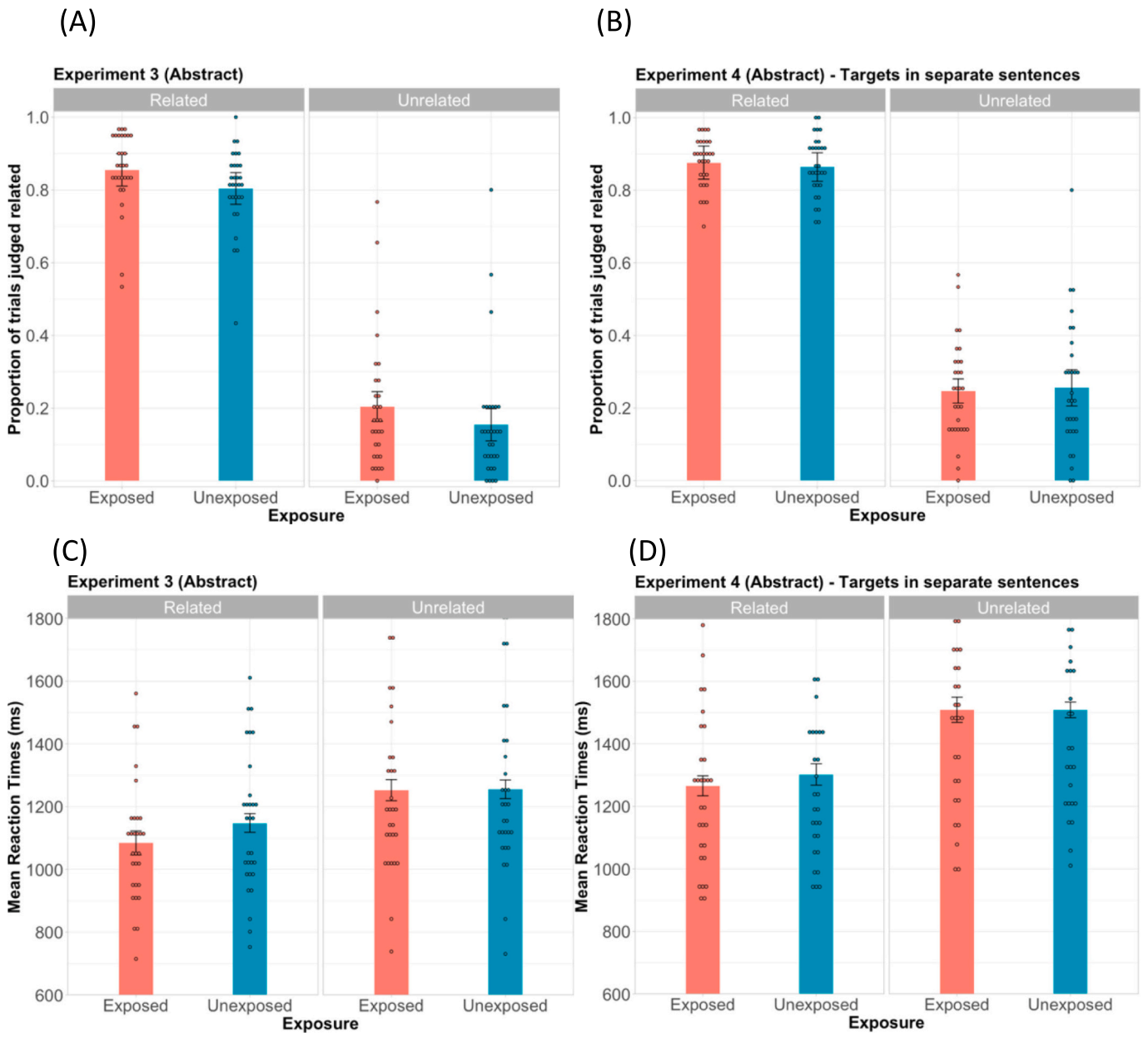

Fig. 4. A and B show the mean proportion of trials being judged related across Relatedness and Exposure conditions in Experiments 3 and 4 respectively, while Fig. 3C and D show the mean RT across Relatedness and Exposure conditions in Experiments 3 and 4 respectively. Each dot represents the mean of an individual participant. Error bars represent 95 % within-subject confidence intervals.

should still bring about the exposure effects. To distinguish between these possibilities, Experiment 4 made use of the design of Experiment 3, maintaining the same delay interval, number of sentences during the reading phase, and relatedness judgement task. The only modification was the use of a different set of sentences.

5.1. Materials and procedures

Based on the 120 abstract pairs from Experiments 1B/2B/3, we constructed a new set of sentences where words from a target pair are presented in separate sentences. To illustrate, consider the unrelated pairs *'weakness—formation'* and *'privacy—export'*. In the earlier experiments, words in each pair co-occurred within the same sentence (e.g., "**Privacy** is paramount for the **export** of personal and confidential goods"). In Experiment 4, *privacy* no longer appeared in the same sentence as *export*, but alongside *weakness* instead (e.g., "Her **weakness** lay in valuing connection over **privacy**, often leaving her secrets exposed."). And similarly, *export* did not appear with *privacy* but with *formation* instead (e.g., "The strategic **formation** of trade alliances boosted the **export** capabilities of the country"). Both the related and unrelated sentences were created to approximate the sentences in Experiments 1B/2B/3 (although note that there were no longer any 'related sentences' as a target word now appeared with another word from a different pair). A total of 240 sentences were created (available on OSF), with an average length of 16.7 words ($SD = 3.19$).

As per the earlier experiments, each target word was encountered in two sentences during the reading phase, and each participant read a total of 120 sentences, half of which followed by a comprehension question. The sentences were presented at a random order; this means that the *weakness-privacy* sentence was separated from the *formation-export* sentence by a random number of intervening sentences. In other words, words from a target pair (e.g., *privacy* and *export*) are now encountered with a significantly greater temporal separation than in Experiments 1B/2B/3, where the two words were read within five words

of each other in the same sentence.

At relatedness judgement, which took place 12 h after the reading phase, participants made speeded judgement to a total of 120 pairs like '*privacy—export*'. As per Experiment 3, of the 120 pairs, 30 belonged to the 'exposed+related' condition, 30 to the 'unexposed+related' condition, 30 to the 'exposed+unrelated' condition, and 30 to the 'unexposed+unrelated' condition. All the task parameters were identical to those in Experiment 1B/2B/3. If the exposure effects observed in Experiments 1–3 do *not* depend on words in a pair co-occurring in the same sentence, we would expect the exposed pairs to receive more related judgments than the unexposed pairs. But if the exposure effects do depend on co-occurrence in the same sentence, there should be little or no difference between the exposed and unexposed pairs here.

5.2. Participants

A total of 35 adults from Prolific took part in Experiment 4. Following the same exclusion criteria as Experiment 3, five participants were excluded from further analysis: Three for missing over 10 % trials in relatedness judgement, one for reporting to have a history of dyslexia, and one for reporting to be a non-native speaker of English. The final sample size was therefore 30 (15 females, 15 males; $M_{age} = 27.8$; $SD_{age} = 4.54$).

5.3. Results

The judgement and RT data are summarised in Fig. 4B and D, respectively. The mean proportion of trials judged related in each of the four conditions is as follows: 'exposed+related' = 0.875 ($SD = 0.12$), 'unexposed+related' = 0.864 ($SD = 0.11$), 'exposed+unrelated' = 0.247 ($SD = 0.09$), 'unexposed+unrelated' = 0.256 ($SD = 0.13$). We fitted the judgement data to a generalised mixed-effect model as before. The results, summarised in Table 2, revealed a main effect of Relatedness ($p < .001$). However, unlike earlier experiments, Exposure was not significant ($p = .704$), nor was the Exposure × Relatedness interaction ($p = .360$).

Turning to the RT data, there was a main effect of Relatedness ($p < .001$), while neither Exposure ($p = .10$) nor the interaction ($p = .127$) was significant, consistent with the judgement data. Exposure showing a null effect in both the judgement and RT data argues against the possibility that experiencing words from a target pair in separate sentences could link the unrelated words in episodic memory and bring about the exposure effects observed in Experiments 1 to 3.

As an additional analysis, and to better understand the null findings, we combined the data from Experiments 3 and 4 to test whether the effect of Exposure was statistically different between these experiments. We fitted the judgement and RT data each to a mixed-effect model, with Relatedness (related vs. unrelated) x Exposure (exposed vs. unexposed) x Experiment (3 vs. 4) as the fixed effects. Recall that words from a target pair were read in the same (Exp 3) or separate (Exp 4) sentences, so if Exposure interacts with Experiment, it suggests that the exposure effects in Experiment 1 to 3 was likely a result of the target words co-occurring in the same context, not a consequence of them merely being present in the reading phase.

In brief, this is supported by the judgement data, where Exposure interacts with Experiment (B = 0.113, SE = 0.036, $z = 3.13$, $p = .002$). No other interaction effects were significant ($ps > 0.34$; the full model outputs are available on OSF.). In the RT data, the Exposure by Experiment interaction did not reach significance (B = −0.005, SE = 0.002, $z = -1.88$, $p = .061$).

5.4. Discussion

In contrast to Experiments 1 to 3, where words from a target pair appeared within the same sentence, Experiment 4 presented these words an equal number of times but in separate sentences. The Exposure effects consistently observed in Experiments 1 to 3 were no longer evident in Experiment 4, indicating that co-occurrence, rather than merely encountering the target words during the reading phase, may be crucial to linking/pushing unrelated words closer in semantic space.

Returning to the exposure effects observed in Experiments 1 to 3, there are some uncertainties with regards to their nature. Specifically, our participants read the target pairs in sentence exposure and then made judgments on these same pairs in the test phase. This provided a retrieval cue that overlaps strongly with the earlier reading episode, contributing to the observed exposure effects. Nevertheless, the episodic context account posits that the context-specific representation formed during language comprehension should guide and shape later judgments even when the retrieval cue does not fully match the original exposure. In the context of our study, the theory predicts that reading an unrelated pair in a sentence should lead to a context-specific representation that will subsequently influence how *one of the* words is interpreted, even if the other word is absent. To further test this hypothesis, we conducted Experiment 5, where we incorporated an innovative word arrangement task (Walsh & Rissman, 2023). This task does not rely on the co- (or successive) presentation of words in a target pair, allowing us to examine whether the context-specific representation induced at reading could bias subsequent lexical processing even when the overlap between the retrieval cue at test and the original exposure is reduced.

6. Experiment 5

This experiment, pre-registered prior to data collection (https://aspredicted.org/t5wg3.pdf), used the same abstract word pairs and sentences from Experiments 1B/2B/3. The delay interval between study and test was 12 h, as in Experiments 3 and 4 (see Fig. 5 for an overview of experimental procedure). The relatedness judgement task was identical to that of Experiment 3 except the word order in a pair was swapped. For instance, if *privacy* preceded *export* in sentence reading, they were shown as *privacy—export* in relatedness judgement in Experiments 1 to 3, but here, it was shown as *export—privacy*. This was motivated by the fact that if their *meanings* were bound together in episodic memory, surface features such as their ordering should have little or no influence on the exposure effects observed in Experiments 1 to 3 (see also Ratcliff & McKoon, 1978).

In addition to swapping the word order, Experiment 5 incorporated a word arrangement task, where participants arranged lists of words on a blank canvas based on semantic relatedness. Here, we outline the details of this task, known as the Similarity-based Word Arrangement Task (SWAT; Walsh & Rissman, 2023).

6.1. Similarity-based Word Arrangement Task (SWAT)

A limitation of Experiments 1 to 3 is that participants encountered target pairs (e.g., *privacy – export*) embedded within meaningful sentences prior to making explicit relatedness judgments about those same pairs. As discussed earlier, this overlap may have contributed (at least partially) to the exposure effects observed in Experiments 1 to 3. Consequently, it remains unclear whether the context-specific representations induced at sentence reading are only activated in the presence of the original pairing, or whether they can guide lexical interpretation independently of such a potent cue. This is important, because the episodic context account predicts that a context-specific representation, once formed, has the potential to shape how words are subsequently interpreted, regardless of whether the original pairing is overtly recalled or cued. In other words, the account posits that the influence of context-specific representations on lexical processing should manifest across a relatively broad range of conditions.

To test whether the exposure effects we observed in Experiments 1 to 3 can be observed even in the absence of a strong retrieval cue, we adopted a word arrangement task (SWAT), pioneered by Walsh and Rissman (2023). In this computer-based task, participants are asked to

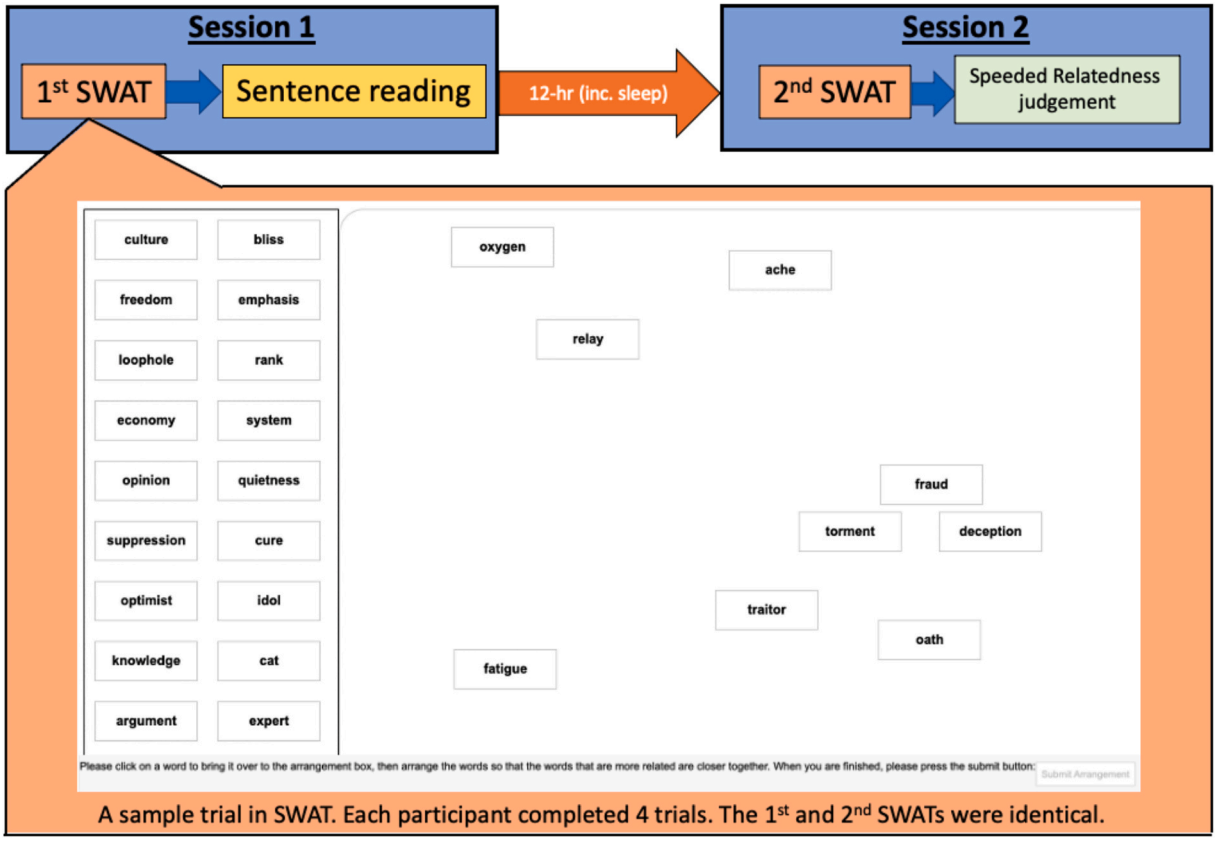

Fig. 5. Experimental procedure of Experiment 5, along with a sample trial in SWAT.

drag-and-drop words with more related meaning closer on a canvas (e.g., Kriegeskorte et al., 2008; Kriegeskorte & Mur, 2012; Richie et al., 2020); however, importantly, the two target words are *not* presented in the same SWAT trial. For example, if a participant read a sentence containing *privacy* and *export* in the reading phase, these two words were split up, with one word in one SWAT trial and the other in a separate one, meaning that participants did not make an explicit judgement about their relatedness to each other, unlike relatedness judgement. This approach eliminated the need to present participants with the original pairings again. In the next subsections, we explain how the degree of relatedness between two target words, which were not shown/judged together, is estimated from SWAT.

In a SWAT trial, 60 target and 2 catch words were shown on the left side of the screen (see Fig. 5 for an example). The set of words in each trial was fixed but their starting order was randomised across participants, and so was the trial order. Participants were instructed to drag-and-drop the 62 words onto a blank canvas, arranging them so that words with more related meanings are placed closer to each other. Once positioned, any word can be moved again. Participants must place all 62 words on the canvas before proceeding to the next, and no time limit was imposed. Once a trial was submitted, the x- and y-coordinates for each word were recorded.[5] There were a total of four trials in each SWAT.

Participants completed SWAT twice, once before sentence reading in Session 1 (baseline measure of relatedness), and once 12 h later in Session 2. The two SWATs were identical, allowing us to track how semantic relatedness at baseline might have changed 12 h after sentence reading. Note that the independent variables in SWAT and relatedness judgement were different: In relatedness judgement, the two (within-participant) independent variables were Relatedness (Related vs. Unrelated) and Exposure (Exposed vs. Unexposed). In SWAT, they were Relatedness (Related vs. Unrelated) and Session (1 vs. 2). In other words, there were no unexposed control items in SWAT, unlike relatedness judgement. However, as explained in the exploratory analysis, we overcame this issue by making use of the 'implicit control items' that are inherently provided by SWAT.

6.1.1. *Word selection in SWAT*

Each SWAT trial showed 60 target words (+ 2 catch words). How these words are selected is best explained with reference to some examples (see Fig. 6).

Reiterating briefly, in sentence reading, participants read 60 word pairs (i.e., 120 target words). In SWAT, each word in those pairs appeared once in two separate trials, resulting in four trials of 60 words. For example, the word *privacy*, which was paired with *export* in sentence reading (see Fig. 6 A), appeared in Trials 1 and 2 (see Fig. 6B). Importantly, *privacy* never appeared alongside *export* in either trial, meaning that they were never judged together. Notably, both words appeared with the same subset of words (e.g., *fact, legend, fiction, myth*), and they serve as the backbone of the SWAT paradigm, explained in more detail in the next subsection. Finally, bear in mind that a target word never appeared in any SWAT trial alongside another subset of words, so for example privacy never appeared with *quantity*, *jeopardy*, or *patent*. These untested pairs (e.g., *privacy—quantity, privacy—jeopardy*) are important in an exploratory analysis, where they serve as the implicit control pairs.

On top of the 60 target words in a trial, we also included two highly related concrete words (e.g., *dog, cat*) as catch items (Note: different

[5] A programming error resulted in recording only the x- and y-coordinates of the first 60 words, omitting the last two. As the starting order within a trial was randomised, this error occurred randomly, and hence, should have minimal impact on the overall results.

Pair #	Word 1	Word 2	Relatedness
1	privacy	export	Unrelated
2	mystery	quantity	Unrelated
3	dynasty	jeopardy	Unrelated
4	question	patent	Unrelated
5	fact	fiction	Related
6	legend	myth	Related
7	faith	soul	Related
8	grammar	language	Related

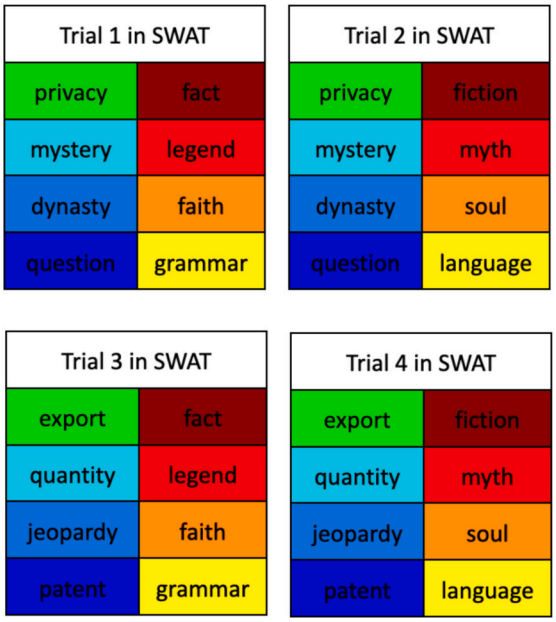

Fig. 6. (A) Word pairs read in sentence reading. (B) Words shown in each trial of SWAT.

catch items in each trial). If participants were paying attention in the task, these catch words should be placed very closely to each other on the canvas, and we pre-registered to use this as an exclusion criterion.[6] In each trial, given 60 target and 2 catch words, there were a total of 62 × 61 = 3782 pairwise (Euclidean) distances that could be computed.

6.1.2. *Estimating relatedness: Imputation*

It is possible to estimate the degree of relatedness between words in a target pair via K-nearest neighbours imputation (Troyanskaya et al., 2001). This procedure was detailed in Walsh and Rissman (2023), so we only provide the most important information here. In brief, the x- and y-coordinates of each word were normalised using the recorded screen dimensions for each participant; then, we computed the Euclidean distance between each of the measured word pairs within a trial (e.g., *privacy—fact, privacy—mystery*; see Fig. 6B). However, since words in a target pair were, by design, never presented together in the same trial, it is not possible to directly compute their Euclidean distance. As a result, this methodology generated an incomplete representational distance matrix. The key innovation of SWAT is that it is possible to *infer* the distance between words in a target pair, because they each appeared with the same subset of words: for example, for the pair, *privacy—export*, each appeared alongside *fact, legend, fiction,* and *myth* (see Fig. 6B). Using each of these pairwise distances, we can then reconstruct the key metrics—the distance between words in a target pair. To do so, we recorded the pixel locations of each word during the SWAT trials, calculated the Euclidean distance between each pair of words, and employed an evidence-weighted average to combine the normalised distances from the four SWAT trials in each session (Kriegeskorte & Mur, 2012). Then, the semantic distance of a target pair was imputed with the KNNImputer function from Python's sci-kit learn package, using the K-nearest neighbours method (Pedregosa et al., 2011) with 40 nearest neighbours and the "distance" weighting function (Walsh & Rissman, 2023). This imputation procedure was performed separately on Session 1 and 2 for each participant. Following this, we converted the distance values to a relatedness value by taking "1 – distance" for ease of interpretation, with higher values indicating greater relatedness. This resulted in a range of relatedness values (for the target pairs) from 0.9666 to 1 (M = 0.9883, SD = 0.004) in Session 1, and 0.9711 to 1 (M = 0.9890, SD = 0.003) in Session 2. Although the absolute range of these values is small due to the normalisation required to combine the SWAT trials, they are highly comparable to Walsh and Rissman's (2023) (e.g., their range in Session 2 was 0.9733 to 0.9972). Given the relatively compressed range of values that results from the distance calculations, a change of e.g., 0.02 in relatedness value, even though small in absolute terms, represents a substantial change in relative terms.

[6] If the catch pair in a trial is not among the top 10 closest pairs in the trial, the participant is said to have failed that catch pair. We pre-registered to exclude participants from the SWAT analysis if they failed the catch pair in ≥ 2 of the 4 trials.

7. Methods

7.1. *Participants*

We did not know a priori the effect size of Session in SWAT, so we assumed an effect of Cohen's d = 0.4, which has been argued to be the smallest effect size of interest in psychological research (Brysbaert, 2019). To detect this with 80 % power in a within-participant design, 52 participants are needed (Brysbaert, 2019). We used this as a guide and recruited 62 participants from Prolific for Experiment 5. Twelve of them did not return for Session 2, so the final sample size was 50 (26 females, 24 males; M_{age} = 28.8; SD_{age} = 4.39). We were unable to recruit more participants due to budgetary constraints. Note that this sample size gave us over 95 % statistical power to replicate the exposure effect (r = 0.596) in the Unrelated condition observed in Experiment 3's relatedness judgement.

7.2. *Procedures and materials*

The experimental procedure is shown in Fig. 5. The experiment was conducted online, as in the previous experiments. SWAT was programmed using jsPsych (De Leeuw, 2015). Session 1, which started between 8 PM and 10 PM, began with a baseline SWAT, which, on average, required 30 min to complete. Afterwards, participants read 60 target pairs, each read twice, embedded in a meaningful sentence. The word pairs, sentences, and reading procedures were identical to those from Experiments 1B/2B/3. Session 1 required approximately 50–55 min to complete.

Session 2 took place ~12 h later between 8 AM and 10 AM the

following day. Here, participants completed SWAT again, which was identical to baseline SWAT. To reiterate, the independent variables in SWAT were Relatedness (Related vs. Unrelated) and Session (1 vs. 2), both of which were manipulated within-participants. Session 2 ended with speeded relatedness judgement, which was identical to those in the previous experiments, except the word order in a pair was swapped here. Session 2 required approximately 40–45 min to complete.

7.3. Results and discussion

7.3.1. SWAT

Following our pre-registered exclusion criteria, two participants were excluded, because they placed more than 35 % of the words at the same position, suggesting they misunderstood the task instructions and/or were not sufficiently engaged with the task. No participants were excluded on the basis of the catch pairs. The following analysis is therefore based on 48 participants.

Fig. 7 shows the change in estimated relatedness values between sessions (Session 2 minus Session 1). A positive change score indicates that words in that condition became more related in Session 2. Note that the control condition (rightmost bar in Fig. 7) will be explained lower down.

As pre-registered, we fitted the estimated relatedness values of the 60 target pairs from each participant in both sessions to a linear mixed-effects model. This model had Session (1 vs. 2), Relatedness (Related vs. Unrelated), and their interactions as the fixed effects. Session and Relatedness were coded using sum contrast (Session 1 = +1, Session 2 = −1; Related = +1, Unrelated = −1). The R package 'buildmer' was used to automatically select the maximal random-effect structure capable of converging using order elimination. The model we report in Table 4 contains this structure: (1 + Session + Relatedness | Participant.ID) + (1 + Session | Word.Pair).

There was a main effect of Relatedness ($p < .001$), indicating that word pairs in the related (vs. unrelated) condition had greater relatedness values in both sessions ($Mdn_{related} = 0.9891$ vs. $Mdn_{unrelated} = 0.9882$). This is reassuring and validates the imputation procedure used to estimate relatedness between words that were never shown together in a SWAT trial. Furthermore, and importantly, there was also a main effect of Session ($p < .001$), indicating that relatedness values for both the related and unrelated target pairs were higher in Session 2 ($Mdn = 0.9892$) than in baseline ($Mdn = 0.9886$). We take this as evidence suggesting that sentence reading led to the meanings of the target pairs being linked up in an episodic context-specific representation (regardless of whether they are related in the first place), pushing them closer in semantic space for at least 12 h. Crucially, as words in a target pair were never shown together in SWAT, a main effect of Session here indicates that a strong retrieval cue (i.e., the original pairing) is not necessary for these episodic context-specific representations to exert an influence. This highlights the relatively pervasive and enduring nature of these representations in shaping lexical processing.

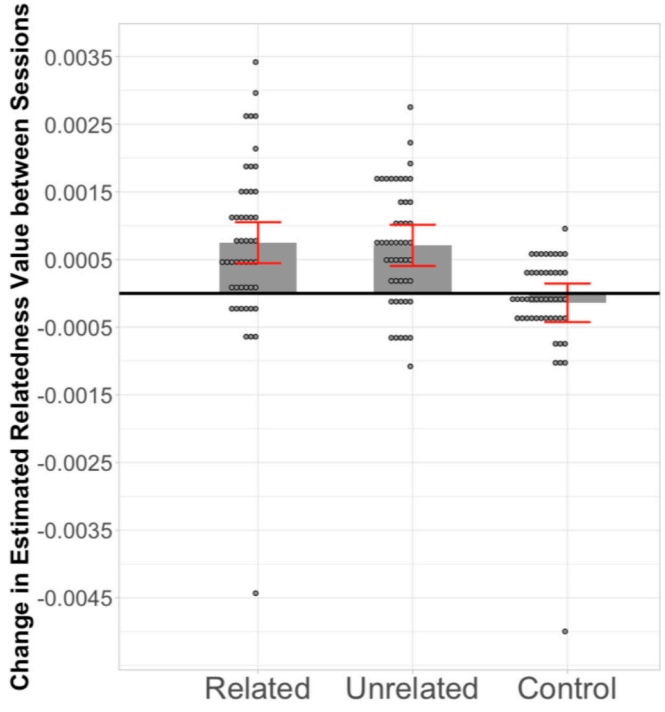

Fig. 7. Change in mean relatedness values across sessions (Session 2 minus Session 1) in each type of pair, with positive values indicating greater relatedness following sentence reading. Each dot represents the mean of an individual participant in that condition, and the error bars represent 95 % within-subject confidence intervals.

Table 4
Outputs from the confirmatory mixed-effect model examining the effects of Session (1 vs. 2) and Relatedness (Related vs. Unrelated) on the estimated relatedness values of the target pairs in SWAT.

Fixed effects	*B*	*SE*	*t*	*p*
Intercept	988.60	0.210	4706.509	<0.001
Relatedness	0.448	0.10	4.490	<0.001
Session	−0.362	0.07	−5.185	<0.001
Relatedness x Session	−0.001	0.06	−0.151	0.88

Note. *$p < .05$. The beta and standard error have been multiplied by 1000 to improve interpretability.

7.3.2. Relatedness judgement

The pre-registered exclusion criteria were the same as Experiments 1 to 3. None of the 50 participants met our exclusion criteria, so the analysis here is based on all 50 participants. The judgement and RT data are summarised in Fig. 8 A and B respectively. In brief, the overall pattern of results was roughly the same as in Experiments 1 to 3.

We used the same analysis approach as before, and the model outputs are summarised in Table 2. In the judgement data, there were main effects of Exposure ($p < .001$) and Relatedness ($p < .001$). Their interaction was not significant ($p = .619$). Therefore, even though we swapped the word order in the pairs, we successfully replicated the prior finding that sentence exposure increased the likelihood of the exposed (vs. unexposed) pairs being judged as related (Related: $M_{exposed} = 92.9$ % vs. $M_{unexposed} = 85.1$ %; Unrelated: $M_{exposed} = 25.6$ % vs. $M_{unexposed} = 14.7$ %), and that this effect survived for at least 12 h.

Turning to the RT data, reiterating briefly, both correct and incorrect trials were included in this confirmatory analysis (e.g., for an unrelated pair, it may have been judged unrelated by some participants and related by others; the RTs for both were not differentiated, as pre-registered). There were main effects of Exposure and Relatedness ($ps < 0.001$), but they were qualified by a significant interaction ($p < .001$). A pairwise comparison using 'emmeans' replicated the findings in Experiment 1–3, showing that Exposure had a significant effect in the Related condition ($z = -12.79$, $p < .001$), such that the exposed items ($M = 1027$ ms) were responded to faster than the unexposed counterparts ($M = 1288$ ms). Interestingly, in contrast to the null findings in Experiment 1–3, Exposure also showed a significant effect in the Unrelated pairs ($z = -2.40$, $p = .016$), with those in the exposed (vs. unexposed) condition being judged faster ($M_{exposed} = 1288$ ms vs. $M_{unexposed} = 1319$ ms). In other words, pairs in the exposed (vs. unexposed) condition—regardless of their relatedness—were judged at a faster speed after sentence reading. Potentially, this means that sentence exposure bound up the exposed unrelated words together, enhancing the speed with which a decision was made. It is not clear at present why in the previous RT data, Exposure did not have a significant simple effect in the Unrelated pairs. This may be a consequence of the previous experiments

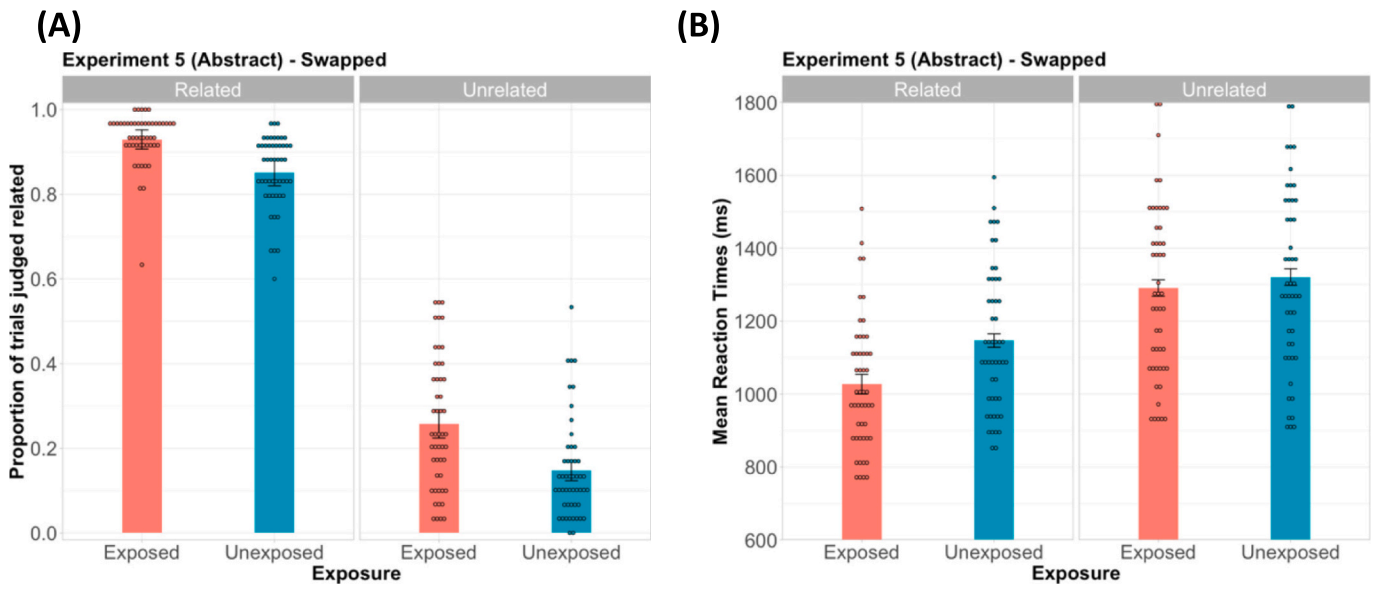

Fig. 8. A shows the mean proportion of trials being judged related and Fig. 7B shows the mean RT, summarised across Relatedness and Exposure conditions in Experiment 5. Each dot represents the mean of an individual participant. Error bars represent 95 % within-subject confidence intervals.

being underpowered, as they had smaller sample sizes (≤30 participants vs. 50 in Exp 4). It might also be related to the fact that participants in this experiment had an additional exposure to the individual target words (although not in pairs) due to second SWAT test, which took place before relatedness judgement. Regardless, a significant Exposure effect in the Unrelated condition here complements a strong body of evidence from relatedness judgement and SWAT, further supporting that sentence exposure can link up the meanings of two unrelated words in episodic memory.

7.4. Exploratory analysis: Implicit control pairs in SWAT

In SWAT, we observed an increase in the estimated relatedness values for the target word pairs from Session 1 to 2, a finding that we attributed to sentence exposure. However, an alternative explanation is possible: Since participants completed SWAT twice, a practice effect might have caused them to arrange the words more efficiently in Session 2, potentially inducing a *global* increase in relatedness values. To test this possibility, we conducted an exploratory analysis to examine if relatedness values increased across sessions among the 'implicit control pairs'.

As explained earlier in word selection, a target word was never shown with a small subset of target words from other pairs in SWAT. For example, for the pair '*privacy—export*', *privacy* was never shown alongside control words like *quantity, jeopardy*, and *patent* in any SWAT trials (see Fig. 6B). This means that *privacy—quantity, privacy—jeopardy*, and *privacy—patent* could serve as the "implicit control pairs", allowing us to test whether their semantic relatedness, as estimated from imputation, might have changed between Sessions 1 and 2. Since these implicit control pairs comprise words that are semantically unrelated (e.g., *privacy—quantity*), this exploratory analysis excluded target pairs from the related condition and compared the implicit control against the unrelated target pairs only.[7]

Within an individual participant, there are 870 implicit control pairs per session (30 unrelated target words x 29 control items). Their relatedness values were estimated using the same imputation procedure as in the target pairs. The rightmost bar in Fig. 7 shows the change in estimated relatedness values for the implicit control pairs between sessions, and importantly, the 95 % confidence interval crosses 0, unlike the target pairs.

We first compared the estimated relatedness values for the implicit control pairs against the unrelated target pairs in a linear mixed-effect model with Session (1 vs. 2), Pair Type (Implicit control vs. Unrelated target pairs), and their interaction as the fixed effects We anticipated a Session by Pair Type interaction if the effect of Session differed significantly between the two types of pairs. This was indeed the case ($p < .001$) (see Appendix B for model output).

To unpack the interaction, we fitted the estimated relatedness values for the implicit control and unrelated target pairs to two separate linear mixed-effect models. Session (1 vs. 2) was the sole fixed effect. The effect of Session was not significant in the control pairs ($B = -0.071$, $SE = -0.073$, $t = -0.968$, $p = .333$) but was significant in the unrelated target pairs ($B = 0.704$, $SE = 0.164$, $t = 4.293$, $p < .001$).

Overall, these exploratory findings suggest that completing SWAT twice did not significantly boost the estimated relatedness values for the implicit control pairs, indicating that our findings from the target pairs cannot be attributed to a practice effect or a global increase in relatedness values across sessions. Instead, this analysis provides compelling evidence that sentence reading was responsible for the increase in relatedness values among the target pairs.

8. General discussion

This paper used words with limited pre-existing associations (e.g.,

[7] We obtained the LSA-cosines for all the implicit control pairs, using the pre-trained LSA semantic space from Günther et al. (2015). The median was 0.184. This is comparable to that for the unrelated target pairs, whose median was 0.177 (see the Materials section of Experiments 1/2). Note that the median LSA-cosine among the related target pairs was substantially higher (0.453); therefore, it is reasonable that we compared the implicit control pairs with the unrelated, rather than the related, target pairs.

privacy—export) to shed light on the involvement of episodic memory in naturalistic language comprehension. We reasoned that if episodic memory is involved, unrelated words read together in a sentence would have their meanings linked together in a context-specific episodic representation, which could subsequently bias language users to perceive these words as more related and/or make faster judgement.

In a series of experiments, we asked participants to first read unrelated word pairs in meaningful sentences (e.g., "Privacy is paramount for the export of personal and confidential goods") before completing speeded relatedness judgement 5 min (Experiment 1), 20 min (Experiment 2), and 12 h (Experiments 3 and 5) later. Across all the delay intervals, unrelated words that were read together in a meaningful sentence were more likely to be judged as related, compared to their unexposed counterparts. Experiment 4, which presented words from a target pair in separate sentences, provided additional evidence that the exposure effects were likely due to the target words co-occurring in the same sentence, not simply to them being encountered before in the reading phase. Finally, in Experiment 5, we employed a word arrangement task (SWAT; Walsh & Rissman, 2023) that allowed us to infer the levels of relatedness between words in a target pair even when they were never shown together. The SWAT task, in line with speeded relatedness judgement, revealed that 12 h after sentence reading, the estimated relatedness of the target pairs increased, compared to baseline. Importantly, this increase was not observed among the implicit control pairs, whose relatedness values remained stable across sessions.

In summary, the results from two distinct outcome measures converged to demonstrate that the perceived relatedness between target words was enhanced after they were read together in meaningful sentences, providing support for the notion that these words are linked together and/or pushed closer in episodic memory, and this can influence subsequent lexical processing. This robust body of evidence extends the literature in at least three significant ways.

First, while most, if not all, relevant prior studies relied solely/primarily on concrete words (e.g., Curtis et al., 2022; McKoon & Ratcliff, 1986), ours used both concrete (Exp 1 A, 2 A) and abstract words (Exp 1B, 2B, 3, 4). The increase in perceived relatedness following sentence reading was equally evident in both types of words, at least after a 5- and 20-min delay interval. This provides evidence for the episodic context account, which predicts that the episodic representation derived from language comprehension can influence the processing of *all* content words (Mak, Curtis, et al., 2023).

Second, to the best of our knowledge, we are the first to demonstrate that reading unrelated words together (e.g., *privacy – export*) in a sentence can subsequently influence their perceived semantic relatedness even after a 12-h delay interval (including sleep). The longevity of this effect is discussed in greater detail in the next subsections.

Third, prior experiments reporting exposure-induced priming between unrelated words (e.g., Carroll & Kirsner, 1982; McKoon & Ratcliff, 1980; Neely & Durgunoğlu, 1985; Prior & Bentin, 2003) bear similarities to our Experiments 1–4, where participants read some word pairs before completing a task where the target words were presented as intact pairs or in close proximity to each other (e.g., successive trials in lexical decision). In other words, the test phase provided the original pairing, which could have served as a potent retrieval cue to the prior exposure phase. This raises the question of whether the original pairings were necessary to drive the observed exposure effects at test. Our SWAT task in Experiment 5, which did not involve simultaneous or successive presentation of the target pairs, offers what may be the first evidence in the relevant literature that the exposure effect is not contingent on the original pairings being shown at test. We demonstrated that when words in a target pair were separated and individually compared to a completely different set of words, their inferred levels of relatedness increased after sentence exposure. The fact that we observed the exposure effects in relatedness judgement and SWAT—two tasks of very different nature—highlights that the context-specific representations induced at reading can perhaps influence lexical interpretation across a broad range of linguistic circumstances. Potentially, then, these representations are routinely shaping how we interpret language on-line and provide a mechanism through which our semantic knowledge is refined throughout the lifespan (see Mak, Curtis, et al., 2023 for a discussion and Duff et al., 2020 for a neural perspective).

8.1. *Why do the bindings between unrelated words survive for as long as 12 h?*

Some memory and psycholinguistic theories propose that when a word (or item) is encountered, its surrounding context is also automatically encoded (e.g., Hayes et al., 2007; Jones et al., 2017; Jones & Estes, 2012; Nation, 2017; Nelson & Shiffrin, 2013). Consistent with these theories, incidental word-learning studies have shown that even though participants were told to read for leisure, the nature of the linguistic context surrounding a novel word had a measurable effect on how well the word was later remembered (Dong et al., 2024; Hulme et al., 2023; Johns et al., 2016; Mak et al., 2021b; see Mak & Nation, in preparation for a review), supporting the notion that a word's contextual information is automatically, and perhaps obligatorily, encoded during language comprehension. Results from the current studies are in line with these word-learning studies, demonstrating that even when we encounter *highly familiar* words (e.g., *privacy*, *export*), they are automatically encoded with their surrounding context, forming a context-specific representation. Furthermore, as demonstrated in the SWAT task, these representations are activated even in the absence of a strong retrieval cue (i.e., the original pairing), highlighting the possibility that these representations are automatically retrieved[8] to aid lexical interpretation of either word when encountered in a wide range of contexts.

Here, we turn to the longevity of our exposure effects: why would our episodic system retain the bindings between unrelated words for as long as 12 h, even though participants were not asked to remember them? Wouldn't it be more efficient if these bindings are pruned shortly afterwards so that there is more room for retaining important information? We propose that there are at least two reasons why the episodic network involved in language comprehension may retain these bindings.

First, retaining seemingly unrelated bindings for some time allows language users to better deal with the burstiness inherent in language (e.g., Myslín & Levy, 2016). For example, a corpus analysis on American newspapers showed that words that have recently appeared have a higher chance of reappearing in the near future than words featured a long time ago (Anderson & Schooler, 1991). This bursty pattern, as argued by Anderson and colleagues (Anderson & Schooler, 1991; Schooler & Anderson, 1997), may exert pressure on human memory to prioritise recent information, which may, in turn, enable us to be more efficient when processing repeated information. Therefore, if the words *privacy* and *export* are encountered now, even if they are unrelated in the first place or remembering them brings no obvious benefit, our episodic memory system may have been 'trained' to hold onto them for some time in case it may become relevant or useful in the near future. Relatedly, this tendency may also reflect the brain's strategy to ensure that memory traces survive long enough to be consolidated into longer-term memory (e.g., Cychosz et al., 2024).

Second, language is constantly evolving (e.g., Hills & Adelman, 2015), as evidenced by content words losing existing meanings and gaining new ones from time to time (e.g., Blank, 1999). The ability to retain and adapt to new and seemingly unrelated pairings may enable individuals to cope with the dynamic nature of language (Mak, Curtis, et al., 2023). For example, before the COVID-19 pandemic, 'furlough' and 'social distancing' were largely unrelated concepts, but the pandemic created a new context in which they became interconnected (e.g., Davis, 2023). The ability to retain previously unrelated links for a

[8] Although it is likely that it could be suppressed.

significant amount of time is likely crucial for establishing these connections in the mental lexicon (e.g., Mak, 2019). This process enables our lexicon to adapt to novel linguistic contexts and integrate new meanings as they emerge.

In summary, our findings are in line with theories suggesting that contextual information of a word is automatically (and perhaps obligatorily) encoded, and that such information could be retrieved (perhaps also automatically) to aid lexical interpretation, even in the absence of a potent retrieval cue (e.g., the original pairing). We propose that our memory/language system may be primed to hold onto meaningful yet previously unrelated bindings for a significant amount of time, enabling us to efficiently navigate the burstiness and clustering inherent in language and adapt effectively to its perpetual evolution.

8.2. Could our findings be supported by implicit memory?

We attributed our exposure effects to episodic memory linking the meanings of the unrelated words together during language comprehension. Could our findings be supported by memory systems other than episodic memory? If we assume that the semantic system is strictly concerned with *context-free* memory (Tulving, 1993), we think it is unlikely that our findings—where unrelated words were linked via context—are supported by semantic memory (see also Dagenbach et al., 1990). While we favour an episodic memory account, an implicit memory account could also be considered.

Implicit memory refers to hippocampus-independent memory that influences our behavior without conscious recollection. In an experiment by Graf & Schacter, 1985; see also Schacter & Graf, 1986), participants engaged in a stem completion task where they were shown JAIL—STR_____ and filled in the blank using any word starting with STR (e.g., *strawberry, strike, strim, strip, strong* were all correct answers). There were two groups of participants: one completing only the stem completion task, and another also performing a sentence production task beforehand, where participants constructed a meaningful sentence using a pair of unrelated words like *jail* and *stranger*. Importantly, among participants who engaged in sentence production, half were amnesic, exhibiting marked deficits in episodic memory, while the other half were healthy controls matched on age, IQ, and education level. In the subsequent stem completion, when given JAIL—STR_____, the amnesic patients filled the blank with *stranger* at a substantially higher rate (~32 %) than the control participants who did not take part in sentence production (~13 %) and at the same rate as the healthy controls (~32 %). These findings indicate that sentence production forged a novel connection between the unrelated words; however, given the episodic impairments in the amnesic patients, it was argued that those novel connections were supported by implicit, rather than episodic, memory (Graf & Schacter, 1985). However, a follow-up analysis of the data from Graf and Schacter (1985), together with findings from a replication study (Schacter & Graf, 1986b), showed that implicit memory for new associations was present only in mildly amnesic individuals, not in those with severe amnesia. This implies that any new bindings between unrelated words observed in mildly amnesic patients might actually be due to their residual episodic memory capabilities (Tulving et al., 1991). These conclusions were further validated by two subsequent studies conducted by Shimamura and Squire (1989), which modelled after Graf and Schacter's experiments but failed to replicate their findings in severely amnesic patients. Such neuropsychological evidence makes it hard to argue that our exposure effects were solely supported by implicit memory. Future neuropsychological studies may be useful for understanding the contribution of hippocampus-(in)dependent memory to the flexibility of lexical knowledge/organization (e.g., Covington & Duff, 2025; Duff & Brown-Schmidt, 2012; Duff et al., 2020).

8.3. Semantic vs. Associative similarity

An anonymous reviewer raised an important question regarding the mechanism by which our target words became more related after sentence exposure: Does this occur through increased semantic or associative similarity (e.g., Cree & Armstrong, 2012; Mirman et al., 2017)? Semantically similar words share overlapping meanings and features (e.g., *chair – table*), while associatively similar words co-occur frequently in language but lack shared meaning/features (e.g., *pillar – society*). Our experiments cannot definitively determine whether the observed exposure effects reflect episodic memory influencing semantic and/or associative similarity. However, we contend that the latter may be the primary driver. In our experiments, words from an unrelated pair (e.g., *privacy – export*) share little to no shared meaning/features. It, therefore, seems unlikely that just two exposures in the reading phase could significantly alter their underlying meanings to increase semantic similarity. Second, while episodic memory is generally thought to play a crucial role in associative learning (e.g., Giovanello et al., 2004; Mayes et al., 2004), it remains unclear whether episodic memory traces can immediately alter long-term semantic representations or their relations (see Gaskell et al., 2019). Finally, computational research suggests that associative relations—arising from co-occurrence in language—may be a key organizing principle for the mental lexicon (e.g., Jones & Mewhort, 2007; Kumar, 2021; Landauer & Dumais, 1997). While our results primarily reflect short-term associative changes, they may also represent the early stages of this broader mechanism, whereby initially unrelated words become more associatively similar through repeated co-occurrence. In sum, we are inclined to conclude that the observed exposure effects were primarily driven by an increase in associative similarity.

8.4. Future directions

The idea that episodic memory contributes to the language system is not new and has been articulated in previous theories concerned with single-word processing (e.g., Goldinger, 1998; McKoon & Ratcliff, 1986; Tenpenny, 1995) and novel word learning (e.g., Davis & Gaskell, 2009). What makes the episodic context account distinct is that it posits the routine involvement of episodic memory in *naturalistic language comprehension* (Gaskell et al., 2019; Mak, Curtis, et al., 2023). To further ascertain the involvement (and limits) of episodic memory in language comprehension, several promising research avenues can be explored.

8.4.1. The role of sleep

Since the context-specific representation derived from language comprehension may be of an episodic nature, the episodic context account predicts that it should be susceptible to sleep-related memory effects, just like any other newly acquired episodic memories (Gaskell et al., 2019). There is some evidence supporting this prediction. For example, Mak, Curtis, et al. (2023) showed that experiencing a word-class ambiguous word like *loan* in a specific word class (e.g., verb) in a sentence primed participants to later use that word in the same word class as before, and importantly, this priming effect was stronger after a night's sleep than after an equivalent amount of daytime wakefulness. This suggests that the context-specific representation derived from language comprehension may be susceptible to sleep-related memory effects, providing support to its episodic nature. Potentially, future studies can add a sleep manipulation to the current design (e.g., see Mak, 2024 for a methodological description): If the level of perceived relatedness between unrelated words is greater after a nap (vs. an equivalent amount of wakefulness), it will add further strength to the argument that the unrelated words were bound together in episodic memory.

Notably, in Experiments 3 and 4, participants started Session 1 in the evening and Session 2 in the morning the next day, so most, if not all, participants would have had a period of overnight sleep. The observation of an increase in perceived relatedness between unrelated words in these experiments is intriguing, especially in the context of the synaptic homeostasis hypothesis (Tononi & Cirelli, 2006, 2014). This hypothesis proposes that a key function of sleep is to prune unimportant

information, providing a refreshed cognitive state for new information and learning. If this is correct, it suggests that the episodic connections between unrelated words were not deemed unimportant by the memory/language system. This begs the question of what makes a piece of information unimportant and what bits of language input may be pruned by sleep. These are important questions that future studies may attempt to address.

Finally, relevant to the discussion of sleep-mediated effects, we acknowledge that Experiments 3 and 4 were conducted at a specific time of day (i.e., reading in evening, and test in morning), while our Experiments 1 and 2 did not control for the time of day. The discourse processing literature indicates a time-of-day variation in reading strategies, with readers tending to engage in more literal processing in the morning and more inferential processing in the evening (Natale & Lorenzetti, 1997; Oakhill, 1986). While it is unclear whether such variations influence the Exposure effects observed in our study, it is plausible that reading in the morning could promote more literal processing, potentially strengthening the binding between target words and enhancing the observed effects. This represents an intriguing empirical question for future investigations.

8.4.2. *Co-occurrence: shared context or temporal proximity?*

Experiment 4 provided evidence that our observed exposure effects are likely a result of sentential co-occurrence, rather than the target words being present in the reading phase. However, for words that co-occur in a sentence, not only do they typically share the same context, but they are also read in close temporal proximity to each other (e.g., in Exp 1, 2, 3, 5, words from a target pair were separated by no more than five words in the reading phase). Our findings cannot tease apart the relative contribution of shared contexts and temporal proximity, as they were confounded in our experiments. Findings from Ratcliff and McKoon (1978), however, seem to support that shared context may be more important. They showed that unrelated words experienced within a propositional structure, defined as a verb and its arguments (Kintsch, 1974), are more likely to prime each other than unrelated words experienced in separate propositional structures, even when surface distance (i.e., temporal proximity) was held constant. Potentially then, if the target unrelated words in our experiments were placed in different propositional structures or separated by an event boundary (Radvansky & Copeland, 2010) while the surface distance remains the same as in the current experiments, the exposure effects may reduce or even disappear. Future studies are needed to test this prediction, and to push the current findings further, future studies can use neuroimaging (e.g., Ezzyat & Davachi, 2011) to investigate how propositional structure/event boundary may influence the binding of unrelated words on a neural level.

8.4.3. *Context overlap*

Our experiments were conducted online. While we did not gather information about participants' physical environment, it is likely that most of them completed the study and test phases in the same environment. This environmental consistency could have acted as a retrieval cue (e.g., Godden & Baddeley, 1975), partially driving the exposure effects we observed. To deepen our understanding of what the context-specific episodic representation captures during language comprehension, future studies can consider changing the contextual environment between study and test. This would provide insights into the extent to which environmental context is captured in the representation and how influential it is to our findings.

9. Conclusion

Across five experiments, we found compelling evidence that experiencing a pair of unrelated words in a meaningful sentence resulted in these words being linked and/or pushed closer in semantic space, increasing their levels of perceived relatedness—a robust effect that persists for at least 12 h after initial exposure (including sleep). We interpreted these findings as suggesting that episodic memory contributed to the generation of a context-specific representation that links words in a discourse together, and in turn, these representations may bias how words (and perhaps discourse) are subsequently interpreted. The SWAT task in Experiment 5, which did not involve simultaneous or successive presentation of the target pairs, provided evidence highlighting that the effect we observed does not depend on the availability of an overlapping retrieval cue—a first in the memory/psycholinguistics literature. All these suggest that not only is a context-specific episodic representation automatically generated during naturalistic reading, but it may also be automatically retrieved to shape lexical interpretation across a range of conditions (e.g., outcome tasks of different nature). Overall, our findings support a role of episodic memory in language comprehension, and we argued that (one of) its key contributions is by endowing the mental lexicon with the adaptability to deal with the bursty and dynamic nature of language.

CRediT authorship contribution statement

Matthew H.C. Mak: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lewis V. Ball:** Writing – review & editing, Resources, Project administration, Investigation. **Alice O'Hagan:** Methodology, Data curation, Conceptualization. **Catherine R. Walsh:** Writing – review & editing, Resources, Methodology. **M. Gareth Gaskell:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Acknowledgements

This research was supported by an Economic and Social Research Council grant (ES/T008571/1) awarded to Gareth Gaskell and Jennifer Rodd. Findings from this paper were presented at an Experimental Psychology Society Meeting in July 2024, at the Asia Virtual Psycholinguistics Forum in August 2024. Experiment 1 A was conducted as the third author's final-year BSc project.

Appendix A

Concrete words in Experiments 1 A and 2 A.

Pair	Word 1	Word 2	Relatedness	Counterbalanced Version
1	road	truck	Related	1
2	hen	cage	Related	1
3	graffiti	poster	Related	1
4	baby	girl	Related	1

(*continued on next page*)

(*continued*)

Pair	Word 1	Word 2	Relatedness	Counterbalanced Version
5	tape	staple	Related	1
6	hoover	mower	Related	1
7	oven	sink	Related	1
8	lettuce	tomato	Related	1
9	needle	pin	Related	1
10	stocking	mittens	Related	1
11	flag	emblem	Related	1
12	coin	dollar	Related	1
13	pot	plant	Related	1
14	soccer	tennis	Related	1
15	finger	claw	Related	1
16	bee	eagle (Exp 1 A), fly (Exp 2 A)	Related	1
17	outlet	keyhole	Related	1
18	exam	lecturer	Related	1
19	apple	lime	Related	1
20	tumbler	chalice	Related	1
21	binder	clipboard	Related	1
22	panda	mouse	Related	1
23	colander	spatula	Related	1
24	bell	tower	Related	1
25	badge	medal	Related	1
26	pillow	quilt	Related	1
27	seaweed	fish	Related	1
28	egg	pancake	Related	1
29	prisoner (Exp 1 A), school (Exp 2 A)	pupil	Related	1
30	mascot	idol	Related	1
31	limousine	debate	Unrelated	1
32	candle	helmet	Unrelated	1
33	wheel	wallpaper	Unrelated	1
34	necklace	fence	Unrelated	1
35	pickle	star	Unrelated	1
36	clown	duck	Unrelated	1
37	drawer	bone	Unrelated	1
38	thermometer	shed	Unrelated	1
39	helicopter	retirement	Unrelated	1
40	feet	pinecone	Unrelated	1
41	baseball	refrigerator	Unrelated	1
42	harbour	popcorn	Unrelated	1
43	mango	marker	Unrelated	1
44	boldness	pumpkin	Unrelated	1
45	rat	antidepressant	Unrelated	1
46	misconception	mammal	Unrelated	1
47	emperor	apron	Unrelated	1
48	vacuum	ocean	Unrelated	1
49	genetics	squirrel	Unrelated	1
50	towel	award	Unrelated	1
51	chimney	planet	Unrelated	1
52	parcel	scam	Unrelated	1
53	fish	windmill	Unrelated	1
54	bar	hippo	Unrelated	1
55	contribution	goggles	Unrelated	1
56	glass	rampage	Unrelated	1
57	tube	trunk	Unrelated	1
58	sword	microwave	Unrelated	1
59	banana	elevator	Unrelated	1
60	eraser	countryside	Unrelated	1
1	marathon	tournament	Related	2
2	zoo	aquarium	Related	2
3	strawberry	raspberry	Related	2
4	match	lighter	Related	2
5	stairs	escalator	Related	2
6	burger	pork	Related	2
7	ring	bracelet	Related	2
8	moon	star	Related	2
9	soda	sandwich	Related	2
10	brush	comb	Related	2
11	warrant	bill	Related	2
12	laundry	dishes	Related	2
13	dormitory	bedroom	Related	2
14	shop	market	Related	2
15	cloud	fog	Related	2
16	pram	trolley	Related	2
17	grave	crypt	Related	2
18	trail	forest	Related	2
19	sandbox	swing	Related	2
20	grill	furnace	Related	2

(*continued on next page*)

(*continued*)

Pair	Word 1	Word 2	Relatedness	Counterbalanced Version
21	paper	pen	Related	2
22	muzzle	mask	Related	2
23	car	mechanic	Related	2
24	butterfly	centipede	Related	2
25	whistle	referee	Related	2
26	dentist	teeth	Related	2
27	grass	lily	Related	2
28	envelope	package	Related	2
29	juice	liquor	Related	2
30	turnip	salad	Related	2
31	guitar	atlas	Unrelated	2
32	elephant	kitchen	Unrelated	2
33	voucher	lizard	Unrelated	2
34	elevator	barn	Unrelated	2
35	domino	mop	Unrelated	2
36	microphone	washcloth	Unrelated	2
37	owl	perfume	Unrelated	2
38	lipstick	giraffe	Unrelated	2
39	zipper	melon	Unrelated	2
40	clergyman	bear	Unrelated	2
41	cowboy	hairdryer	Unrelated	2
42	acorn	sofa	Unrelated	2
43	deodorant	wrench	Unrelated	2
44	goose	fertility	Unrelated	2
45	needle	harmonica	Unrelated	2
46	holiday	curtain	Unrelated	2
47	scooter	meat	Unrelated	2
48	marble	snail	Unrelated	2
49	sandal	crow	Unrelated	2
50	banner	toothpick	Unrelated	2
51	astronaut	scarf	Unrelated	2
52	drone	cigarette	Unrelated	2
53	farmer	episode	Unrelated	2
54	train	sock	Unrelated	2
55	clock	rhino	Unrelated	2
56	award	marmalade	Unrelated	2
57	puddle	tweezers	Unrelated	2
58	waiter	trampoline	Unrelated	2
59	stamp	toilet	Unrelated	2
60	shell	notebook	Unrelated	2

Abstract words in experiments 1B, 2B, 3, and 4.

Pair	Word 1	Word 2	Relatedness	Counterbalanced Version
1	oxygen	air	Related	1
2	torment	abuse	Related	1
3	mistake	blunder	Related	1
4	ache	symptom	Related	1
5	system	mechanism	Related	1
6	worry	anguish	Related	1
7	knowledge	intellect	Related	1
8	culture	society	Related	1
9	cure	immunity	Related	1
10	behavior	discipline	Related	1
11	pitch	tone	Related	1
12	quietness	whisper	Related	1
13	argument	quarrel	Related	1
14	toothache	doom	Related	1
15	oath	loyalty	Related	1
16	traitor	treason	Related	1
17	sense	taste	Related	1
18	idea	theory	Related	1
19	spirit	afterlife	Related	1
20	deception	bribe	Related	1
21	salary	promotion	Related	1
22	franchise	brand	Related	1
23	idol	inspiration	Related	1
24	anthem	verse	Related	1
25	measure	length	Related	1
26	socialist	democracy	Related	1
27	outbreak	epidemic	Related	1
28	sorrow	grief	Related	1
29	freedom	liberty	Related	1

(*continued on next page*)

(*continued*)

30	economy	pension	Related	1
31	trend	prosecution	Unrelated	1
32	incentive	progress	Unrelated	1
33	fatigue	miracle	Unrelated	1
34	analyst	frustration	Unrelated	1
35	expert	boredom	Unrelated	1
36	loophole	status	Unrelated	1
37	bliss	botany	Unrelated	1
38	rank	scapegoat	Unrelated	1
39	mankind	habit	Unrelated	1
40	amateur	altitude	Unrelated	1
41	emphasis	recognition	Unrelated	1
42	traction	glamour	Unrelated	1
43	heredity	nutrient	Unrelated	1
44	truce	weekend	Unrelated	1
45	suppression	caution	Unrelated	1
46	menace	infinity	Unrelated	1
47	scope	paradox	Unrelated	1
48	drama	narcotic	Unrelated	1
49	opinion	diabetes	Unrelated	1
50	relay	approval	Unrelated	1
51	curse	toxin	Unrelated	1
52	morality	accent	Unrelated	1
53	optimist	upgrade	Unrelated	1
54	fraud	mood	Unrelated	1
55	pause	sector	Unrelated	1
56	inkling	delirium	Unrelated	1
57	aptitude	glimpse	Unrelated	1
58	heir	deceit	Unrelated	1
59	scandal	tribute	Unrelated	1
60	enterprise	panic	Unrelated	1
1	zoology	research	Related	2
2	tempo	velocity	Related	2
3	psychology	mind	Related	2
4	setback	adversity	Related	2
5	nuisance	annoyance	Related	2
6	shock	trauma	Related	2
7	conclusion	hypothesis	Related	2
8	rhyme	vowel	Related	2
9	memory	recall	Related	2
10	climate	atmosphere	Related	2
11	worship	religion	Related	2
12	entrepreneur	investment	Related	2
13	charlatan	error	Related	2
14	power	responsibility	Related	2
15	location	sonar	Related	2
16	quantum	atom	Related	2
17	news	story	Related	2
18	forum	seminar	Related	2
19	force	gravity	Related	2
20	joke	comedy	Related	2
21	coward	humiliation	Related	2
22	legend	myth	Related	2
23	geology	calcium	Related	2
24	law	justice	Related	2
25	genius	wisdom	Related	2
26	faith	soul	Related	2
27	fact	fiction	Related	2
28	management	hierarchy	Related	2
29	genocide	onslaught	Related	2
30	grammar	language	Related	2
31	dynasty	jeopardy	Unrelated	2
32	mystery	quantity	Unrelated	2
33	replacement	criticism	Unrelated	2
34	heroism	history	Unrelated	2
35	agility	age	Unrelated	2
36	excuse	league	Unrelated	2
37	honesty	predicament	Unrelated	2
38	prophet	violation	Unrelated	2
39	privacy	export	Unrelated	2
40	comment	insecurity	Unrelated	2
41	generation	practice	Unrelated	2
42	dose	empathy	Unrelated	2
43	abdication	glory	Unrelated	2
44	edition	consent	Unrelated	2
45	failure	trance	Unrelated	2
46	weakness	formation	Unrelated	2

(*continued on next page*)

(*continued*)

47	disposition	hesitation	Unrelated	2
48	ego	passion	Unrelated	2
49	comparison	blackmail	Unrelated	2
50	taboo	sanity	Unrelated	2
51	bluff	warrant	Unrelated	2
52	duty	voice	Unrelated	2
53	definition	curiosity	Unrelated	2
54	question	patent	Unrelated	2
55	scheme	remedy	Unrelated	2
56	satire	divorce	Unrelated	2
57	fusion	zero	Unrelated	2
58	parole	personality	Unrelated	2
59	underworld	disdain	Unrelated	2
60	elite	apathy	Unrelated	2

Appendix B

Table B1
Outputs from the exploratory mixed-effect model examining the effects of Session (1 vs. 2) and types of pair (Unrelated target vs. implicit control) on the estimated relatedness values. Note that the beta and SE were multiplied by 1000 for ease of interpretation.

Fixed effects	*B*	*SE*	*t*	*p*
Intercept	988.10	0.154	6394.218	<0.001*
Session	−0.142	0.079	−1.800	0.072
Type (Unrelated target vs. Implicit control)	0.090	0.074	1.292	0.201
Session x Type	−0.213	0.055	−3.860	<0.001*

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106086.

Data availability

We have made public our data/scripts/materials on Open Science Framework (https://osf.io/ca75m/).

References

Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K. H. T. (2019). Sleep reduces the testing effect-But not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 45*(2), 272–287. https://doi.org/10.1037/xlm0000576

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Ball, L. V., Mak, M. H. C., Ryskin, R., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2025). The contribution of learning and memory processes to verb-specific syntactic processing. *Journal of Memory and Language, 141*(104), 595. https://doi.org/10.1016/j.jml.2024.104595

Blank, A. (1999). Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. In A. Blank, & P. Koch (Eds.), *Historical semantics and cognition* (pp. 61–90). Berlin/New York: Mouton de Gruyter.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition, 2*(1). https://doi.org/10.5334/joc.72

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Carroll, M., & Kirsner, K. (1982). Context and repetition effects in lexical decision and recognition memory. *Journal of Verbal Learning and Verbal Behavior, 21*(1), 55–69. https://doi.org/10.1016/S0022-5371(82)90445-5

Covington, N. V., & Duff, M. C. (2025). Hippocampus supports long-term maintenance of language representations: Evidence of impaired collocation knowledge in amnesia. *Cortex, 182*, 71–86.

Cree, G. S., & Armstrong, B. C. (2012). Computational models of semantic memory. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 259–282). Cambridge: Cambridge University Press.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64*(2), 119–132. https://doi.org/10.1016/j.jml.2010.10.001

Curtis, A. J., Mak, M. H. C., Chen, S., Rodd, J. M., & Gaskell, M. G. (2022). Word-meaning priming extends beyond homonyms. *Cognition, 226*, 105–175. https://doi.org/10.1016/j.cognition.2022.105175

Cychosz, M., Romeo, R. R., Edwards, J., & Newman, R. S. (2024, February 2). *Bursty, irregular speech input to children predicts vocabulary size*. https://doi.org/10.31234/osf.io/6x3d7

Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 581–591. https://doi.org/10.1037/0278-7393.16.4.581

Davis, C. P. (2023). Emergence of Covid-19 as a Novel concept shifts existing semantic spaces. *Cognitive Science, 47*(1), Article e13237. https://doi.org/10.1111/cogs.13237

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioral evidence. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*, 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dong, Y., Mak, M. H., Hepach, R., & Nation, K. (2024). Learning new words via reading: The influence of emotional narrative context on learning novel adjectives. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218241308221

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience, 6*, 69. https://doi.org/10.3389/fnhum.2012.00069

Duff, M. C., Covington, N. V., Hilverman, C., & Cohen, N. J. (2020). Semantic memory and the Hippocampus: Revisiting, reaffirming, and extending the reach of their critical relationship. *Frontiers in Human Neuroscience, 13*(January), 1–17. https://doi.org/10.3389/fnhum.2019.00471

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science, 22*(2), 243–252. https://doi.org/10.1177/09567976103937

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition, 182*, 109–126. https://doi.org/10.1016/j.cognition.2018.09.007

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and readers generalize their experience with word meanings across modalities. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 44*(10), 1533–1561. https://doi.org/10.1037/xlm0000532

Gilbert, R. A., & Rodd, J. M. (2022). Dominance norms and data for spoken ambiguous words in British English. *Journal of Cognition, 5*(1), 1–14. https://doi.org/10.5334/joc.194

Giovanello, K., Schnyer, D. M., & Verfaellie, M. (2004). A critical role for the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus, 14*(1), 5–8.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251. https://doi.org/10.1037//0033-295x.105.2.251

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501–518. https://doi.org/10.1037/0278-7393.11.3.501

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on latent semantic analysis. *Behavior Research Methods, 47*(4), 930–944. https://doi.org/10.3758/s13428-014-0529-0

Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus, 17*(9), 873–889. https://doi.org/10.1002/hipo.20319

Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition, 143*, 87–92. https://doi.org/10.1016/j.cognition.2015.06.009

Hulme, R. C., Begum, A., Nation, K., & Rodd, J. M. (2023). Diversity of narrative context disrupts the early stage of learning the meanings of novel words. *Psychonomic Bulletin & Review, 30*(6), 2338–2350. https://doi.org/10.3758/s13423-023-02316-z

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and toward logit mixed models. *Journal of Memory and Language, 59*, 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review, 23*, 1214–1220. https://doi.org/10.3758/s13423-015-0980-7

Jones, L. L., & Estes, Z. (2012). Lexical priming: Associative, semantic, and thematic influences on word recognition. In *, Vol. 2. Visual Word Recognition* (pp. 44–72). Psychology Press. https://doi.org/10.4324/9780203106976-11.

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. In *, 67. Psychology of learning and motivation* (pp. 239–283). Academic Press. https://doi.org/10.1016/bs.plm.2017.03.008.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114*(1), 1.

Kassambara, A. (2023). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R Package Version 0.7.2 https://CRAN.R-project.org/package=rstatix.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, New Jersey: Lawrence Erlbaum.

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology, 3*, 245. https://doi.org/10.3389/fpsyg.2012.00245

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 249. https://doi.org/10.3389/neuro.06.004.2008

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review, 28*(1), 40–80.

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

Lenth, R. (2023). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.5 https://CRAN.R-project.org/package=emmeans.

Mak, M. H. C. (2019). Why and how the co-occurring familiar object matters in Fast Mapping (FM)? Insights from computational models. *Cognitive Neuroscience, 10*(4), 229–231. https://doi.org/10.1080/17588928.2019.1593121

Mak, M. H. C. (2024). Data from "a registered report testing the effect of sleep on DRM false memory: Greater lure and veridical recall but fewer intrusions after sleep". *Journal of Open Psychology Data, 12*(1), 6. https://doi.org/10.5334/jopd.98

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023). Episodic memory and sleep are involved in the maintenance of context-specific lexical information. *Journal of Experimental Psychology: General, 152*(11), 3087–3115. https://doi.org/10.1037/xge0001435

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2024). Recall and recognition of discourse memory across sleep and wake. *Journal of Memory and Language, 138* (104), 536. https://doi.org/10.1016/j.jml.2024.104536

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021a). Lexical connectivity effects in immediate serial recall of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(12), 1971–1997. https://doi.org/10.1037/xlm0001089

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021b). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language, 118*(October 2020), Article 104203. https://doi.org/10.1016/j.jml.2020.104203

Mak, M. H. C., & Nation, K. (2024). *A review of how word learning during reading is influenced by the linguistic context: From contextual constraint to contextual diversity*.

Mak, M. H. C., O'Hagan, A., Horner, A. J., & Gaskell, M. G. (2023). A registered report testing the effect of sleep on Deese-Roediger-McDermott false memory: greater lure and veridical recall but fewer intrusions after sleep. *Royal Society Open Science, 10* (12). https://doi.org/10.1098/rsos.220595

Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired-associate learning. *Psychonomic Bulletin & Review, 27*, 1059–1069. https://doi.org/10.3758/s13423-020-01773-0

Mayes, A. R., Holdstock, J. S., Isaac, C. L., Montaldi, D., Grigor, J., Gummer, A., … Norman, K. A. (2004). Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus, 14*(6), 763–784.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 463–480. https://doi.org/10.1016/s0022-5371(79)90255-x

McKoon, G., & Ratcliff, R. (1980). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior, 19* (4), 369–386. https://doi.org/10.1016/s0022-5371(80)90267-4

McKoon, G., & Ratcliff, R. (1986). Automatic activation of episodic information in a semantic memory task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(1), 108–115. https://doi.org/10.1037/0278-7393.12.1.108

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin, 143*(5), 499.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 863. https://doi.org/10.1037//0278-7393.21.4.863

Myslín, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition, 147*, 29–56. https://doi.org/10.1016/j.cognition.2015.10.021

Natale, V., & Lorenzetti, R. (1997). Influences of morningness-eveningness and time of day on narrative comprehension. *Personality and Individual Differences, 23*(4), 685–690. https://doi.org/10.1016/s0191-8869(97)00059-7

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning, 2*(1). https://doi.org/10.1038/s41539-017-0004-7, 0-1.

Neely, J. H., & Durgunoğlu, A. Y. (1985). Dissociative episodic and semantic priming effects in episodic recognition and lexical decision tasks. *Journal of Memory and Language, 24*(4), 466–489. https://doi.org/10.1016/0749-596x(85)90040-3

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review, 120*(2), 356. https://doi.org/10.1037/a0032020

Oakhill, J. (1986). Effects of time of day on the integration of information in text. *British Journal of Psychology, 77*(4), 481–488. https://doi.org/10.1111/j.2044-8295.1986.tb02212.x

Parker, A. J., Taylor, J. S. H., & Rodd, J. M. (2023, August 3). *Readers use recent experiences with word meanings to support the processing of lexical ambiguity: Evidence from eye movements*. https://doi.org/10.31234/osf.io/b49vk

Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PLoS One, 7* (3), Article e33079. https://doi.org/10.1371/journal.pone.0033079

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning Research, 12*, 2825–2830.

Prior, A., & Bentin, S. (2003). Incidental formation of episodic associations: The importance of sentential context. *Memory & Cognition, 31*, 306–316. https://doi.org/10.3758/bf03194389

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

Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 210. https://doi.org/10.1037/a0017258

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17*(4), 403–417. https://doi.org/10.1016/S0022-5371(78)90238-4

Ratcliff, R., & McKoon, G. (1981a). Does activation really spread? *Psychological Review, 88*(5), 454–462. https://doi.org/10.1037/0033-295X.88.5.454

Ratcliff, R., & McKoon, G. (1981b). Automatic and strategic priming in recognition. *Journal of Verbal Learning and Verbal Behavior, 20*(2), 204–215.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*(3), 191–201. https://doi.org/10.3758/BF03197692

Reggin, L. D., Muraki, E. J., & Pexman, P. M. (2021). Development of abstract word knowledge. *Frontiers in Psychology, 12*(686), 478. https://doi.org/10.3389/fpsyg.2021.686478

Richie, R., White, B., Bhatia, S., & Hout, M. C. (2020). The spatial arrangement method of measuring similarity can capture high-dimensional semantic structures. *Behavior Research Methods, 52*, 1906–1928. https://doi.org/10.3758/s13428-020-01362-y

Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science, 15*(2), 411–427. https://doi.org/10.1177/1745691619885860

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language, 87*, 16–37. https://doi.org/10.1016/j.jml.2015.10.006

Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68* (2), 180–198. https://doi.org/10.1016/j.jml.2012.08.002

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 432–444. https://doi.org/10.1037/0278-7393.12.3.432

Schacter, D. L., & Graf, P. (1986b). Preserved learning in amnesic patients: Perspectives from research on direct priming. *Journal of Clinical and Experimental Neuropsychology, 6*, 727–774. https://doi.org/10.1080/01688638608405192

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology, 32*(3), 219–250. https://doi.org/10.1006/cogp.1997.0652

Shimamura, A. P., & Squire, L. R. (1989). Impaired priming of new associations in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 721. https://doi.org/10.1037//0278-7393.15.4.721

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2*, 339–363. https://doi.org/10.3758/bf03210972

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10*(1), 49–62. https://doi.org/10.1016/j.smrv.2005.05.002

Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron, 81*(1), 12–34. https://doi.org/10.1016/j.neuron.2013.12.025

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics, 17*(6), 520–525. https://doi.org/10.1093/bioinformatics/17.6.520

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2*(3), 67–70. https://doi.org/10.1111/1467-8721.ep10770899

Tulving, E., Hayman, C. A., & Macdonald, C. A. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 595. https://doi.org/10.1037//0278-7393.17.4.595

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition, 22*(1), 111–126.

Van Heuven, Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for. *British English. Quarterly journal of experimental psychology, 67*(6), 1176–1190.

Voeten, C. C. (2023). *Buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R Package Version 2.11.

Walsh, C. R., & Rissman, J. (2023). Behavioral representational similarity analysis reveals how episodic learning is influenced by and reshapes semantic memory. *Nature Communications, 14*(1), 7548. https://doi.org/10.1038/s41467-023-42,770-w

Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. *arXiv*, 1–22. preprint arXiv:1308.5499.