UNIVERSITY of York

This is a repository copy of *Word-meaning priming extends beyond homonyms*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/187682/

Version: Published Version

Article:
Curtis, Adam J, Mak, Matthew H C, Chen, Shuang et al. (2 more authors) (2022) Word-meaning priming extends beyond homonyms. Cognition. 105175. ISSN 0010-0277

https://doi.org/10.1016/j.cognition.2022.105175

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Cognition 226 (2022) 105175

Contents lists available at ScienceDirect

Cognition

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

Word-meaning priming extends beyond homonyms

Adam J. Curtis [a], Matthew H.C. Mak [a], Shuang Chen [b], Jennifer M. Rodd [c], M. Gareth Gaskell [a,*]

[a] Department of Psychology, University of York, UK
[b] Department of Psychology, Zhejiang Normal University, China
[c] Division of Psychology and Language Sciences, University College London, UK

ARTICLE INFO

Keywords:
Lexical ambiguity
Priming
Language comprehension
Memory
Plasticity

ABSTRACT

When a homonym (e.g., *bark*) is encountered in a sentential context that biases its interpretation towards a less frequent meaning, subsequent interpretations of the word are more likely to favour that subordinate meaning. Such word-meaning priming effects have been shown to be maintained via sleep-related consolidation, leading some to suggest that declarative memory systems play a crucial role in language comprehension, providing a relatively enduring contextually bound memory trace for the ambiguous word. By this account, word-meaning priming effects should be observable for all words, not just homonyms. In three experiments, participants were exposed to non-homonym targets (e.g., "balloon") in sentences that biased interpretation towards a specific aspect of the word's meaning (e.g., *balloon-helium* vs. *balloon-float*). After a ~ 10–30 min delay, the targets were presented in relatedness judgement and associate production tasks to assess whether the sentential contexts enhanced access to the primed aspect of the word's meaning. The results reveal that word-meaning priming effects do extend to non-homonyms. Indeed, there was also some evidence of a more generalised priming that did not rely on prior presentation of the non-homonym itself. We argue that context-specific interpretations of words are maintained during recognition in order to facilitate comprehension over longer periods.

1. Introduction

Research over the last 15–20 years has demonstrated the remarkable flexibility of human language systems in adapting to new circumstances, such as a stranger's accent, or an unfamiliar word (e.g., Gaskell & Dumay, 2003; Hanulíková, van Alphen, van Goch, & Weber, 2012; Norris, McQueen, & Cutler, 2003; Shtyrov, Nikulin, & Pulvermüller, 2010). This research has inevitably led to a stronger focus on the way in which memory systems support language processing by providing necessary plasticity and longevity. Mnemonic processes such as rapid encoding and consolidation have been implicated in the processing of novel linguistic units (Dumay & Gaskell, 2007; Wang et al., 2017). However, more recent research suggests that the role of these mnemonic mechanisms in language processing may not be limited to cases where specific new linguistic units are acquired. Instead, it may be that there is a broader role for these processes in the comprehension of highly familiar language.

Language comprehension involves the rapid retrieval of a word's contextually appropriate meaning. For homonyms, we have to retrieve one meaning from multiple semantically unrelated alternatives. For example, when reading a sentence such as "He threw the bag into the trunk of the car", the correct interpretation of the word "trunk" must be selected from several other contextually inappropriate meanings (e.g., tree trunk, elephant trunk). Over the lifespan, we likely accumulate substantial lexical knowledge regarding the relative frequencies of the alternative meanings of such words. For example, participants tend to be biased towards a word's more frequent meaning in tasks such as word association (Twilley, Dixon, Taylor, & Clark, 1994). However, these preferences for the alternative meanings of ambiguous words are not stable; rather, they are adjusted based on new usage information. For example, Rodd, Lopez Cutrin, Kirsch, Millar, and Davis (2013) demonstrated that word-meaning preferences for lexically ambiguous words could vary depending on recent usage. Rodd et al. (2013) exposed participants to spoken homophones that were embedded in sentential contexts favouring the subordinate meaning. For example, hearing "Bark is found on the trunk of many trees" would bias the interpretation of "bark" towards its subordinate meaning "the tough exterior of a woody plant" and away from its dominant meaning "the sound made by a dog". After a 20-min delay, participants were given an associate production task where a cue word (e.g., "bark") was shown and participants

* Corresponding author at: Department of Psychology, University of York, Heslington, York YO10 5DD, UK
E-mail address: gareth.gaskell@york.ac.uk (M.G. Gaskell).

https://doi.org/10.1016/j.cognition.2022.105175
Received 4 February 2022; Received in revised form 12 May 2022; Accepted 14 May 2022
Available online 27 May 2022
0010-0277/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

had to generate an associated word. Relative to an unprimed baseline, presentation of the biasing sentential contexts resulted in an increase in the proportion of associate responses that were semantically congruent with a word's subordinate meaning. This finding, referred to as *word-meaning priming*, demonstrates that a single encounter with a lexically ambiguous word in a disambiguating context may be enough to alter its lexical-semantic representation, at least temporarily. Thus, any recent experience with a word appears to have important consequences for how the word is subsequently interpreted. Subsequent studies addressed the reliability, generality and longevity of this priming effect (Gilbert, Davis, Gaskell, & Rodd, 2018, 2021; Rodd et al., 2016).

One explanation of this word-meaning priming effect is that it reflects the immediate updating of long-term, lexical form-meaning representations (Gilbert et al., 2018). Under this model, the form of a homonym is associated with its different meanings, with more dominant meanings having stronger weightings. Exposure to a word in a disambiguating context strengthens the association between the word-form and a specific meaning. Thus, the alteration of weights between form and meaning results in an increased likelihood of a primed meaning being accessed at test. Although plausible, this *immediate alteration* account has been questioned by more recent research investigating the longer-term retention of such information. Building on the paradigm from Rodd et al. (2013), Gaskell, Cairney, and Rodd (2019) tested participants both shortly after encoding (8 mins) and after a 2- or 12-h delay spent either awake or asleep. When tested after a short delay of 8 min, priming was present across participant groups. However, after 2 or 12 h, evidence for priming was only present for participants in the sleep group. Thus, the longer-term retention of knowledge regarding the statistical frequencies of a word's meanings, and the ability of such information to bias subsequent interpretation, appear to be influenced by a period of sleep-associated consolidation.

The proposal that word-meaning priming is the consequence of immediate adjustments of long-term lexical-semantic representations does not provide an easy explanation for the consolidation effects observed in Gaskell et al. (2019). Gaskell et al. instead proposed that such findings may be explicable by assuming that rapid encoding and subsequent systems consolidation contribute to the on-line processing and longer-term retention of lexical information (Duff & Brown-Schmidt, 2012, 2017). When a homonym is encountered in a disambiguating context, rather than long-term representations being immediately altered, a new but more temporary memory for the comprehension episode is formed in which the word is bound to its sentential context. The binding of a word to its context could contribute to, or support, an event representation or situation model of the discourse (Graesser, Millis, & Zwaan, 1997; van Dijk & Kintsch, 1983; Zacks, Speer, & Reynolds, 2009; Zacks, Speer, Swallow, Braver, & Reynolds, 2007), facilitating wider comprehension of linguistic material across sentences or utterances. As part of this process, a contextually bound representation of a word would provide an additional source of information alongside long-term lexical knowledge that could influence the subsequent interpretation of that same word. Thus, the formation of a temporary event representation could account for word-meaning priming effects in the minutes or possibly hours following exposure (Rodd et al., 2016). On the other hand, the long-term retention of this new knowledge (i.e., to inform future interpretations of the word) may be reliant on a period of sleep-associated consolidation, which would allow it to be integrated into long-term memory. If such a consolidation period is not provided soon enough, the representation may be subject to hippocampal trace decay, reducing the likelihood of any priming effects to emerge. However, if integrated into long-term stores over a period of sleep, this knowledge would be less susceptible to loss and thus would be able to influence comprehension over longer delay periods.

Gaskell et al. (2019) referred to this alternative account as a contextual binding account, but here we describe it as an *episodic context* account, to avoid confusion with Yonelinas, Ranganath, Ekstrom, and Wiltgen (2019). The account implies a central role for declarative memory systems in the comprehension of familiar language. Importantly, in linking episodic memory to discourse comprehension, the scope of the account is rather different from the immediate alteration account. According to an episodic context account, all sentences successfully comprehended by a reader or listener (cf. Gilbert et al., 2021) should lead to the generation of a contextually bound representation in memory to aid future comprehension. If this is the case, then word-meaning priming should not be restricted to lexically ambiguous words; it should be possible to find priming for any word where the sentence context modifies or contributes to the meaning of the word in some way. In view of this, the present study investigates whether word-meaning priming effects can be observed for non-homonyms (also known as lexically unambiguous words; e.g., Rayner & Duffy, 1986). In three experiments, we used sentences that highlighted particular aspects of a non-homonym's meaning (e.g., particular semantic features, senses or states). Previous eye-tracking and neuroimaging studies (Altmann & Kamide, 1999; Hindy, Altmann, Kalenik, & Thompson-Schill, 2012) provided evidence that the comprehension of such words in sentence context may indeed be linked to the differential activation of certain features or aspects of a word's meaning. For example, when a word such as “balloon” is encountered in isolation, it may activate associated concepts such as *helium* and *float* equally well. However, if comprehension involves the binding of a word to a given context, encountering “balloon” within a specific sentential context would likely result in the generation of an event model that is more congruent with a specific aspect of the word's meaning. For example, reading “The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky” would likely generate an event model more compatible with the concept of *helium* than with *float*. The subsequent interpretation of a new instance of “balloon” in isolation may then be biased towards the primed context (e.g., *helium*). If priming effects can be observed for non-homonyms via the manipulation of sentential context, this would indicate that word-meaning priming effects are not dependent on words having multiple unrelated meanings and would provide further support for an episodic context account. The present study investigates word-meaning priming for non-homonyms within the same day over similar periods for which homonyms have shown word-meaning priming (~10–20 min).

All experiments used a set of non-homonymic target words (e.g., “balloon”) paired with probe words that were related to a particular aspect of the target word's meaning (e.g., *helium*). For Experiment 1, during an exposure phase, participants read the target words embedded in sentential contexts that biased their interpretation either towards that specific aspect of their meaning (probe-consistent condition) or away from it (probe-inconsistent condition). For some target-probe pairs, no sentence was presented at exposure (unprimed condition). Priming was indexed using two tasks, administered straight after the exposure phase: (i) speeded semantic relatedness decisions, in which participants judged whether the target word was semantically related to the probe word (e.g., balloon-helium), and (ii) associate production, in which participants were presented with the targets as cues in isolation and were asked to generate an associated word (e.g., balloon-). If the prior exposure to the target word in the probe-consistent condition resulted in an enduring increase in the availability of its contextually relevant semantic features then at test this is predicted to result in faster and more accurate semantic relatedness judgements in the probe-consistent condition, compared to both the probe-inconsistent condition and the unprimed baseline. Similarly in the associate production task, participants were predicted to generate more responses relating to the primed aspect of the words meaning in the probe-consistent condition than in the other two conditions. Alternatively, if word-meaning priming is specific to homonyms, no effect of exposure context would be predicted.

For Experiment 1, the delay between exposure and test for any particular target word was about 10 min. In Experiment 2, we increased the delay to ~20–30 min and also tested whether priming could be found when the target word was replaced by a semantically similar word

during exposure (e.g., "vessel" for "balloon"). This allowed us to test whether any priming was specifically tied to a specific word-form, or whether it might be a more general semantic priming. Experiment 3 was similar to Experiment 2, but refined the sentence materials so that they related to less accessible aspects of the target word meanings. All experiments were run online.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Seventy-one participants were recruited via Prolific (www.prolific.co). Two participants were excluded based on language background and three participants were excluded for low accuracy on the relatedness judgement task, exceeding a time limit of 75 min, and guessing the experimental manipulation, respectively. The final sample of 66 participants (30 females; $M \pm SD$ age in years = 29.5 ± 6.48) were all native English speakers with normal or corrected-to-normal eyesight and none reported history of any developmental disorders. The target sample size was selected to be in line with the Brysbaert and Stevens (2018) recommendation of ~1600 observations per condition (66 participants × 24 trials per condition = 1584 observations). Participants gave informed consent for the experiment, which was approved by the Research Ethics Committee of the Department of Psychology, University of York.

2.1.2. Materials

The experiment used 72 target words that had a single "entry" in the Wordsmyth Advanced online dictionary (Parks, Ray, & Bland, 1998). This makes them non-homonyms in terms of lexicographers' judgements, which tend to agree well with participant responses (Rodd, Gaskell, & Marslen-Wilson, 2002). Additionally, none of the target words were listed in the University of Alberta homograph norms (Twilley et al., 1994). Although the target words all had a single entry (i.e., a single core meaning), we did not attempt to restrict our selection to words that were listed as having only one sense, but the distinction between (related) senses and (unrelated) meanings is theoretically important (Armstrong & Plaut, 2016), and different senses are not understood to be in competition during word recognition in the way that unrelated meanings are (Rodd, 2020). Of the 72 target words, 52 had more than one sense according to the Wordsmyth database and the average number of senses across all items was 2.63. Target word frequency for these items, indexed by the SUBTLEX-UK database (van Heuven, Mandera, Keuleers, & Brysbaert, 2014), ranged from 0.11–48.26 per million ($M = 13.31$, $SD = 11.69$), with all words selected to be familiar to a typical participant (e.g., the lowest frequency words were eraser, bathtub and quilt). Target word length ranged from 4 to 13 letters ($M = 6.28$, $SD = 1.75$). Each of these target words was paired with a probe word that was semantically related to a specific aspect of the target word's meaning (e.g., balloon-helium). All target-probe pairs were listed as associated in the USF free association norms database (Nelson, McEvoy, & Schreiber, 1998).

As mentioned, each target word was used to construct two sentence contexts. The probe-consistent sentence included the target word embedded within a context that biased its interpretation towards the probe word. For example, reading "The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky", in which the target word is "balloon", may generate an interpretation more compatible with *helium* than with *float*. The probe-inconsistent sentence included the target word embedded within a context that biased its interpretation away from the probe word and towards a different associate. For example, reading "People looked up at the hot air balloon drifting across the sky", would likely generate an interpretation more compatible with *float* than with *helium*. Note that although the two context sentences were intended to relate to different aspects of meaning of the target words, they generally (65 out of 72) related to the same sense in the Wordsmyth dictionary. In total, there were 144 test sentences (72 pairs) with a mean word count of 14.2 ($SD = 3.6$). For the purposes of counterbalancing, the items were randomly divided into three lists. The lists were matched on mean target word length $F(1, 70) = 0.24$, $p = .625$, mean target word frequency, $F(1, 70) = 1.13$, $p = .292$, and mean sentence word count, $F(1, 142) = 2.41$, $p = .123$. These lists were then used to construct six counterbalanced sets of stimuli with each item appearing in each condition equally often across participants.

In order to assess attention during the exposure phase and to ensure participants were fully comprehending the sentences, each sentence was also paired with a semantically related thematic word. For example, for the target-probe pair "balloon-helium" the thematic word was "inflate". We also created 32 filler sentences ($M \pm SD$ word count = 14.28 ± 2.83) that had semantically unrelated thematic words (See Appendix A).

2.1.3. Design

The study used a within-subjects design. For each participant, a target word could fall into one of three conditions: probe-consistent, probe-inconsistent or unprimed. For the probe-consistent condition, priming sentences biased interpretation of the target words towards the probe words. For the probe-inconsistent condition, priming sentences biased interpretation of the target words away from the probe word. The unprimed condition was a control in which target words were not presented at exposure.

The use of these three conditions effectively provided two forms of control to compare with the probe-consistent condition. The unprimed condition is the standard control in a word-meaning priming condition and provides a simple baseline to assess priming (e.g., Rodd et al., 2013). The probe-inconsistent condition assesses the specificity of any priming, which is particularly important for the semantic relatedness test. Feasibly, just the presence of the target word in any context during exposure could lead to the activation of all aspects of that word's meaning, meaning that a later semantic relatedness judgement for that word would be facilitated. However, if the specific sentence context of the word is important during exposure then facilitation of semantic judgement responses should be found for the probe-consistent but not the probe-inconsistent condition. For this reason, it was important that our analysis tested whether the probe-consistent condition differed from both types of control.

2.1.4. Procedure

The experiment was constructed using Gorilla Experiment Builder (https://gorilla.sc/) and was run online. Participants were required to complete the study on either a desktop computer or laptop and were instructed to ensure that they were in a quiet environment where they would not be disturbed for the duration of the study. Recent research (e.g., Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2021) has demonstrated that display durations and measured RTs show a reasonable level of accuracy when experiments are run online. More directly relevant, word-meaning priming experiments involving speeded RTs have proved to be sensitive to the exposure manipulation (e.g., Gilbert et al., 2021), so we were confident that any extra variability associated with environment or equipment would not impact on our research goals. The procedure of Experiment 1 is summarised in Fig. 1.

The experiment began with an exposure phase where participants read sentences containing the target non-homonyms. Immediately after the exposure phase, participants completed two post-tests in a fixed order, namely semantic relatedness judgement and associate production. The ordering of the two tests was important in that the first test would likely act as a further source of bias on the outcome of the second test regardless of the order. That said, the potential for bias from the relatedness judgement task is fixed and present in all three condition (i.e., the same probe is used in all conditions). The same cannot be said of the associate production task, where the biasing effect would depend on the nature of the generated associate, which we expected to vary across

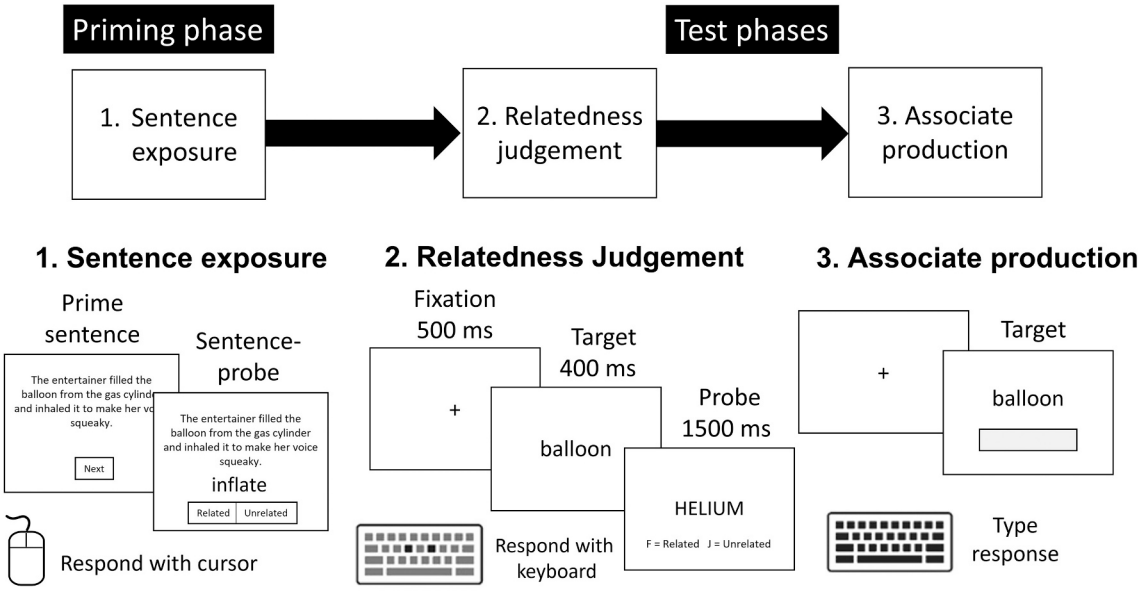

Fig. 1. Schematic of trial progression, Experiment 1.

conditions. Therefore, the chosen ordering ensured that any biasing effect would be equal across conditions and unable to act as a confound in the results of the second task.

2.1.4.1. Exposure phase. Participants read 80 sentences, 48 of which were test sentences (24 probe-consistent, 24 probe-inconsistent), while the remaining 32 were filler sentences. Each trial began with a sentence on the top half of the screen. After reading the sentence, the participants could press a "Next" button with their cursor and a thematic word would appear beneath the sentence. Here, participants judged whether the sentence and the thematic word were semantically related by pressing either a "Related" or an "Unrelated" button with their cursor. Note, for the test trials, the thematic words were always related to the sentences, whereas for the fillers the thematic words were always unrelated to the sentences. This was done to ensure that the test sentences were always processed under similar task conditions. The 80 sentences were split equally across two blocks. Trial order was randomised within blocks; block order was counterbalanced across participants. There were no time limits placed on the exposure phase.

2.1.4.2. Semantic relatedness judgement. This was a speeded task in which participants judged whether two words were semantically related. The task consisted of 72 test trials (in which target-probe pairs were related) and 48 filler trials (in which target-probe pairs were unrelated), giving a total of 120 trials. Note, the target-probe pairs used on the filler trials were not in any way linked to the filler sentences used during the exposure phase. The task was spread across two blocks with target words from Block 1 of the exposure phase appearing in Block 1 of the relatedness-judgement task, and target words from Block 2 of the exposure phase appearing in Block 2 of the relatedness judgement task.

The task began with six practice trials (3 related, 3 unrelated). In these, participants received immediate feedback. However, no feedback was given in the experimental blocks. Each trial began with a 500 ms fixation cross followed by the target word for 400 ms. After a 200 ms blank screen, the probe word appeared. Participants judged whether or not the probe was related to the target by pressing the "F" (=Related) or "J" (=Unrelated) key. Participants were given 1500 ms to respond. If they did not respond, the text "Too Slow!" appeared for 1500 ms.

2.1.4.3. Associate production. This was a free association task in which the participants produced an associate for each of the 72 target words. After a single practice trial, each trial began with a 500 ms fixation cross after which a target word appeared. Participants were instructed to type in the first word that came to mind upon seeing the target word. Participants could proceed to the next trial by pressing the "Return" key on their keyboard. They were not allowed to proceed without giving an answer. The task also included three attention checks in which participants were required to recall the target word from the previous trial.

2.2. Results

2.2.1. Analysis approach

All data were analysed in a mixed-effects environment using the lme4 package (version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015) in R (version 4.1.1; R Core Team, 2021). Generalised linear models (relatedness judgement accuracy and associate production) were conducted using the *glmer* function which provides *p*-values based on an asymptotic Wald's test. Linear models (relatedness judgement RT) were conducted using the *lmer* function. To derive p-values for the linear models we used a second package: *lmerTest*. *lmerTest* provides p-values based on a *t*-test that calculates degrees of freedom using Satterthwaite's method. For all tasks, there is one independent variable, Priming, which was manipulated within-participants and had three levels: probe-consistent, probe-inconsistent, and unprimed. These conditions were compared using treatment contrasts, with the probe-consistent serving as the reference level. This allows us to test (1) whether there was a word-meaning priming effect (probe-consistent vs. unprimed) and (2) whether any word-meaning priming effect was context-specific (probe-consistent vs. probe-inconsistent). As described in the Design section, both these contrasts are particularly important for the semantic relatedness judgement analysis.

For the relatedness judgement task there were two outcome measures: accuracy and response time (RT), with accuracy coded as a binary variable. Trials with a response time greater than 1500 ms were removed from all analyses and only correct responses were included in the RT analysis.

For the associate production task, the dependent variable of interest was whether a generated associate was related to either sentence context (i.e., probe-consistent or probe-inconsistent), coded as a binary variable. This was judged by a group of third-party raters ($n = 12$), recruited via Prolific and Sona Systems. The raters were given the associate words generated by the participants in the main associate production task. In

each trial, an associate word was shown underneath two sentences. For example, the word "pop" (an associate generated in the main experiment) was shown underneath two sentences: One included the target word in the probe-consistent context (e.g., "The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky"), the second included the target word in the probe-inconsistent context (e.g., "People looked up at the hot air balloon drifting across the sky"). The sentences appeared above/below one another on the screen at the same time as the associate word. The position of the probe-consistent sentence (i.e., above or below the probe-inconsistent sentence) was counterbalanced across raters. The raters were given three options: (i) the response word is related to the first sentence (ii) the response word is related to the second sentence, (iii) the response word is equally related to both sentences. In order for a response to be coded as favouring the probe-consistent context it was required to be judged as related to the probe-consistent sentence by at least 7 of the 12 raters.[1]

Analysis of the associate production task, and of accuracy in the relatedness judgement task, used binary logistic mixed effects models. Response times in the relatedness judgement task were analysed using a linear mixed effects model. Model building began by computing a random intercepts-only model with all fixed effects and random intercepts for items and subjects. Random-slopes were then integrated into the model using a forward best-path approach (Barr, Levy, Scheepers, & Tily, 2013). Only random slopes that were justified by the data with an α-level of $p < .2$ were included in the model.

2.2.2. Relatedness judgement

2.2.2.1. Accuracy. Mean accuracy rates in the three conditions are summarised in the left panel of Fig. 2. As predicted, accuracy was significantly greater in the probe-consistent condition ($M = 92.7\%$, $SD = 6\%$) than in the unprimed condition ($M = 89.6\%$, $SD = 7\%$) ($\beta = -0.45$, SE = 0.13, $z = -3.38$, $p < .001$; see Table 1). Likewise, accuracy was greater in the probe-consistent condition ($M = 92.7\%$, $SD = 6\%$) than in the probe-inconsistent condition ($M = 88.8\%$, $SD = 8.1\%$) ($\beta = -0.56$, SE = 0.13, $z = -4.22$, $p < .001$).

2.2.2.2. Response times. Mean RTs in the three conditions are summarised in the right panel of Fig. 2. The residuals of the model violated the assumption of normality. To resolve this, the RTs were subjected to a log transformation. RTs were significantly faster in the probe-consistent condition ($M = 622$ ms, $SD = 81$ ms) when compared with both the unprimed ($M = 683$ ms, $SD = 93$ ms) ($\beta = 1.25$, SE = 0.3, $t = 4.03$, $p < .001$) and probe-inconsistent conditions ($M = 679$ ms, $SD = 87$ ms) ($\beta = 1.19$, SE = 0.3, $t = 3.84$, $p < .001$).

2.2.3. Associate production

For this analysis, a further six participants were excluded, giving a sample size of 60. Two were excluded for failing to finish the task within the experiment time limit and four were excluded as post-test questions revealed they misunderstood the requirements of the task.

The proportion of associate responses that favoured the probe-consistent context in the three conditions are summarised in Fig. 3. As predicted, participants were more likely to produce associates favouring the probe-consistent context in the probe-consistent ($M = 57\%$, $SD = 16.2\%$) than in the unprimed condition ($M = 48.6\%$, $SD = 16.4\%$) ($\beta = -0.43$, SE = 0.09, $z = -5.06$, $p < .001$; see Table 1). Moreover, associates favouring the probe-consistent context were also more common in the probe-consistent ($M = 57\%$, $SD = 16.2\%$) than in the probe-inconsistent condition ($M = 49.2\%$, $SD = 14.3\%$) ($\beta = -0.4$, SE = 0.09, $z = -4.74$, $p < .001$).

2.3. Discussion

Across the two tasks and three dependent variables of Experiment 1, the results were clear. The speeded semantic relatedness judgement task showed that exposure to a sentence context containing a target word made it easier (i.e., faster and more accurate) to determine that the target word was related to a sentence context-consistent probe, compared with an inconsistent context or the unprimed control case. Similarly, when later asked to produce an associate of the target word, the associate was more likely to be related to the biasing sentence context, as compared with the other two conditions.

The comparison with both control conditions is important here. The comparison with the unprimed condition shows that seeing the target word in a probe-consistent context changes the way in which that target word is treated, but it does not identify whether the context itself is important. It could be that the effects are found regardless of the nature of the context. However, the comparison with the probe-inconsistent context shows that the nature of the context is also important, and that this helps to guide the subsequent interpretations. Taken together, these results suggest that it is possible to prime selected aspects of meaning for non-homonyms. That is, exposure to a non-homonym in a particular context leads to changes in the way that the word is subsequently evaluated about 10 min later, biasing its interpretation towards the prior context. This word-meaning priming has been found previously for classical lexically ambiguous words (i.e., words with two or more unrelated meanings such as"bank") (Gaskell et al., 2019; Gilbert et al., 2021; Rodd et al., 2013), but appears to generalise to a wider range of words including those with just one meaning. This result would not be predicted if priming reflects the immediate updating of the relative dominance levels of different unrelated meanings of lexically ambiguous words (Gilbert et al., 2018).

Perhaps rather than adjusting dominance levels of semantically unrelated *meanings* of a word, word-meaning priming here relates to the adjustment of the relative prominence or accessibility of different semantically related *senses* of the target words. As mentioned, most of the words that we used were listed in Wordsmyth as having more than one sense. However, for the majority of the items (62 out of 72) the sentence-context manipulation related to the same sense of the words. Therefore, if word-meaning priming in Experiment 1 was a consequence of enhancing the availability of the primed meaning there should be no difference in performance between the probe-consistent and probe-inconsistent conditions. In fact, when we restricted our analysis to just these items (exploratory analyses available at https://osf.io/yg428/), we found exactly the same pattern of performance. Therefore, it is difficult to see how the data of Experiment 1 can be explained in terms of immediate alteration of dominance levels, regardless of whether they relate to separate meanings or senses of a word.

On the other hand, our results sit comfortably with an episodic context account of word-meaning priming (Gaskell et al., 2019). If exposure to a word in a particular sentential context results in the creation or updating of a contextually bound representation, as proposed by the episodic context account, then the congruence between the context and the probe word could reasonably bias responses at test. Contextual binding provides a natural explanation for the fact that word-meaning priming can be seen both for words with multiple meanings and for words with just a single meaning.

However, it is worth considering the possibility that these effects are a consequence of a more general form of semantic priming rather than anything specific to the target words. For example, given the sentence "The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky", in which the target word is "balloon", it could be the case that semantic cues related to *entertainment*, *filling*, *squeaky*,

[1] The threshold of 7 out of 12 is fairly arbitrary, but ensures that the number of probe-consistent responses is greater than the number of responses for the other two options combined. Tweaking the threshold, or using a continuous measure could conceivably have led to a more sensitive measure of priming, but when we found that this threshold revealed priming effects in Experiment 1 we opted to keep it fixed for the remaining experiments.

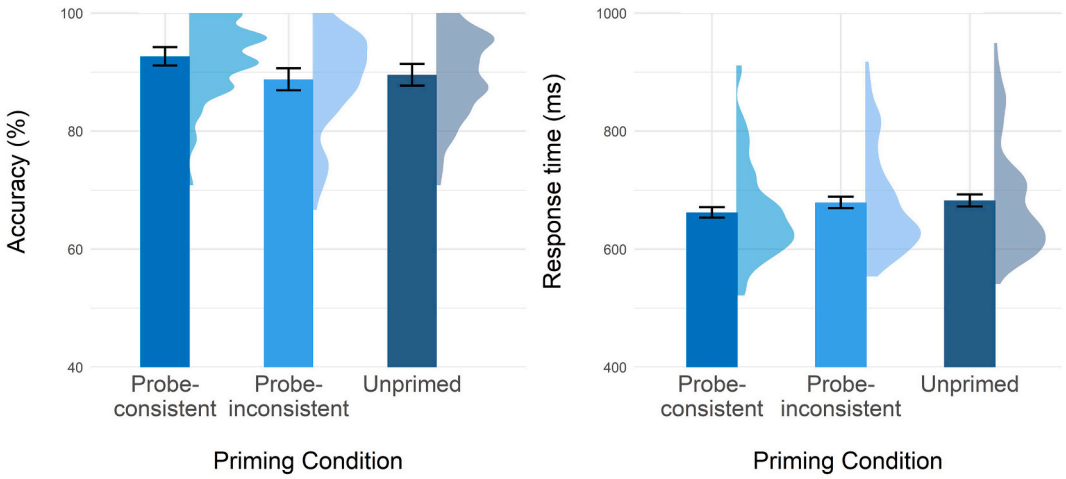

Fig. 2. Accuracy (left) and RTs (right) in the relatedness judgement task, Experiment 1. Error bars represent 95% within-subject confidence intervals (Morey, 2008). Density functions represent the distribution of participant means in each condition.

Table 1
Summary of the mixed-effects models for the two tests in Experiment 1.

Task/ dependent variables		Estimates	Std error	*z* / *t* values	*p*
Relatedness judgement/ Accuracy	Intercept	3.2	0.2	16.37	<0.001*
	Probe-consistent vs. Unprimed	−0.45	0.13	−3.38	<0.001*
	Probe-consistent vs. Probe-inconsistent	−0.56	0.13	−4.22	<0.001*
Relatedness judgement/ RTs	Intercept	2.81	0.008	366.48	<0.001*
	Probe-consistent vs. Unprimed	1.25	0.3	4.03	<0.001*
	Probe-consistent vs. Probe-inconsistent	1.19	0.3	3.84	<0.001*
Associate production	Intercept	0.39	0.16	2.42	0.015*
	Probe-consistent vs. Unprimed	−0.43	0.09	−5.06	<0.001*
	Probe-consistent vs. Probe-inconsistent	−0.4	0.09	−4.74	<0.001*

Note. The estimates and standard errors for the RT analysis (excluding the intercept) have been multiplied by 100 for ease of interpretability. This includes those values as reported in the text.

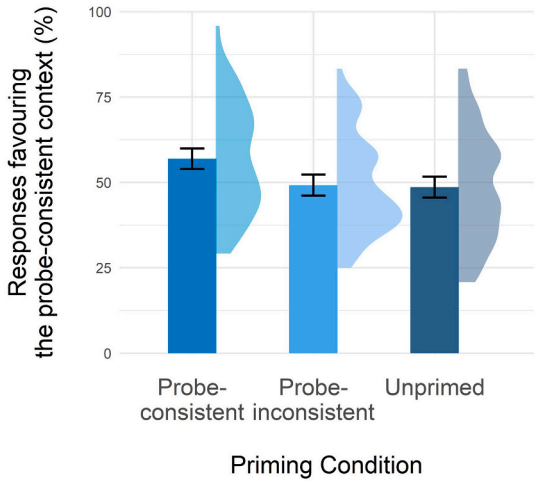

Fig. 3. Proportion of responses favouring the probe-consistent context in the associate production task, Experiment 1. Error bars represent 95% within-subject confidence intervals (Morey, 2008). Density functions represent the distribution of participant means in each condition.

voice and so on are enough to prime the later semantic relatedness and word association tests involving the word "balloon".

Semantic priming is often viewed as a short-lived phenomenon that relies on residual activation during recognition, and typically lasts of the order of hundreds of milliseconds (e.g., Hoedemaker & Gordon, 2017). However, there is also evidence that longer-term semantic priming can be found when attention is focused on meaning. For example, Becker, Moscovitch, Behrmann, and Joordens (1997) found semantic priming using an animacy judgement task with up to eight intervening words between prime and target. They argued that this kind of priming reflected semantic learning rather than residual activation, although others have argued that long-term priming can also be based on memory of specific cognitive operations (Was, Woltz, & Hirsch, 2019). We provide a fuller discussion of the evidence related to semantic priming in the General Discussion, but nonetheless, by comparison to the current paradigm, the lag used to demonstrate long-term semantic priming is often less than a minute, whereas word-meaning priming has been observed in some cases 24 h after exposure (Gaskell et al., 2019).

Rodd et al. (2013) assessed this question directly by presenting participants with two sets of sentences; the first contained homophones in sentential contexts that biased their interpretation towards one of its multiple meanings (i.e., a standard word-meaning priming condition). The second type of sentence was identical to the first except that the critical homophone was replaced by a semantically related word that did not substantially alter the interpretation of the sentence. Although priming was present for both sets of sentences after a short delay (an average across items of ~3 mins between exposure and test), it was only present for sentences that included a homophone after a 20-min delay. Thus, Rodd et al. concluded that word-meaning priming is distinct from more general semantic priming and has greater longevity.

Considering the results from Rodd et al. (2013), it might seem reasonable to assume that the word-meaning priming effects observed in Experiment 1 for non-homonyms were also distinct from semantic priming. However, the average interval between exposure and test in Experiment 1 was in fact intermediate between the two delay periods used in Rodd et al. (2013). For the relatedness judgement task there was a *median* ± *IQR* delay of 7.68 ± 3.57 min, and for the associate production task the delay was 13.93 ± 5.92 min. Conceivably, these delays could have been short enough for some residual semantic priming to influence performance. Therefore, Experiment 2 investigated whether the word-meaning priming effects in Experiment 1 are indeed word-

specific or whether they could be accommodated by a semantic priming account.

3. Experiment 2

Experiment 2 had three goals. First, we wanted to test whether the word-meaning priming effects in Experiment 1 could be replicated using a fully pre-registered design and analysis (https://aspredicted.org/blind.php?x=e9bn6y). Second, we wanted to determine whether word-meaning priming for non-homonyms is distinct from a more general form of semantic priming. To this end, we made a change to the design for Experiment 2. Having shown that the word-meaning priming effect was context specific, we no longer needed the secondary baseline of the probe-inconsistent condition, which was replaced by a semantic priming condition along the lines of Rodd et al. (2013) in which the target word of the prime sentence was replaced with a different word with a similar meaning. This semantic priming condition is referred to as the target-replaced condition from here onwards. Third, we wished to test for both word-meaning priming and semantic priming effects with an exposure-test delay more comparable to the 20 min condition of Rodd et al. (2013) and so a short filler task was added after the exposure phase.

Our pre-registered predictions were that if semantic priming followed Rodd et al. (2013) in terms of its time-course, we should see word-meaning priming but not semantic priming. Specifically:

1. For relatedness judgement, accuracy and response times should be improved for target words in the probe-consistent condition relative to those in the target-replaced and unprimed conditions.
2. For associate production, responses favouring with the primed context should be more common in the probe-consistent condition relative to those in the target-replaced and unprimed conditions.

3.1. Methods

3.1.1. Participants

As pre-registered, and consistent with Experiment 1, a total of 66 participants were recruited from either the University of York ($n = 28$) or Prolific ($n = 38$) (www.prolific.co), all of whom completed the study online. Two participants were excluded from further analysis for reporting that they were not native English speakers. Therefore, the final sample size was 64 (47 females; $M \pm SD$ age in years $= 23.8 \pm 6.21$). These participants were native English speakers with normal or corrected-to-normal eyesight and none reported history of any developmental disorders. Participants gave informed consent as per Experiment 1.

3.1.2. Materials

Like Experiment 1, the experiment made use of 72 target words and three conditions. The probe-consistent and unprimed conditions were as in Experiment 1. For the target-replaced condition, the sentence frames were identical to those in the probe-consistent condition, but the critical target word was replaced by a semantically related word that, as far as possible, did not substantially change the interpretation of the sentence. For example, for the target-probe pair "balloon-helium" the sentence in the probe-consistent condition was "The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky", which became "The entertainer filled the vessel from the gas cylinder and inhaled it to make her voice squeaky" in the target-replaced condition (See Appendix A).

In an attempt to maximise the size of the word-meaning priming effect for greater sensitivity we ran an item-by-item analysis of the priming effect size in Experiment 1. Our thinking was that some sentences might not be suitably precise or selective in terms of biasing towards a particular aspect of the target word's meaning. Six items were replaced for this reason. Furthermore, there were some items for which the associate responses in the control (unprimed) condition were often judged as being similar to the probe-consistent sentence. For these items it would be hard for the probe-consistent sentences to have any biasing effect because the baseline level of priming was already high. On this basis, 21 items were modified by having the probe-consistent and probe-inconsistent contexts switched and the probe word replaced by an alternative associate (see Appendix A). For this partially new set of materials, target word frequency ranged from 0.11–48.26 per million ($M = 14.64$, $SD = 12.65$), target word length ranged from 4 to 10 letters ($M = 6.2$, $SD = 1.6$) and sentence word count ranged from 7 to 23 words ($M = 14.24$, $SD = 3.78$). The items were then randomly split into three lists matched on target word length $F(1, 70) = 1.17$, $p = .283$, target word frequency, $F(1, 70) = 0.08$, $p = .779$, and sentence word count $F(1, 70) = 0.37$, $p = .545$. These lists were then used to construct six counterbalanced sets of stimuli.

3.1.3. Procedure

The general procedure of this experiment was near-identical to Experiment 1. As before, the experiment started with an exposure phase, followed by two tasks: a relatedness judgement task, and an associate production task. In contrast to Experiment 1, we added a filler task after the exposure phase to introduce a short delay between exposure and test. During the filler task, participants watched a 10-min clip of "Shaun the Sheep" (https://www.shaunthesheep.com/), chosen for its minimal linguistic content.

3.2. Results

The analysis approach followed the pre-registered plan and was near-identical to that described in Experiment 1.

For both tasks, the independent variable, Priming, had three levels: probe-consistent, target-replaced and unprimed. Similar to Experiment 1, treatment contrasts allowed us to determine whether there was word-meaning priming (probe-consistent vs. unprimed) and whether any priming found for this condition differed from semantic priming (target-replaced vs. unprimed).

The introduction of the filler task made the interval between exposure and test in Experiment 2 more comparable to the 20-min condition in Rodd et al. (2013) for both the relatedness judgement task (*median* $\pm$ *IQR* $= 20.12 \pm 4.87$ min) and the associate production task (*median* $\pm$ *IQR* $= 26.47 \pm 6.63$ min).

3.2.1. Relatedness judgement

3.2.1.1. Accuracy. Mean accuracy rates in the three conditions are summarised in the left panel of Fig. 4. Accuracy was greater in the probe-consistent condition ($M = 86.6\%$, $SD = 8.8\%$) than in the unprimed condition ($M = 83.8\%$, $SD = 11.8\%$) ($\beta = -0.29$, SE $= 0.11$, $z = -2.51$, $p = .012$; see Table 2). Likewise, accuracy in the probe-consistent condition was greater than that in the target-replaced condition ($M = 84.6\%$, $SD = 10.6\%$) ($\beta = -0.24$, SE $= 0.12$, $z = 2.06$, $p = .039$).

3.2.1.2. Response times. Mean RTs in the three conditions are summarised in the right panel of Fig. 4. Log transformation successfully normalised the residual distribution. In contrast to Experiment 1, there was no significant RT difference between the probe-consistent ($M = 706$ ms, $SD = 94$ ms) and unprimed conditions ($M = 716$ ms, $SD = 94$ ms) ($\beta = 0.49$, SE $= 0.35$, $t = 1.41$, $p = .159$, see Table 2). The difference between the probe-consistent ($M = 706$ ms, $SD = 94$ ms) and target-replaced conditions ($M = 710$ ms, $SD = 99$ ms) ($\beta = 0.09$, SE $= 0.35$, $t = 0.26$, $p = .789$) was also not significant.

3.2.2. Associate production

The percentage of associate responses that favoured the probe-

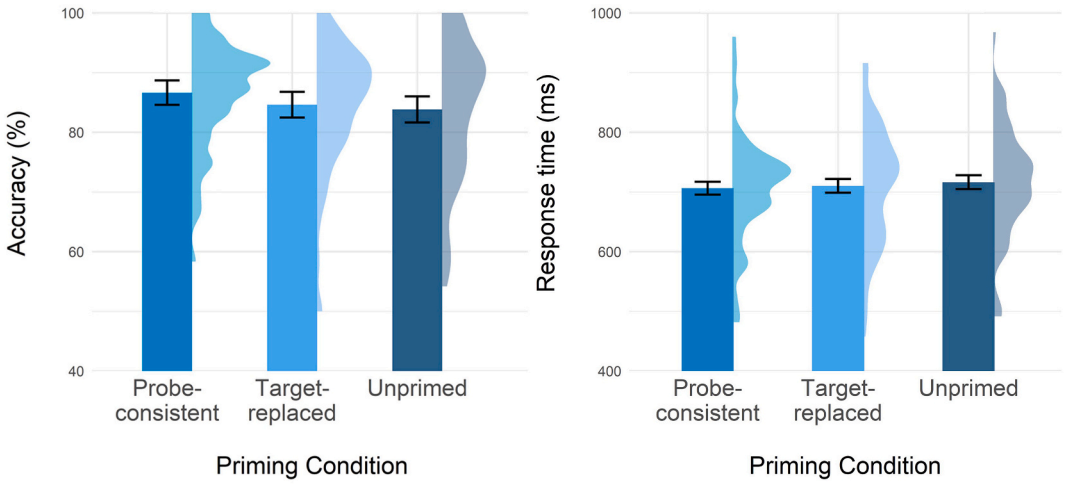

Fig. 4. Accuracy (left) and RTs (right) in the relatedness judgement task for Experiment 2. Error bars represent 95% within-subject confidence intervals (Morey, 2008). Density functions represent the distribution of participant means in each condition.

Table 2
Summary of the mixed-effects models for the two tests in Experiment 2.

Task/dependent variables		Estimates	Std error	*z* / *t* values	*p*
Relatedness judgement/ Accuracy	Intercept	2.54	0.2	12.63	<0.001*
	Probe-consistent vs. Unprimed	−0.29	0.11	−2.51	0.012*
	Probe-consistent vs. Target-replaced	−0.24	0.12	−2.06	0.039*
Relatedness judgement/ RTs	Intercept	2.84	0.008	334.2	<0.001*
	Probe-consistent vs. Unprimed	0.49	0.35	1.41	0.159
	Probe-consistent vs. Target-replaced	0.09	0.35	0.27	0.789
Associate production	Intercept	−0.21	0.19	−1.09	0.278
	Probe-consistent vs. Unprimed	−0.17	0.09	−2.01	0.044*
	Probe-consistent vs. Target-replaced	−0.07	0.09	−0.82	0.414

Note. The estimate and standard error of the RT analysis have been multiplied by 100 (excluding the intercept) to improve interpretability. This includes those values as reported in the text.

consistent context in the three conditions are summarised in Fig. 5. Participants were more likely to produce associates favouring the probe-consistent context in the probe-consistent (M = 46.1%, SD = 17.7%) than in the unprimed condition (M = 42.6%, SD = 18.9%) (β = −0.17, SE = 0.09, z = −2.01, p = .044; see Table 2). However, the percentage of associates favouring the probe-consistent context did not differ between the probe-consistent (M = 46.1%, SD = 17.7%) and target-replaced conditions (M = 44.3%, SD = 19.6%) (β = −0.07, SE =0.09, z = −0.82, p = .414).

3.3. Discussion

Experiment 2 once again showed good evidence that word-meaning priming can be observed for non-homonyms. In a pre-registered conceptual replication, with a longer delay between sentence exposure and test, we found word-meaning priming effects both for the word production test and for accuracy in the relatedness judgements. The effect for RTs was in the predicted direction but not significant. Nonetheless,

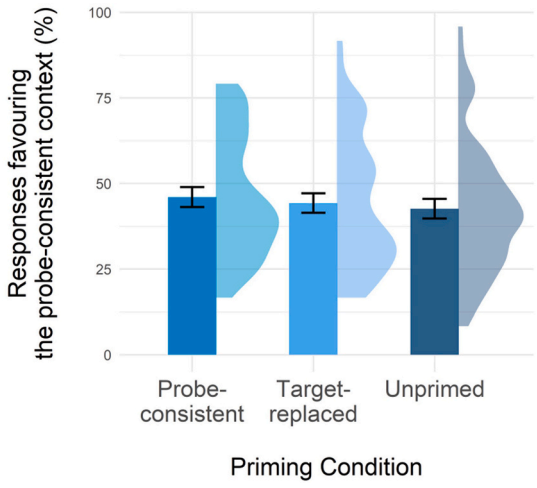

Fig. 5. The percentage of responses that favoured the probe-consistent context in the associate production task, Experiment 2. Error bars represent 95% within-subject confidence intervals (Morey, 2008). Density functions represent the distribution of participant means in each condition.

across the overall pattern of performance in the two tests we see strong evidence suggesting that encountering non-homonyms in a particular sentential context can influence the way they are interpreted after a delay of 20–30 min.

A further goal in Experiment 2 was to determine whether word-meaning priming could be dissociated from semantic priming at this lag. To this end, we assessed whether priming could be observed when the target word in the probe-consistent context was replaced by a word with a similar meaning. In this respect, our results were more mixed. For accuracy in the relatedness judgement task, there was a clear pattern of performance, with participants being more accurate when responding to target-probe pairs in the probe-consistent condition than those in the target-replaced condition. This finding suggests that word-meaning priming contains a word-specific component, and is not solely driven by semantic priming. However, for the other dependent variables performance in the target-replaced condition tended to be more intermediate between the other two conditions, and word-meaning priming was not significantly stronger than semantic priming. Overall, the evidence that word-meaning priming effects can be found for non-homonyms is substantial, but the evidence that these effects are separable from semantic priming effects is less compelling. Our goal for the next

experiment (Experiment 3) was to provide further evidence relating to this question, ideally from a base of a stronger word-meaning priming effect so that differences between word-meaning and semantic priming could be more easily observed.

It is possible that the default preferred interpretation of a target word may influence the size of the priming effect it produces. Previous word-meaning priming studies have shown that priming effects are larger for meanings that are more strongly subordinate (Gaskell et al., 2019; Rodd et al., 2013). This indicates that it is easier to improve accessibility of an ambiguous word's meaning if that meaning is relatively inaccessible in the first place. We hypothesised that the same principle holds for the different senses or aspects of meaning of non-homonyms. To investigate this, we ran some exploratory analyses investigating the relationship between the baseline interpretation of the unprimed target words and the strength of priming in the probe-consistent condition (See Appendix B). In general, there was a modest effect in the predicted direction. When a prime sentence guided the reader towards an interpretation of the target word that was otherwise less preferred, it tended to show larger priming effects, particularly in Experiment 1.

This analysis provided an opportunity to refine our materials further in an attempt to boost the overall strength of priming. Experiment 3 therefore aimed to replicate Experiment 2 while controlling for the baseline preferred interpretations of the target words. By pairing the target words with probe words that related to dispreferred interpretations of the target word's meaning, we aimed to increase the strength of priming effects in order to enhance our assessment of the relationship between word-meaning and semantic priming.

4. Experiment 3

Experiment 3 used a pre-registered design similar to that used in Experiment 2. (https://aspredicted.org/blind.php?x=av3i4b). The primary goal of Experiment 3 was a replication of Experiment 2 with materials that better controlled the baseline preferred interpretations of the target words. That is, target words were to be paired with probe words that related to the less-preferred, subordinate, interpretation of the target word's meaning. As such, the design of Experiment 3 was identical to that of Experiment 2 except most of the target-probe word pairs and their accompanying sentence frames were altered so that they primed less accessible interpretations of the target words. The pre-registered predictions were identical to Experiment 2.

4.1. Methods

4.1.1. Participants

A total of 73 participants were recruited using Prolific (www.prolific.co). As per our pre-registered exclusion criteria, five participants were excluded because their accuracy was less than 70% on the relatedness judgement task. A further two participants were excluded for reporting that their first language was not English. This left us with a final sample of 66 participants of whom 42 identified as female, 22 as male and 2 as other ($M \pm SD$ age in years $= 27.18 \pm 7.25$). These participants were native English speakers with normal or corrected-to-normal eyesight and none reported history of any developmental disorders. Participants gave informed consent before the study began.

4.1.2. Materials

Experiment 3 required items in which the probe word related to a dispreferred interpretation of the target word and the probe-consistent context guided the reader towards this interpretation. Using the data from Experiments 1 and 2 we determined that there were 13 items which could be carried over to Experiment 3 unchanged as they appeared to consistently prime dispreferred interpretations. For the remaining 65 targets which were used across Experiments 1 and 2, a new probe and corresponding priming sentence were developed. To determine whether these new materials primed dispreferred interpretations of the target words we ran a rating pre-test similar to the post-tests used to determine consistency in the associate production tasks of the previous experiments. For each new target-probe pair, no more than 40% of participants in this pre-test produced an associate that favoured the context of the newly developed priming sentence (*range* = 0–39.5%, $M = 11\%$, $SD = 11\%$). This indicates that the new probe words relate to dispreferred interpretations of the target words' meanings.

Target word frequency for this new set of materials ranged from 0.11–48.26 per million ($M = 14.53$, $SD = 12.89$), target word length ranged from 4 to 13 letters ($M = 6.25$, $SD = 1.79$) and sentence word count ranged from 7 to 21 words ($M = 12.61$, $SD = 3.17$). These items were split into three lists. All three lists were matched on target word length $F(1, 70) = 0.006$, $p = .936$, mean target word frequency, $F(1, 70) = 0.12$, $p = .731$, and mean sentence word count $F(1, 70) = 0.35$, $p = .558$. These lists were then used to construct six counterbalanced sets of stimuli.

4.1.3. Procedure

Other than the change of materials, the procedure of Experiment 3 was as in Experiment 2. As before, the experiment had three conditions: probe-consistent, target-replaced, and unprimed. The experiment started with an exposure phase followed by an interval in which participants watched a 10-min video. Following this, participants completed a relatedness judgement and an associate production task.

4.2. Results

The analysis approach followed the pre-registered plan as in Experiment 2. Analysis of the interval between exposure and test in Experiment 3 revealed that the median interval duration was comparable to that of Experiment 2 for both the relatedness judgement ($median \pm IQR = 19.42 \pm 4.06$ min) and the associate production task ($median \pm IQR = 25.6 \pm 6.26$ min).

First, we investigated whether the materials used in Experiment 3 differed from those in Experiments 1 and 2 in terms of the tendency of the probe words to relate to less preferred interpretations of the target words. This was done by looking at the tendency of participants to produce probe-consistent associates in the unprimed condition of the associate production task. For each item we calculated an unprimed score defined as the proportion of associate responses that favoured the probe-consistent context in the unprimed condition. A one-way analysis of variance revealed that unprimed scores differed between the three experiments, $F(2,213) = 8.96$, $p < .001$. Post-hoc Holm corrected t-tests revealed that unprimed scores in Experiment 1 and 2 did not significantly differ, $t(141.53) = 1.49$, $p = .127$. However, unprimed scores in Experiment 3 were lower than in Experiment 1, $t(139.94) = 4.18$, $p < .001$, and in Experiment 2, $t(141.42) = 2.74$, $p = .017$, confirming that the materials in Experiment 3 were priming dispreferred interpretations of the target words more consistently than those in Experiments 1 and 2.

4.2.1. Relatedness judgement

4.2.1.1. Accuracy. Accuracy rates in the three conditions are summarised in the left panel of Fig. 6. Accuracy was significantly greater in the probe-consistent ($M = 82.4\%$, $SD = 11.5\%$) than in the unprimed condition ($M = 79\%$, $SD = 11.7\%$) ($\beta = -0.29$, SE $= 0.1$, $z = -2.86$, $p = .004$; see Table 3). Accuracy in the target-replaced condition was intermediate between the other two conditions ($M = 80.5\%$, $SD = 10.9\%$), and not significantly different from the probe-consistent condition ($\beta = -0.15$, SE $= 0.1$, $z = -1.41$, $p = .159$).

4.2.1.2. Response times. Mean RTs in the three conditions are summarised in the right panel of Fig. 6. After log transformation, the residuals of the RT model were still non-normal with some participants showing very fast RTs. Thus, we removed RTs < 300 ms from the model,

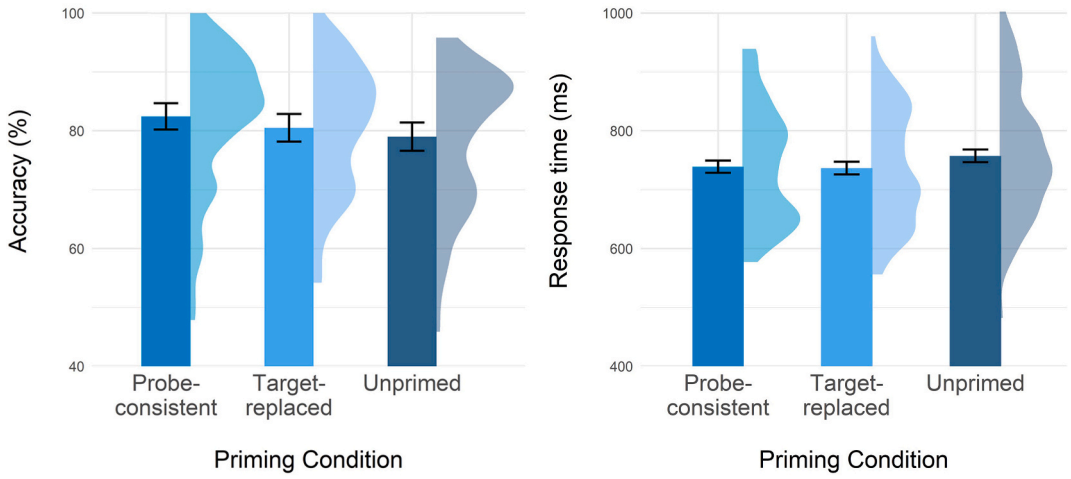

Fig. 6. Accuracy (left) and RTs (right) in the relatedness judgement task for Experiment 3. Error bars represent 95% within-subject confidence intervals (Morey, 2008). Density functions represent the distribution of participant means in each condition.

Table 3
Summary of the mixed-effects models for the two tests in Experiment 3.

Task/dependent variables		Estimates	Std error	*z* / *t* values	*p*
Relatedness judgement/ Accuracy	Intercept	2.04	0.18	11.14	<0.001*
	Probe-consistent vs. Unprimed	−0.29	0.1	−2.86	0.004*
	Probe-consistent vs. Target-replaced	−0.15	0.1	−1.41	0.159
Relatedness judgement/ RTs	Intercept	2.85	0.008	345.7	<0.001*
	Probe-consistent vs. Unprimed	0.96	0.35	2.71	0.003*
	Probe-consistent vs. Target-replaced	−0.41	0.35	−1.15	0.251
Associate production	Intercept	−0.61	0.16	−3.7	<0.001*
	Probe-consistent vs. Unprimed	−0.27	0.09	−3.19	0.001*
	Probe-consistent vs. Target-replaced	−0.03	0.08	−0.41	0.682

Note. The estimate and standard error of the RT analysis have been multiplied by 100 (excluding the intercept) to improve interpretability. This includes those values as reported in the text.

which successfully normalised the residuals. This decision was based on evidence that participants are unlikely to be able to make conscious semantic decisions in less than 300 ms (Pylkkänen & Marantz, 2003). RTs were significantly faster in the probe-consistent (M = 732 ms, SD = 103 ms) than in the unprimed condition (M = 750 ms, SD = 107 ms) (β = 0.96, SE = 0.35, t = 2.71, p = .007; see Table 3). However, there was no significant difference in RTs between the probe-consistent (M = 732 ms, SD = 103 ms) and target-replaced conditions (M = 724 ms, SD = 98 ms) (β = −0.41, SE = 0.35, z = −1.15, p = .251).

4.2.2. Associate production

For this analysis, one participant was excluded, giving a sample size of 65. This participant was excluded for providing the target word as a response on each trial (e.g., target: balloon, response: balloon). The proportion of associate responses favouring the probe-consistent context in the three conditions are summarised in Fig. 7. Participants were more likely to produce associates favouring the probe-consistent context in the probe-consistent (M = 38.3%, SD = 15.7%) than in the unprimed

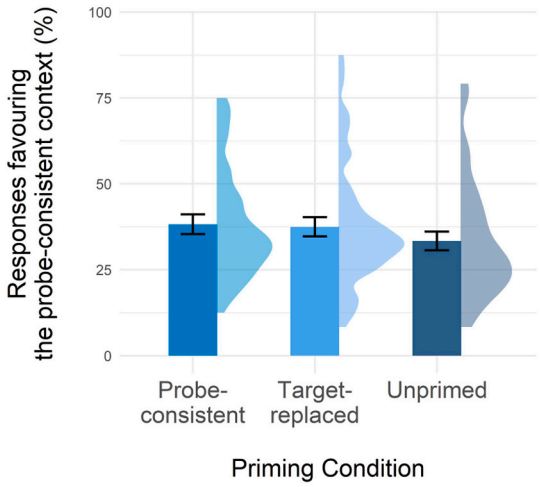

Fig. 7. The proportion of responses that favoured the probe-consistent context in the associate production task, Experiment 3. Error bars represent 95% within-subject confidence intervals (Morey, 2008). Density functions represent the distribution of participant means in each condition.

condition (M = 33.4%, SD = 16.9%) (β = −0.27, SE = 0.09, z = −3.19, p = .001; see Table 3). However, there was no significant difference on this measure between the probe-consistent condition and the target-replaced condition (M = 37.5%, SD = 17.4%) (β = −0.03, SE =0.08, z = −0.41, p = .682).

Inspection of the preceding RT and associate production analyses reveals a significant priming effect for the probe-consistent condition as compared with the unprimed baseline, in line with our predictions. However, there also appears to be a difference between the target-replaced and unprimed conditions, suggesting the presence of a semantic priming effect. In view of this, we used the R package *emmeans* to run pairwise post-hoc contrasts with Holm-Bonferroni correction for multiple comparisons.

These contrasts revealed that there was no significant difference between the target-replaced and unprimed conditions for accuracy in the relatedness judgement task (β = 0.15, SE = 0.1, z = 1.45, p = .295). However, a significant semantic priming effect was present when comparing the target-replaced and unprimed conditions in both the relatedness judgement RT (β = −1.36, SE = 0.36, z = −3.83, p < .001) and associate production analyses (β = 0.24, SE = 0.09, z = 2.78, p = .011).

4.3. Discussion

The primary aim of Experiment 3 was to provide further evidence on the nature of word-meaning priming using more precisely developed materials. By using materials in which the prime sentence guided interpretation of the target word towards an otherwise dispreferred aspect of its meaning, we expected the word-meaning priming effect would be strengthened, providing more of an opportunity to determine whether it could be distinguished from semantic priming. The word-meaning priming effect was indeed robust across all three dependent variables in this experiment, although numerically speaking it was only slightly larger than in Experiment 2. Interestingly, this somewhat enhanced effect was also seen in semantic priming. There were no significant differences between probe-consistent and target-replaced conditions in any of the measures, and although our planned analyses did not directly compare target-replaced and unprimed conditions, post hoc comparisons did suggest that some semantic priming was present. This, in its own right, is a noteworthy finding as semantic priming is generally considered to be short-lived, observed across the time-scale of only a few minutes (McNamara, 2005). Here, we observed semantic priming after a delay of more than 25 min. We will revisit this point in the General Discussion where we flesh out the broader implications of the present results. We now move on to a combined analysis of the data from Experiments 2 and 3.

5. Combined analysis - Experiments 2 and 3

To get the clearest indication of the pattern of results across Experiments 2 and 3, we ran an exploratory analysis on the combined datasets. These analyses followed the same analysis approach as the previous analyses; however, the structure of the model differed slightly due to the alteration of the materials between Experiments 2 and 3. In total, there were 13 target-probe pairs that appeared in both experiments, 54 target-probe pairs had a target that appeared in both experiments but the probe differed between Experiments 2 & 3, and each experiment had 5 target-probe pairs that were unique to that experiment. To address this, our model for the combined analyses treated each unique target-probe pair as an individual item, meaning some items in the combined analysis shared a target word (e.g., Experiment 2: balloon-helium, Experiment 3: balloon-rubber). Furthermore, the models were constructed so that both items and participants were nested within experiments. However, the model building process revealed that having Experiment as a nesting variable led to a singular fit or a failure to converge for all analyses and so Experiment was removed as a variable from our models.

5.1. Relatedness judgement

In line with the analysis of response times in Experiment 3, this analysis removed all trials with an RT less than 300 ms.

5.1.1. Accuracy

Mean accuracy rates in the three conditions across Experiments 2 and 3 are summarised in Fig. 8a. Across Experiments 2 and 3, accuracy in the probe-consistent condition (M = 84.5%, SD = 10.4%) was significantly greater than in both the unprimed (M = 81.4%, SD = 11.9%) (β = −0.29, SE = 0.08, z = −3.78, p < .001; see Table 4) and target-replaced conditions (M = 82.6%, SD = 10.9%) (β = −0.18, SE = 0.08, z = −2.38, p = .017).

5.1.2. Response times

Mean RTs in the three conditions across Experiments 2 and 3 are summarised in Fig. 8b. Log transformation of the data successfully

Table 4
Summary of the mixed-effects models for the two tests in the combined analysis of Experiments 2 and 3.

Task/dependent variables		Estimates	Std error	z / t values	p
Relatedness judgement/ Accuracy	Intercept	2.3	0.14	16.59	<0.001*
	Probe-consistent vs. Unprimed	−0.29	0.08	−3.78	<0.001*
	Probe-consistent vs. Target-replaced	−0.18	0.08	−2.38	0.017*
Relatedness judgement/ RTs	Intercept	2.85	0.6	472.11	<0.001*
	Probe-consistent vs. Unprimed	0.72	0.25	2.91	0.004*
	Probe-consistent vs. Target-replaced	−0.15	0.25	−0.6	0.552
Associate production	Intercept	−0.33	0.13	−2.59	0.01*
	Probe-consistent vs. Unprimed	−0.22	0.06	−3.71	<0.001*
	Probe-consistent vs. Target-replaced	−0.05	0.06	−0.88	0.377

Note. The estimate and standard error of the RT analysis have been multiplied by 100 (excluding the intercept) to improve interpretability. This includes those values as reported in the text.

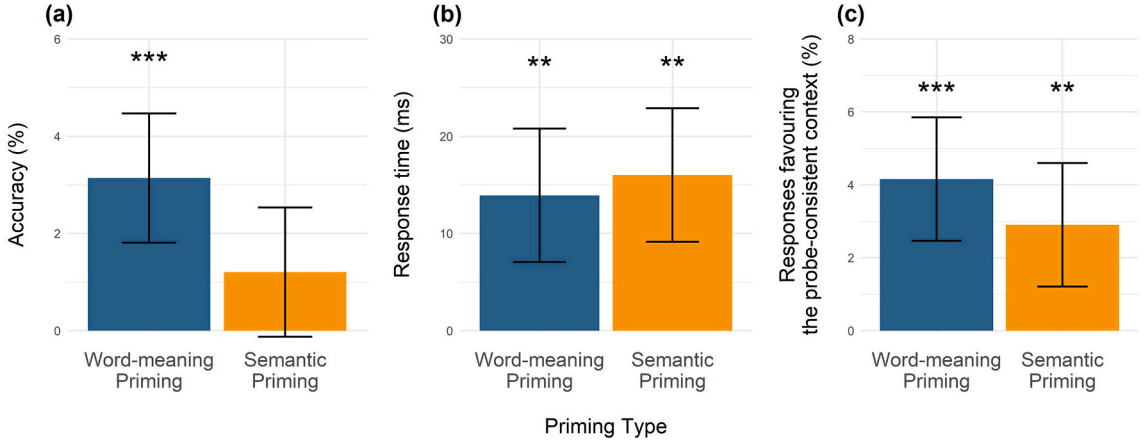

Fig. 8. Difference from the unprimed condition for word-meaning (probe-consistent) and semantic (target-replaced) priming in the combined analysis of Experiments 2 & 3 for (a) relatedness judgement accuracy (b) relatedness judgement response time (c) associate production. Error bars represent 95% within-subject CI (Morey, 2008). Asterisks represent significance from unprimed after Holm-Bonferroni correction for multiple comparisons: * < 0.05, ** < 0.01, *** < 0.001.

normalised the residual distribution. RTs were significantly faster in the probe-consistent (M = 719 ms, SD = 99 ms) than in the unprimed condition (M = 733 ms, SD = 102 ms) (β = 0.72, SE = 0.25, z = 2.91, p = .004). However, there was no significant difference between RTs in the probe-consistent (M = 719 ms, SD = 99 ms) and target-replaced conditions (M = 717 ms, SD = 99 ms) (β = −0.15, SE = 0.25, z = −0.6, p = .552).

5.2. Associate production

The proportion of associate responses that favoured the probe-consistent context in the three conditions across Experiments 2 and 3 are summarised in Fig. 8c. Participants were more likely to produce associates favouring the probe-consistent context in the probe-consistent (M = 42.1%, SD = 17.1%) than in the unprimed condition (M = 38%, SD = 18.5%) (β = −0.22, SE = 0.06, z = −3.71, p < .00; see Table 4). However, the proportion of associates favouring the probe-consistent context did not differ between the probe-consistent (M = 42.1%, SD = 17.1%) and target-replaced conditions (M = 40.9%, SD = 18.8%) (β = −0.05, SE =0.06, z = −0.88, p = .377).

As with Experiment 3, we also ran post-hoc contrasts with Holm-Bonferroni correction for multiple comparisons on the combined data to investigate whether any of the semantic priming effects were significant.

There was no significant difference between the target-replaced and unprimed conditions for accuracy in the relatedness judgement task (β = 0.1, SE = 0.07, z = 1.4, p = .162). However, there was a significant semantic priming effect when comparing the target-replaced and unprimed conditions in both the relatedness judgement RT (β = −0.87, SE = 0.25, z = −3.48, p = .002) and associate production analyses (β = 0.17, SE = 0.06, z = 2.83, p = .009).

Overall, the results from the combined analysis of Experiments 2 and 3 were very similar to the results of Experiment 3 alone further verifying the notion that word-meaning and semantic priming do not substantially differ.

6. General discussion

The present study investigated whether word-meaning priming effects, previously observed for homonyms, generalise to non-homonyms. Across three experiments, we consistently found that an encounter with a non-homonym in a context that refined its interpretation towards a specific aspect of its meaning led to a preference for a similar interpretation at a later time point (up to 20–30 min later). Evidence for this was observed in both speeded semantic judgement and free association. Thus, word-meaning priming effects do generalise to non-homonyms. As outlined in the introduction, this finding is predicted by an episodic context account of language comprehension (Gaskell et al., 2019). Such an account asserts that the comprehension of *any* sentence should lead to the generation of a contextually bound representation of the sentence at some level of abstraction that can be used to facilitate subsequent discourse comprehension and assist in updating of lexical knowledge.

This episodic context account was initially used to explain prior evidence of word-meaning priming for homonyms. In such cases there is a clear learning goal for the comprehender: to update or fine-tune the balance of different meanings of the homonym so that they are appropriately weighted for the next time that word is encountered. This fine-tuning ensures that the accessibility of the different unrelated meanings of homophones are optimised, based on the statistics of the language environment. The episodic context account proposed that successful comprehension of a sentence or utterance would lead to a memory of the comprehension episode that could provide a secondary source of knowledge to aid in the interpretation of later words and sentences, alongside standard lexical knowledge. The implication of experimental data from Gaskell et al. (2019) showing that word-meaning priming was preserved after time asleep more than time awake was that consolidation could provide a mechanism for the new episodic representation to gradually adjust lexical knowledge to ultimately ensure that meaning frequencies for homonyms reflect the statistical properties of the language.

Gaskell et al. (2019) speculated that the episodic representations that help to update lexical meaning frequencies of homonyms might also support the maintenance of an event model or situation model (Kintsch, 1988) to facilitate discourse comprehension over longer periods of linguistic input (e.g., across pages of a novel, or to understand conversations lasting minutes or longer). The current results perhaps enhance this prospect. For the target non-homonyms that we used here, there is arguably less to learn about their meaning from their sentential context during exposure, so the lexical updating value of the episodic memory may be lesser in these cases (we will return to this point in more detail). Nonetheless, there is an important contribution that such memories could make to the mental record of the discourse. So conceivably the episodic contextually rich memories of linguistic input that support word-meaning priming could have quite a flexible purpose. In cases where lexical knowledge needs updating then they could maintain new information long enough for it to be consolidated offline. At the same time, they could help maintain discourse information to allow language comprehension to operate effectively across timespans ranging from seconds and minutes through to hours and potentially days and weeks. This latter property fits well with the observation that language is prone to clustering at a range of levels, such that expectation of repetition is rational (e.g., Myslín & Levy, 2016).

The notion of the language comprehension system making use of episodic information is not new (cf. Dumay & Gaskell, 2007; Goldinger, 1998; Jiang & Forster, 2001; Witzel & Forster, 2012). For example, memories for novel words are not fully integrated into the lexicon prior to a period of consolidation (e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). These new memories may well be initially episodic in nature, but often prove to be well able to allow comprehension to proceed smoothly (e.g., Borovsky, Elman, & Kutas, 2012). Furthermore, there is a growing consensus that the mental lexicon involves both episodic and abstractionist aspects of knowledge, and that the balance between the two depends on learning circumstances and experience (e.g., Kapnoula & Samuel, 2019; Pufahl & Samuel, 2014). Our proposal extends the range of circumstances in which episodic context is called into play. Rather than responding to situations of novelty (e.g., a new word) our argument is that episodic representations of utterances are routinely recruited during language comprehension, so that subsequent discourse comprehension and long-term lexical knowledge can be optimised.[2]

This characterization of the role of episodic knowledge in language comprehension fits well with what we understand from studies of language comprehension when episodic memory formation is heavily compromised through hippocampal damage. It is clear that simple language comprehension tasks involving encapsulated sentences can operate well in patients with amnesia, and that moment-by-moment components of sentence comprehension (e.g., prediction) are largely intact (Brown-Schmidt, Cho, Nozari, Klooster, & Duff, 2021). Problems emerge, however, when information needs to be integrated across longer domains, such as multiple sentences or paragraphs (Covington, Kurczek, Duff, & Brown-Schmidt, 2020; Kurczek, Brown-Schmidt, & Duff, 2013). Similarly, neuroimaging data suggest a distinction between the core language network crucial for successful comprehension of individual sentences, and an episodic network that is increasingly

[2] Just how routine this recruitment is remains an open question. Our paradigm made use of sentence probes that encouraged the reader to fully comprehend the sentences. Gilbert et al. (2021) showed that in situations where comprehension may be less complete word-meaning priming effects diminished, suggesting that good comprehension may be a requirement for the generation of an enduring episodic memory.

recruited over longer timescales such as paragraphs (Blank & Fedorenko, 2020).

Alongside the probe-consistent and unprimed conditions, each experiment had a third condition. In Experiment 1, this involved exposure to the target word but in a context that was inconsistent with the probe. Importantly, this kind of exposure was not effective in priming, showing that the nature of the context matters rather than just prior exposure to a particular word. In the other experiments, the third condition involved the probe-consistent sentence context during the exposure phase, but with the target word replaced by a semantically similar word. This condition allowed us to determine how general the priming effects that we saw were. Specifically, this condition tested whether priming effects are tied to particular word-forms or whether they can be found more broadly for semantically related words. In Rodd et al.'s (2013) original study on homonyms, their Experiment 3 had a similar goal. They used prime sentences for which the ambiguous word (e.g., *post* in "The man accepted the post in the accountancy firm") was replaced with a non-homonym that had a similar meaning (e.g., *job*) to determine whether the prime sentence actually had to include the target word in order to see priming effects, or whether more general semantic priming from the broader sentence context could explain their results. Although semantic priming was observed at a short delay between the priming phase and test (~3 mins), no such priming was found at a longer delay (~20 mins), whereas word-meaning priming was seen at both delays. Rodd et al. used this dissociation to argue that word-meaning priming is not simply semantic priming.

Here, we did not observe such a clear dissociation. In our combined analysis, there was evidence that word-meaning priming was stronger than semantic priming after 20–30 min in the relatedness judgement accuracy data; however, this difference was not observed for relatedness judgement RTs or word association responses. Furthermore, the latter two dependent variables did show semantic priming effects in the absence of the target word from the prime sentence. So here, the overall pattern seems to be that word-meaning priming effects are robust after 20–30 min, but (somewhat weaker) semantic priming effects can also be observed. This does not invalidate the basic finding of word-meaning priming for non-homonyms, but it does raise interesting questions about the generality of such effects.

As mentioned above, semantic priming has often been considered to be a short-lived effect that persists over a period of seconds, or minutes at best (Becker et al., 1997; Collins & Loftus, 1975), and the semantic priming results of Rodd et al. (2013) were unsurprising at the time. However, research subsequent to Rodd et al. has challenged this assumption and has demonstrated long-term facilitation of semantic priming over longer delays. Woltz, Sorensen, Indahl, and Splinter (2015) found that prior exposure to a simple sentence intended to express a familiar proposition (e.g., *consumers purchase products*) could facilitate the completion of a different sentence expressing the same single proposition (e.g., *shoppers buy merchandise*) over delays of up to 15 min. This finding is important as it demonstrates long-term semantic priming for abstract propositional meaning. Woltz et al. discussed how this priming effect could emerge from an attractor-style network model (cf. Becker et al., 1997) that was focused on familiar propositions rather than isolated words (cf. Sharkey, 1990). In many ways, this kind of model is like the immediate alteration account of word-meaning priming described in the Introduction, but dedicated to a higher level of processing.

The Woltz et al. (2015) priming effect is similar in some respects to the priming effects we see in our experiments. The crucial difference, though, relates to prior knowledge. Woltz et al. specifically focused on individual well-known concepts that would have an existing propositional representation in memory. Here, however, the content of the priming sentences was effectively new material (or at least new combinations of existing concepts) as is typical of much of discourse. As such, there would not be a single familiar propositional representation that could be primed in the way specified by Woltz et al. Therefore, we do not believe that Woltz et al.'s theory can explain the semantic priming we found. That said, it is possible that our episodic context account *could* explain the Woltz et al. semantic priming data if the level of representation at which the contextual binding occurs is not at the level of specific word-forms but a more abstract level that involves meaning units. This is equivalent to the "textbase" level of the construction-integration model of Kintsch (1988) at which distinctions between near-synonyms such as *consumers* and *shoppers* would be largely lost. Indeed Woltz et al. acknowledged that the reliance on pre-existing representations in their account was an assumption, and that a more constructive account of priming (e.g., Logan, 1988) would also be compatible with their results.

Returning to our own results, this same more abstract, gist-like version of the episodic context account would likely be able to accommodate both the word-meaning priming effect and the semantic priming effects. The level of representation at which contextual binding might occur was left as an open question by Gaskell et al. (2019). Binding at a superficial word level was seen as one possibility, but the second possibility was that binding would take place at a more gist-like level, compatible with the evidence that enduring memory for text tends to be gist-like (Fisher & Radvansky, 2018; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Mak, Curtis, Rodd, & Gaskell, 2022). Either alternative could explain the word-meaning priming effects for homonyms in Gaskell et al. and for non-homonyms here. But it is easier to explain the semantic priming effect found here if binding occurs at a level for which the often superficial differences between the target word and its replacement in the target-replaced condition are lessened. This interpretation also fits well with recent evidence suggesting that word-meaning priming is strongest following successful comprehension; a more superficial analysis of a sentence containing a target word tends to lead to weaker priming at a later point (Gilbert et al., 2021).

One more point about the semantic priming data should be considered. Given that we have seen good evidence for semantic priming over a timescale of 20–30 min and that this can fit with an episodic context theory that retains contextual memories for words at an abstract level, why was no semantic priming effect observed in Rodd et al.'s (2013) original study of homonyms? This is a difficult question to answer, given that there were substantial differences in the methods and materials across the two studies. The disparity could reflect a real difference in the way in which word-meaning priming operates for homonyms and non-homonyms, but we think that such a conclusion would be premature on current evidence. Rodd et al. used fewer participants (42 in one experiment vs 139 across two experiments) and employed a demanding digit span filler task and both these factors may have had the effect of reducing the likelihood of observing priming effects. Perhaps most importantly, the specificity of the context may well have been different across the two studies. For Rodd et al., it was relatively easy to generate sentences that could select between the unrelated meanings of the target homonyms. Here, we were using the sentences to specify particular aspects of meaning, and so we needed to choose them carefully. This may have had the effect of narrowing the focus of the sentences and so enhancing the likelihood of observing enduring semantic priming effects.

The impetus for the current research was a prediction attributed to the episodic context account of word-meaning priming. Gaskell et al. (2019) speculated that that word-meaning priming might be found for *all* words rather than just lexically ambiguous words. This prediction derived in part from a growing consensus that some degree of semantic disambiguation is required for successful comprehension for most, if not all, words (Rodd, 2020). Most common word are polysemous, i.e., are ambiguous between different highly related word senses (as were most of the target words we used). For example, a word like "school" can refer to either the physical building or a community of people. Successful comprehension of such words requires selection of the contextually appropriate semantic features associated with the intended word sense. Although the presence of multiple (related) senses may in some

circumstances have different effects on lexical access compared with multiple (unrelated) meanings (e.g., Rodd et al., 2002), there is increasing evidence of commonalities in how such words are processed (Rodd, 2020). In addition, even for relatively unambiguous words, contextually relevant aspects of their meanings are selected for integration into the ongoing discourse representation. For example, for the word "pineapple" the features related to its taste or shape are differently relevant in contexts in which it is being eaten or picked up (Tabossi, 1988). Likewise, the appropriate physical features of an onion will depend on whether it has just been chopped or weighed (for review, see Altmann & Ekves, 2019). From our current study we cannot say for sure at what semantic level the enduring memory for word meaning is operating. Most of the target words had multiple senses, but at least in Experiment 1 the sentence context manipulation generally did not relate to different dictionary senses of the words, and the pattern of performance held even when just the same-sense items were analysed. Furthermore, in analyses suggested by an anonymous reviewer, we included the presence of multiple senses, the number of senses and whether or not the contextual manipulation crossed sense boundaries as covariates in the main analyses of priming (exploratory analyses available at https://osf.io/yg428/), and saw little or no evidence of any pattern to the priming. Word-meaning priming appears to be ubiquitous and difficult to explain in terms of an adjustment of the balance between separate meanings or senses of a word.

The current results show, therefore, that just as prior experience with a homonym can boost the future availability of the semantic features that are congruent with currently selected meaning, prior experience with non-homonyms can boost the subset of contextually relevant semantic features. The finding of word-meaning priming for non-homonyms therefore supports the general view that lexical ambiguity is best viewed as a continuum (Rodd, 2020). This view is consistent with connectionist network accounts of word semantics, in which familiar word meanings/senses correspond to basins of stability within a high-dimensional semantic space (e.g., Armstrong & Plaut, 2008; Joordens & Besner, 1994; Rodd, 2020; Rodd, Gaskell, & Marslen-Wilson, 2004). Distributional models of semantics are also consistent with this characterization. These accounts imply that levels of semantic ambiguity can be captured by a word's distribution in their linguistic context, with more diverse contexts linked to greater variability in meaning (McDonald & Shillcock, 2001; Hoffman, Lambon Ralph, & Rogers, 2013, 2021; Hsiao & Nation, 2018; Jones, Johns, & Recchia, 2012; see also Johns, Dye, & Jones, 2016; Mak, Hsiao, & Nation, 2021). The extension of word-meaning priming to non-homonyms fits naturally with all of these accounts, in which the sense/meaning distinction is a matter of degree of ambiguity rather than any hard division. Under this view, while semantic disambiguation demands are necessarily greatest for homonyms, because the reader/listener must select between two (or more) mutually exclusive, unrelated meanings, the need to select contextually relevant lexical features and integrate them into the ongoing discourse is also needed, albeit to a lesser extent, for words with a lower degree of ambiguity.

We still do not know whether all words are susceptible to this kind of priming, but the current extension of word-meaning priming to non-homonymic nouns is an important one. The treatment of homonyms and non-homonyms as different points on a continuum of semantic ambiguity helps to show how our current results could be explained by an immediate alteration account of word-meaning priming. Just as priming effects for homonyms could be explained in terms of an adjustment in the balance of weightings of two separate meanings, so the effects here could be a consequence of a (perhaps finer) adjustment of the distributed semantic representation of a less ambiguous word. Nothing in the current results rules this account out, but our preferred interpretation at the moment is in terms of episodic contextual binding because of its ability to explain sleep effects on the longevity of word-meaning priming for homonyms. One potential avenue for further research to tease these theories apart would be to assess whether word-meaning priming effects are seen for unrelated words that are encountered in a coherent sentential context, given that they would have no pre-existing semantic relationship to adjust (e.g., Mak & Twitchell, 2020). Observation of priming in such circumstances would be strong support for an episodic context account. A second option would be to investigate whether sleep consolidation effects, such as those observed in Gaskell et al. (2019) for homonyms, are also seen for word-meaning priming of non-homonyms. Both these questions are the focus of ongoing research in our lab. More generally, it will be important to characterise the extent to which episodic memories of prior language remain influential in comprehension over the longer term, across multiple instances and in the context of potentially interfering memories.

6.1. Conclusions

Our results mark an important broadening of the scope of word-meaning priming. Rather than just homonyms, our experiments indicate that encountering non-homonyms—and possibly all content words—in a linguistic context leaves a trace in memory that can guide interpretation of those same or similar words 20–30 min later. Like other forms of priming (e.g., syntactic priming) this bias can be seen as a rational adaptive response to the observation that linguistic items tend to be clustered across time rather than randomly spaced (Myslín & Levy, 2016). This behaviour conforms to a prediction of Gaskell et al. (2019) and hence provides evidence in support of an episodic context account, by which successful comprehension of a sentence or utterance leads to the formation of a new memory of that linguistic unit, at a somewhat abstract level of meaning. We propose that this episodic memory trace can then guide the interpretation of subsequent similar sentences, contributes to the understanding of discourse and its retention in memory, and facilitates the later adjustment of lexical representations, perhaps via consolidation over time.

Open science statement

All data and related R scripts are publicly available at: https://osf.io/yg428/?view_only=28e55ea54ab8400895b3e8ddfdf512bb

CRediT authorship contribution statement

Adam J. Curtis: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Matthew H.C. Mak:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Shuang Chen:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – review & editing. **Jennifer M. Rodd:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **M. Gareth Gaskell:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

We have no conflict of interest to disclose.

Acknowledgements

This research was supported by an Economic and Social Research Council grant (ES/T008571/1) to Gareth Gaskell and Jenni Rodd. We thank members of the Sleep, Language and Memory Group at the University of York for their advice and feedback.

Appendix A. Experimental materials

Experiment 1.

Item	Target	Probe	Probe-consistent context	Probe-inconsistent context
1	acorn	seed	If we plant an acorn in the ground and water it, an oak tree will grow.	The family collected a bag of dry acorns and separated the nut meat from the shell to make flour.
2	acre	farm	He drove the tractor out to the fields and ploughed an acre before sunrise.	The garden at the new house is nearly an acre, even though it's in the city.
3	adventure	trip	My friends and I went on an adventure to Scotland together in my mum's old car.	The girl read adventure stories about Tom Sawyer at the library.
4	advertisement	picture	They posted advertisements all over town in the hope that someone would recognize their lost dog.	To increase sales to commuters, we played the advertisement on the radio.
5	ambulance	sickness	If he gets any worse overnight, we must phone for an ambulance.	An ambulance accompanied the search team to get the climbers down from the mountain.
6	ancestor	dead	Finding my ancestor's grave helped me find out what happened to him.	Typewriters are the ancestor of the modern electronic keyboard.
7	ankle	sprain	The doctor said my ankle wasn't broken and suggested rest and ice to reduce the swelling.	Everyone says not to wear them with sandals but they keep my ankles and feet warm and clean.
9	athlete	injury	The athlete fell off her skis and sprained her arm, so she was unable to train.	This athlete has only done cross country before so the 800 m will be her first track event.
10	backpack	student	Ashley slung her backpack on her desk and took out her homework.	The hiker's backpack contained everything she needed to survive the long climb.
11	balloon	helium	The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky.	People looked up at the hot air balloon drifting across the sky.
12	bathtub	relax	I soaked in the bathtub and listened to music for hours.	After our walk I put the muddy dog in the bathtub and scrubbed him with dog shampoo.
13	beard	shave	He bought a new razor because he wanted to get rid of his beard.	He works part time in the winter, giving gifts to children while wearing a long white beard and red coat.
14	bicycle	tyre	He can't use his bicycle these days because he got a puncture last week.	His bicycle needs some oil to ride smoothly.
15	biologist	lab	The biologist stained the specimen before looking at it through the microscope.	He aspired to become a biologist who studies sharks in their natural environment.
16	bone	skeleton	The archaeologists discovered an ancient human bone buried in the tomb.	She almost changed her mind when she saw the long needle that would go into her pelvic bone, but she knew the donation would save lives.
18	breeze	blow	I was enjoying the cool breeze coming from the window, but it sent my papers everywhere.	I could taste the salt in the breeze even a mile away from the beach.
20	cactus	spine	The dog pricked its tongue when it tried to eat the cactus.	Almost all his plants died under the strong sunlight, but not the cactus.
21	cage	trap	We caught the mouse using a cage with a piece of cheese inside.	He bent some thick iron wires and welded them to make a cage.
23	castle	prince	Then Cinderella went to the royal castle to dance at the ball.	The treasure is at the top of the spiral staircase in the highest point of the abandoned castle.
26	chimney	christmas	He jumped down the chimney with a sack full of presents.	The chimney is almost choked up with soot so we'll need to clean it soon.
28	coal	miner	His work is digging coal underground, which is dangerous and tiring.	They piled up some coal and twigs, and then lit it with a match.
29	concert	loud	I wore ear plugs to the rock concert yesterday but my ears were still ringing by the end of it.	We need a page turner to assist our maestro at the Chopin concert this afternoon.
31	cupboard	plates	I opened the cupboard and set the table for dinner.	The cupboard was stuffed with old fishing tackle.
32	dancer	club	They pushed the door seeing a pole dancer under the spotlight and the excited audience cheered for her performance.	She was trained at The Royal Ballet School, gesturing and moving with the elegance of a dancer.
35	dessert	yummy	She served her friends with freshly baked apple pie as dessert, and everyone loved it.	She really wanted some dessert but her personal assistant denied her request as she had to keep fit.
36	device	control	Players use a haptic device such as a joystick to direct the characters in the game.	The device can discriminate between the cancerous and the normal cells.
39	disaster	storm	He heard the thunder in the distance and worried about flooding after last year's disaster.	The train came close to disaster when the brakes failed.
41	dragon	china	The profile of the ancient Great Wall from space looks like a dragon.	Small children often like story books about dragons and other monsters.
46	eraser	blackboard	After the class, the student dusted the chalk off with an eraser.	These answers you pencilled in aren't correct, use an eraser and try again.
47	eyebrows	pluck	She tapered the shape of her eyebrows towards the outer corners with tweezers.	Her eyebrows were arched in supercilious surprise when hearing the news.
48	fangs	snake	Its venom is contained in the fangs, which are long, hollow tubes.	The cinema audience held their breath when his fangs almost touched the neck of the actress.
49	feather	soft	This pillow is filled with feathers for maximum comfort.	To be authentic, he wrote the letter with a feather quill.
50	flask	chemistry	The graduations are marked on the side of the flask so you can know the precise amount of saline solution.	She drank from the little flask she kept in her bag and shuddered when the whisky hit her throat.
51	foam	beer	The bartender was distracted as he poured the pint and so the foam overflowed.	The children only agreed to a bath so they could play with the foam.
52	geometry	teacher	In today's maths lesson, Ms. Dale presented geometry.	The pilot used geometry to systematically work out the safest way to land.
58	gravity	moon	The lunar gravity is only about one-sixth that on earth so an astronaut can carry a heavier bag.	Martha pushed the rock off the edge of the cliff, knowing that gravity would do the rest.
59	herb	medicine	The essence drawn from this herb is good for high blood pressure.	She clipped some mint from the herb garden to make a refreshing tea.
60	hostage	plane	The hijackers held two passengers hostage to get the pilot to change the flight path.	The kidnapper demanded a million pounds to free the hostage.

(*continued on next page*)

(*continued*)

Item	Target	Probe	Probe-consistent context	Probe-inconsistent context
62	infection	amputation	The infection spread up to his knee and they had to remove his leg.	If you have a respiratory infection please cover your mouth when you cough or sneeze.
63	insurance	coverage	Check that your insurance policy includes breakages and damage during removals.	I didn't understand what insurance I needed so I hired someone to help me.
64	intensity	stress	The intensity of the work made me frustrated and anxious.	I coughed at the intensity of my drink, I'd clearly put too much squash in.
65	inventor	genius	The young inventor was known for finding solutions where nobody else could.	The inventor was trying to improve product sales by appealing to businesses.
67	lamp	burn	She lit the oil lamp and watched the small flame grow.	The new lamp was patterned with pictures that were cast about the room when the light shone through.
71	lizard	tongue	This lizard catches insects in the same way frogs do.	He didn't want to pick up the lizard because its skin looked slippery and he was scared he'd drop it.
72	lobster	claw	Ouch, the lobster pinched my finger!	Anton the chef concocts a sensual coupling of lobster and asparagus.
74	luggage	heavy	It was lucky someone helped her lift her luggage onto the train because it was full of books.	They got off the plane and waited for their luggage.
80	museum	painting	We found a Picasso in the museum but we couldn't tell what it was meant to be.	The museum has many historic cultural relics.
81	onion	tears	He had to stop cutting the onion because his eyes were blurry and stinging.	To make the food taste just as he remembered he added onion and coriander.
85	poison	snake	He applied pressure to the bite wound on his arm and hoped there was no poison.	Markov died after poison was secretly placed in his drink.
86	potato	mash	She made sure there were no lumps in the potato before she fed it to her baby.	They went down to the allotment to see what they could harvest, but there was only one potato.
88	predator	shark	The divers knew that they had no hope when they saw the deadliest predator in the ocean.	The police believed that he was a predator who regularly followed people home.
90	quilt	sew	Her grandmother put together a quilt for her birthday with lots of different fabrics and threads.	The child looked completely snug under the thick quilt.
92	receipt	return	We only take goods back if customers can produce the receipt.	After she paid for their meal she put the receipt in her handbag.
93	stamp	collection	He has five albums of stamps, many of which are very rare.	It's best to send your letters first class, so you will need a stamp.
96	ticket	parking	I found a ticket attached to the front window of my car when I came out of the building.	I intend to watch all of my team's games this season because I got a season ticket for my birthday.
98	tray	waitress	He saw a woman holding a tray of glasses pass by and got them more champagne.	She poured water into each cube and then put the tray in the freezer.
99	tulip	bulb	He planted tulips in the garden, making sure they were all the right way up.	They plan to visit the famous tulip kingdom.
101	tunnel	train	The railway company spent a lot of money constructing a tunnel through the mountain and laying new tracks.	He climbed down into the secret tunnel that ran all the way out of the mansion.
104	vacancy	apartment	They wanted somewhere bigger to live so when they heard about a vacancy in their building they jumped at the chance.	We still have vacancies for secretaries but other positions have been filled.
105	vase	rose	He put the Valentine's day bouquet in a vase with some water.	The delicate vase was made by a famous artist and is sure to sell for a high price at auction.
107	victim	murder	The victim was smothered to death with a pillow.	He knew he hadn't stolen anything so he must be the victim of a frame-up.
108	vine	climb	The vine grew up the side of their house, clinging to the wall.	They cut down a vine and used it to tie up their tent.
109	violence	gang	Several suburbs of the city were blighted by organized violence from rival groups.	Many people lose their lives every year as a result of domestic violence.
110	visitor	tourist	As first-time visitors in New York they wanted to see the Statue of Liberty and a show on Broadway.	Throughout his long period in hospital he didn't like visitors coming to see him in case they were shocked.
112	wand	magician	He tapped his wand on the top hat and out hopped a rabbit.	The fairy godmother tapped her wand, and a coach and horses appeared at the back of the stage.
115	weapon	war	They considered using their nuclear arsenal as their ultimate weapon in the hope that the opposition would surrender.	In desperation, the woman picked up a cricket bat as a weapon to protect her family.
117	whiskey	drunk	He'd been drinking whiskey all night, and when he got up he stumbled and his vision blurred.	He wanted to buy a nice whiskey for his boss but it was hard to choose from all the different shapes and sizes.
120	wire	electric	All the wires in the house need to be replaced, so we can't use the plug sockets or lights.	Despite the hole being tiny, he just managed to poke the wire through to tie it to the fencepost.
121	wolf	howl	The lone wolf tilted its head to the sky and called out mournfully for the rest of the pack.	The ravenous wolf snarled as it ripped apart the lamb.
123	worm	fishing	He hooked on the worm as bait and threw out the line in hopes he'd get a catch.	Worms like to live in damp soil and can help plants grow.
124	yogurt	calories	Low fat yogurt can be good for losing weight but you also need to look at sugar content.	She bought fruit-flavored yogurt for her daughter.
125	zone	parking	You can leave your car in the long-stay zone over there.	The central zone of the Earth is the core, and is surrounded by the mantle and then the crust.

Experiment 2 - Underlined words highlight the semantic replacement of the target word.

Item	Target	Probe	Probe-consistent sentence	Target-replaced sentence
2	acre	yard	The garden at the new house is nearly an acre, even though it's in the city.	The garden at the new house is nearly a hectare, even though it's in the city.
5	ambulance	rescue	An ambulance accompanied the search team to get the climbers down from the mountain.	A paramedic accompanied the search team to get the climbers down from the mountain.
6	ancestor	predecessor	Typewriters are the ancestor of the modern electronic keyboard.	Typewriters are the forebear of the modern electronic keyboard.

(*continued on next page*)

(*continued*)

Item	Target	Probe	Probe-consistent sentence	Target-replaced sentence
7	ankle	socks	Everyone says not to wear them with sandals but they keep my ankles and feet warm and clean.	Everyone says not to wear them with sandals but they keep my toes and feet warm and clean.
9	athlete	injury	The athlete fell off her skis and sprained her arm, so she was unable to train.	The competitor fell off her skis and sprained her arm, so she was unable to train.
11	balloon	helium	The entertainer filled the balloon from the gas cylinder and inhaled it to make her voice squeaky.	The entertainer filled the vessel from the gas cylinder and inhaled it to make her voice squeaky.
12	bathtub	relax	I soaked in the bathtub and listened to music for hours.	I soaked in the shower and listened to music for hours.
13	beard	santa	He works part time in the winter, giving gifts to children while wearing a long white beard and red coat.	He works part time in the winter, giving gifts to children while wearing a long white moustache and red coat.
14	bicycle	tyre	He can't use his bicycle these days because he got a puncture last week.	He can't use his bmx these days because he got a puncture last week.
15	biologist	marine	He aspired to become a biologist who studies sharks in their natural environment.	He aspired to become a researcher who studies sharks in their natural environment.
16	bone	skeleton	The archaeologists discovered an ancient human bone buried in the tomb.	The archaeologists discovered an ancient human skull buried in the tomb.
20	cactus	hot	Almost all his plants died under the strong sunlight, but not the cactus.	Almost all his plants died under the strong sunlight, but not the fuchsia.
23	castle	tower	The treasure is at the top of the spiral staircase in the highest point of the abandoned castle.	The treasure is at the top of the spiral staircase in the highest point of the abandoned palace.
26	chimney	Christmas	He jumped down the chimney with a sack full of presents.	He jumped down the fireplace with a sack full of presents.
28	coal	miner	His work is digging coal underground, which is dangerous and tiring.	His work is digging copper underground, which is dangerous and tiring.
29	concert	loud	I wore ear plugs to the rock concert yesterday but my ears were still ringing by the end of it.	I wore ear plugs to the rock show yesterday but my ears were still ringing by the end of it.
31	cupboard	storage	The cupboard was stuffed with old fishing tackle.	The drawer was stuffed with old fishing tackle.
32	dancer	club	They pushed the door seeing a pole dancer under the spotlight and the excited audience cheered for her performance.	They pushed the door seeing a pole girl under the spotlight and the excited audience cheered for her performance.
35	dessert	yummy	She served her friends with freshly baked apple pie as dessert, and everyone loved it.	She served her friends with freshly baked apple pie as pudding, and everyone loved it.
36	device	control	Players use a haptic device such as a joystick to direct the characters in the game.	Players use a haptic tool such as a joystick to direct the characters in the game.
41	dragon	China	The profile of the ancient Great Wall from space looks like a dragon.	The profile of the ancient Great Wall from space looks like a serpent.
46	eraser	blackboard	After the class, the student dusted the chalk off with an eraser.	After the class, the student dusted the chalk off with a rubber.
47	eyebrows	expression	Her eyebrows were arched in supercilious surprise when hearing the news.	Her face was arched in supercilious surprise when hearing the news.
48	fangs	snake	Its venom is contained in the fangs, which are long, hollow tubes.	Its venom is contained in the teeth, which are long, hollow tubes.
49	feather	soft	This pillow is filled with feathers for maximum comfort.	This pillow is filled with cotton for maximum comfort.
50	flask	chemistry	The graduations are marked on the side of the flask so you can know the precise amount of saline solution.	The graduations are marked on the side of the bottle so you can know the precise amount of saline solution.
51	foam	beer	The bartender was distracted as he poured the pint and so the foam overflowed.	The bartender was distracted as he poured the pint and so the bubbles overflowed.
52	geometry	analytical	The pilot used geometry to systematically work out the safest way to land.	The pilot used statistics to systematically work out the safest way to land.
58	gravity	moon	The lunar gravity is only about one-sixth that on earth so an astronaut can carry a heavier bag.	The lunar force is only about one-sixth that on earth so an astronaut can carry a heavier bag.
59	herb	medicine	The essence drawn from this herb is good for high blood pressure.	The essence drawn from this shrub is good for high blood pressure.
60	hostage	plane	The hijackers held two passengers hostage to get the pilot to change the flight path.	The hijackers held two passengers prisoner to get the pilot to change the flight path.
62	infection	amputation	The infection spread up to his knee and they had to remove his leg.	The virus spread up to his knee and they had to remove his leg.
63	insurance	coverage	Check that your insurance policy includes breakages and damage during removals.	Check that your protection policy includes breakages and damage during removals.
64	intensity	stress	The intensity of the work made me frustrated and anxious.	The magnitude of the work made me frustrated and anxious.
65	inventor	market	The inventor was trying to improve product sales by appealing to businesses.	The entrepreneur was trying to improve product sales by appealing to businesses.
67	lamp	burn	She lit the oil lamp and watched the small flame grow.	She lit the oil lantern and watched the small flame grow.
71	lizard	tongue	This lizard catches insects in the same way frogs do.	This chameleon catches insects in the same way frogs do.
72	lobster	broil	Anton the chef concocts a sensual coupling of lobster and asparagus.	Anton the chef concocts a sensual coupling of crab and asparagus.
74	luggage	heavy	It was lucky someone helped her lift her luggage onto the train because it was full of books.	It was lucky someone helped her lift her suitcase onto the train because it was full of books.
80	museum	artifacts	The museum has many historic cultural relics.	The exhibition has many historic cultural relics.
81	onion	spice	To make the food taste just as he remembered he added onion and coriander.	To make the food taste just as he remembered he added garlic and coriander.
85	poison	spider	He applied pressure to the bite wound on his arm and hoped there was no poison.	He applied pressure to the bite wound on his arm and hoped there were no toxins.
86	potato	mash	She made sure there were no lumps in the potato before she fed it to her baby.	She made sure there were no lumps in the vegetables before she fed it to her baby.
88	predator	shark	The divers knew that they had no hope when they saw the deadliest predator in the ocean.	The divers knew that they had no hope when they saw the deadliest carnivore in the ocean.
90	quilt	sew	Her grandmother put together a quilt for her birthday with lots of different fabrics and threads.	Her grandmother put together a blanket for her birthday with lots of different fabrics and threads.
92	receipt	bill	After she paid for their meal she put the receipt in her handbag.	After she paid for their meal she put the coupon in her handbag.
93	stamp	collection	He has five albums of stamps, many of which are very rare.	He has five albums of coins, many of which are very rare.
96	ticket	sport	I intend to watch all of my team's games this season because I got a season ticket for my birthday.	I intend to watch all of my team's games this season because I got a season pass for my birthday.
98	tray	ice	She poured water into each cube and then put the tray in the freezer.	She poured water into each cube and then put the platter in the freezer.

(*continued on next page*)

(*continued*)

Item	Target	Probe	Probe-consistent sentence	Target-replaced sentence
99	tulip	holland	They plan to visit the famous tulip kingdom.	They plan to visit the famous orchid kingdom.
101	tunnel	train	The railway company spent a lot of money constructing a tunnel through the mountain and laying new tracks.	The railway company spent a lot of money constructing a passage through the mountain and laying new tracks.
104	vacancy	apartment	They wanted somewhere bigger to live so when they heard about a vacancy in their building they jumped at the chance.	They wanted somewhere bigger to live so when they heard about a tenancy in their building they jumped at the chance.
105	vase	expensive	The delicate vase was made by a famous artist and is sure to sell for a high price at auction.	The delicate pottery was made by a famous artist and is sure to sell for a high price at auction.
107	victim	murder	The victim was smothered to death with a pillow.	The target was smothered to death with a pillow.
108	vine	climb	The vine grew up the side of their house, clinging to the wall.	The plant grew up the side of their house, clinging to the wall.
109	violence	gang	Several suburbs of the city were blighted by organized violence from rival groups.	Several suburbs of the city were blighted by organized crime from rival groups.
110	visitor	tourist	As first-time visitors in New York they wanted to see the Statue of Liberty and a show on Broadway.	As first-time travellers in New York they wanted to see the Statue of Liberty and a show on Broadway.
112	wand	magician	He tapped his wand on the top hat and out hopped a rabbit.	He tapped his sceptre on the top hat and out hopped a rabbit.
115	weapon	war	They considered using their nuclear arsenal as their ultimate weapon in the hope that the opposition would surrender.	They considered using their nuclear arsenal as their ultimate defense in the hope that the opposition would surrender.
117	whiskey	bottle	He wanted to buy a nice whiskey for his boss but it was hard to choose from all the different shapes and sizes.	He wanted to buy a nice scotch for his boss but it was hard to choose from all the different shapes and sizes.
120	wire	electric	All the wires in the house need to be replaced, so we can't use the plug sockets or lights.	All the cables in the house need to be replaced, so we can't use the plug sockets or lights.
121	wolf	howl	The lone wolf tilted its head to the sky and called out mournfully for the rest of the pack.	The lone canine tilted its head to the sky and called out mournfully for the rest of the pack.
123	worm	fishing	He hooked on the worm as bait and threw out the line in hopes he'd get a catch.	He hooked on the maggot as bait and threw out the line in hopes he'd get a catch.
124	yogurt	calories	Low fat yogurt can be good for losing weight but you also need to look at sugar content.	Low fat milk can be good for losing weight but you also need to look at sugar content.
125	zone	parking	You can leave your car in the long-stay zone over there.	You can leave your car in the long-stay area over there.
126	fuel	pollution	The corporation was renowned for burning oil and coal for fuel.	The corporation was renowned for burning oil and coal for energy.
127	reflection	mirror	She walked into the changing room to look at her reflection.	She walked into the changing room to look at her appearance.
128	argument	anger	Their argument became very heated and they had to be restrained.	Their altercation became very heated and they had to be restrained.
129	needle	injection	The doctor inserted the needle to extract some blood.	The doctor inserted the syringe to extract some blood.
130	professor	expert	The professor knew almost everything there was to know about her field of interest.	The lecturer knew almost everything there was to know about her field of interest.
131	muscle	strength	He has a lot of muscle in his upper body which allows him to lift very heavy weights.	He has a lot of brawn in his upper body which allows him to lift very heavy weights.
132	relief	charity	The philanthropic organisation provides relief to poverty-stricken nations.	The philanthropic organisation provides support to poverty-stricken nations.

Experiment 3 - Underlined words highlight the semantic replacement of the target word.

Item	Target	Probe	Probe-consistent sentence	Target-replaced sentence
1	acorn	squirrel	The bushy-tailed rodent buried the acorn for consumption later.	The bushy-tailed rodent buried the hazelnut for consumption later.
2	acre	yard	The garden at the new house is nearly an acre, even though it's in the city.	The garden at the new house is nearly a hectare, even though it's in the city.
3	adventure	television	There was an adventure programme airing on the BBC.	There was a fantasy programme airing on the BBC.
4	advertisement	agency	The firm offered advertisement services to other companies.	The firm offered publicity services to other companies.
5	ambulance	rush	The ambulance speeded to the scene of the accident; there was little time to spare.	The paramedics speeded to the scene of the accident; there was little time to spare.
6	ancestor	apes	Humans diverged from their closest evolutionary ancestor, the chimpanzee, six million years ago.	Humans diverged from their closest evolutionary cousin, the chimpanzee, six million years ago.
7	ankle	bracelet	She wore a band of jewellery around her ankle.	She wore a band of jewellery around her leg.
9	athlete	sneaker	The athlete was sponsored by a major brand and could only compete while wearing that brand's shoes.	The competitor was sponsored by a major brand and could only compete while wearing that brand's shoes.
10	backpack	book	His backpack was filled with novels, a dictionary and several other texts.	His bag was filled with novels, a dictionary and several other texts.
11	balloon	rubber	Balloons are made from an elastic gum derived from a latex tree.	Inflatables are made from an elastic gum derived from a latex tree.
12	bathtub	slip	The old man fell while getting out of the bathtub.	The old man fell while getting out of the shower.
13	beard	santa	He works part time in the winter, giving gifts to children while wearing a long white beard and red coat.	He works part time in the winter, giving gifts to children while wearing a long white moustache and red coat.
14	bicycle	exercise	Riding a bicycle is a good way to keep aerobically fit.	Riding a crosstrainer is a good way to keep aerobically fit.
15	biologist	frog	The biologist was studying how tadpoles develop into mature amphibians.	The scientist was studying how tadpoles develop into mature amphibians.
16	bone	white	Bone is the same colour as chalk or snow.	Skeletons are the same colour as chalk or snow.
21	cage	metal	The cage and its bars were constructed out of wrought iron.	The cell and its bars were constructed out of wrought iron.
23	castle	old	The castle was ancient and could be dated back to the 10th century.	The fort was ancient and could be dated back to the 10th century.
26	chimney	Christmas	He jumped down the chimney with a sack full of presents.	He jumped down the fireplace with a sack full of presents.
28	coal	diamond	It takes millions of years of heat and pressure for coal to be transformed into a precious stone.	It takes millions of years of heat and pressure for carbon to be transformed into a precious stone.
31	cupboard	food	The cupboard was filled with fruits, vegetables, bread and oats.	The pantry was filled with fruits, vegetables, bread and oats.
32	dancer	tiny	The dancer was very small for her age.	The performer was very small for her age.
35	dessert	sugar	The dessert was incredibly sweet and was likely packed with carbohydrates.	The pudding was incredibly sweet and was likely packed with carbohydrates.

(*continued on next page*)

(*continued*)

Item	Target	Probe	Probe-consistent sentence	Target-replaced sentence
36	device	electric	The device can be powered by plugging it into a mains socket.	The gadget can be powered by plugging it into a mains socket.
41	dragon	scales	The dragon's body was covered in hard, rigid, platelike structures.	The hydra's body was covered in hard, rigid, platelike structures.
46	eraser	delete	The eraser can be used to edit out any information which isn't necessary.	The rubber can be used to edit out any information which isn't necessary.
47	eyebrows	expression	Her eyebrows were arched in supercilious surprise when hearing the news.	Her face was arched in supercilious surprise when hearing the news.
48	fangs	bat	The fangs of this nocturnal flying mammal allow it to drink blood.	The canines of this nocturnal flying mammal allow it to drink blood.
49	feather	weight	Boxers are split into classes based on how heavy they are, John was in the feather class.	Boxers are split into classes based on how heavy they are, John was in the light class.
50	flask	chemistry	The graduations are marked on the side of the flask so you can know the precise amount of saline solution.	The graduations are marked on the side of the bottle so you can know the precise amount of saline solution.
51	foam	mousse	The foam can be applied to damp hair to aid in styling.	The cream can be applied to damp hair to aid in styling.
52	geometry	analytical	The pilot used geometry to systematically work out the safest way to land.	The pilot used statistics to systematically work out the safest way to land.
58	gravity	density	The force of gravity is the same regardless of whether matter is tightly packed together or widely dispersed.	The force of attraction is the same regardless of whether matter is tightly packed together or widely dispersed.
59	herb	marijuana	The herb can be smoked to produce psychedelic effects.	The shrub can be smoked to produce psychedelic effects.
60	hostage	takeover	The terrorists took control of the area and held the residents hostage.	The terrorists took control of the area and held the residents prisoner.
62	infection	ear	She couldn't hear much, as the infection had led to a build up of wax.	She couldn't hear much, as the inflammation had led to a build up of wax.
63	insurance	health	You must get insurance to cover the cost of your hospital bills.	You must get protection to cover the cost of your hospital bills.
64	intensity	sex	The intensity in the room increased as they began to make love.	The emotion in the room increased as they began to make love.
65	inventor	market	The inventor was trying to improve product sales by appealing to businesses.	The entrepreneur was trying to improve product sales by appealing to businesses.
67	lamp	desk	He sat at his workstation and switched on the lamp.	He sat at his workstation and switched on the light.
71	lizard	small	Lizards can be tiny, some can even fit on the end of a fingertip.	Reptiles can be tiny, some can even fit on the end of a fingertip.
72	lobster	steak	The surf and turf came with lobster and a cut of beef.	The surf and turf came with crab and a cut of beef.
74	luggage	moving	Her luggage had been collected and was currently in transit.	Her suitcase had been collected and was currently in transit.
80	museum	dinosaur	There was a huge Tyrannosaurus skeleton in the museum's foyer.	There was a huge Tyrannosaurus skeleton in the building's foyer.
81	onion	spice	To make the food taste just as he remembered he added onion and coriander.	To make the food taste just as he remembered he added garlic and coriander.
85	poison	acid	The poison has a low pH and can easily burn through certain materials.	The venom has a low pH and can easily burn through certain materials.
86	potato	mash	She made sure there were no lumps in the potato before she fed it to her baby.	She made sure there were no lumps in the vegetables before she fed it to her baby.
90	quilt	grandma	The old woman knitted a quilt for her grandchild.	The old woman knitted a blanket for her grandchild.
92	receipt	tax	He opened the receipt to find that the government's deduction from his income was quite substantial.	He opened the payslip to find that the government's deduction from his income was quite substantial.
93	stamp	approval	She put a stamp on the document to confirm that the plans could go ahead.	She put an emblem on the document to confirm that the plans could go ahead.
96	ticket	bill	The ticket outlined the cost of the meal and drinks they had ordered.	The check outlined the cost of the meal and drinks they had ordered.
98	tray	silver	The tray was made out of a shiny grey precious metal.	The plate was made out of a shiny grey precious metal.
99	tulip	holland	They plan to visit the famous tulip kingdom.	They plan to visit the famous orchid kingdom.
101	tunnel	mountain	In order to avoid the high peaks, they built a tunnel by boring a hole through the rock.	In order to avoid the high peaks, they built an underpass by boring a hole through the rock.
104	vacancy	apartment	They wanted somewhere bigger to live so when they heard about a vacancy in their building they jumped at the chance.	They wanted somewhere bigger to live so when they heard about a tenancy in their building they jumped at the chance.
105	vase	ceramic	A vase is most commonly made from materials derived from clay, such as porcelain or pottery.	A flowerpot is most commonly made from materials derived from clay, such as porcelain or pottery.
107	victim	sad	After the incident, the victim was mournful and heartbroken.	After the incident, the subject was mournful and heartbroken.
108	vine	climb	The vine grew up the side of their house, clinging to the wall.	The plant grew up the side of their house, clinging to the wall.
109	violence	peace	The countries signed a treaty, agreeing to end the violence and abstain from future conflict.	The countries signed a treaty, agreeing to end the bloodshed and abstain from future conflict.
110	visitor	company	The firm was expecting a visitor who would potentially invest.	The firm was expecting a guest who would potentially invest.
112	wand	baton	The conductor held a wand-like stick in his hand and used it to direct the orchestra.	The conductor held a rod-like stick in his hand and used it to direct the orchestra.
115	weapon	knife	The weapon was a 6 inch blade with a very sharp edge.	The instrument was a 6 inch blade with a very sharp edge.
117	whiskey	sour	The cocktail was made by mixing whiskey, lemon juice, sugar and egg white.	The cocktail was made by mixing scotch, lemon juice, sugar and egg white.
120	wire	braces	The orthodontist attached wires to her teeth in an attempt to straighten them.	The orthodontist attached brackets to her teeth in an attempt to straighten them.
121	wolf	teeth	A wolf has 12 incisors, 4 canines and 26 molars	A dog has 12 incisors, 4 canines and 26 molars
123	worm	earth	Worms can be found digging their way through dirt and soil.	Grubs can be found digging their way through dirt and soil.
124	yogurt	calories	Low fat yogurt can be good for losing weight but you also need to look at sugar content.	Low fat milk can be good for losing weight but you also need to look at sugar content.
125	zone	space	A supercluster is a zone in the universe which has a large number of galaxies.	A supercluster is an area in the universe which has a large number of galaxies.
126	fuel	power	The fuel provides the city with electricity.	The gas provides the city with electricity.
128	argument	discussion	Their deliberation was very productive with both sides putting forward some strong arguments.	Their deliberation was very productive with both sides putting forward some strong suggestions.
129	needle	haystack	He was looking for a needle in a pile of straw.	He was looking for a syringe in a pile of straw.
130	professor	doctor	The professor was a physician who also worked in hospitals.	The lecturer was a physician who also worked in hospitals.
132	relief	fund	The subsidy had been set up to finance the relief effort following the disaster.	The subsidy had been set up to finance the revival effort following the disaster.

Filler sentences – The same filler sentences were used in all three experiments.

Item	Sentence
1	He tried to take a photo of the waterfall as the sun cast a rainbow through the spray.
2	She could barely see her hand stretched out in front of her through the fog.
3	The feast was laid out in the grandest room of the palace.
4	Beans and pulses are good sources of protein and are high in fibre.
5	The judge told the murderer that he was a monster with no remorse.
6	He rubbed the chair down with sandpaper so the paint would go on evenly.
7	The child's favorite TV show character was a friendly octopus.
8	She hand washed the delicates in cold water with a mild soap.
9	Her French vocabulary was limited so she often needed to look up the meanings of words.
10	They tip-toed into the dark cave and jumped when something flew out over their heads.
11	She discovered how clay beds, which had existed for millions of years, were formed.
12	He watched the children putting their favorite decorations on the tree while his wife prepared the roast dinner.
13	Today, his homework is to calculate the circumference given the diameter.
14	She felt uncomfortable and sneezed after she stepped into the room covered with dust.
15	Some employers are in trouble for paying a lower wage than is legally required.
16	The eagle collected many twigs over the highest branch of the big tree to prepare for a comfortable and safe home.
17	The electrician used some sophisticated electronic tools and set up the circuits for the new house in a matter of days.
18	A full-grown elephant can weigh over 6000 kg.
19	Germs are invisible to the naked eye but can be observed in the lab.
20	He was shaking and his eyes were wide as if he had seen a ghost.
21	To secretly open the envelope she held it over a boiled kettle to loosen the glue.
22	Bob was worried about his heart, and he altered his lifestyle accordingly.
23	They crowded around the screen to watch the missile fired into space.
24	He could feel the tension in his muscles from sitting with poor posture all day.
25	The city threw a parade to welcome in the New Year.
26	The glittering necklace sold for so much because of the perfectly clear, flawless diamond at its center.
27	Penguins rock from side-to-side when they walk because their legs are so short.
28	The comet moves around the sun on a long oval course over a period of 76 years.
29	Dinosaurs first appeared during the Triassic period, between 243 and 233 million years ago.
30	As editor of the daily paper he assigns reporters to investigate interesting events.
31	The small denizens of the wilderness hardly took pains to move out of her path.
32	They were worried about the consequences of smoking so much on his lung health.

Appendix B. Exploratory analyses

An item's preferred interpretation was assessed by calculating the proportion of participant responses in the unprimed condition of the associate production task that were judged to favour the probe-consistent context, referred to here as the unprimed score. A low unprimed score for an item in associate production would indicate that the item's probe-consistent context likely refers to an aspect of the target word's meaning that is, generally speaking, a less preferred interpretation of that word (as participants referred to it in a minority of cases). We used this unprimed score as our predictor to assess its relationship with priming strength. For each item we calculated the z-score of its probe-consistent priming effect for each of the three dependent variables (with positive values indicating stronger priming). These three z-scores were then averaged to give an indication of overall priming strength for each item. As well as using the unprimed score as a predictor of priming strength, we also used the item average of the unprimed score and the same score in the probe-consistent (i.e., primed) condition. This latter measure is more (although perhaps overly) conservative as it guards against any influence of regression to the mean. The results of this analysis are summarised in Fig. 9.

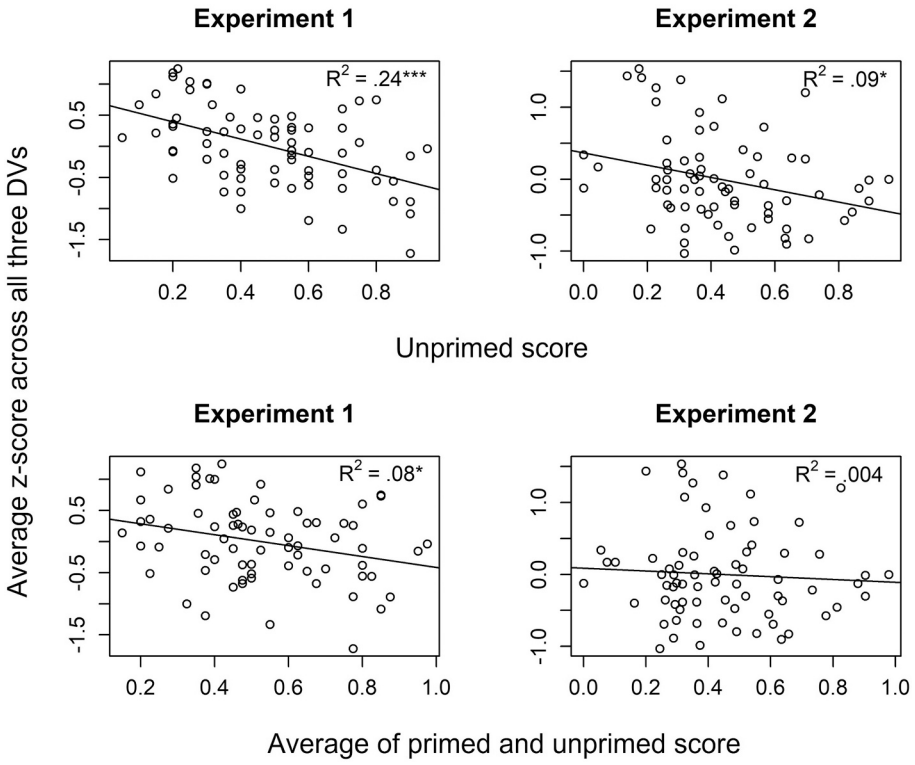

Fig. 9. The relationship between an item's average z-score across all three dependent variables and an item's unprimed score (top row) or the average of an item's primed and unprimed score (bottom row). An item's primed or unprimed score is the proportion of participants who produced an associate which favoured the probe-consistent context in that condition in the associate production task. "***" < 0.001, "**" < 0.01, "*" < 0.05.

References

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review, 126*(6), 817–840. https://doi.org/10.1037/rev0000154

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*(4), 1407–1425.

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 273–278). Austin, TX: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2016). *Semantic ambiguity effects in lexical processing: A neural-network account based on semantic settling dynamics (unpublished manuscript)*. Pittsburgh, PA: Department of Psychology, Carnegie Mellon University.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*(5), 1059–1082.

Blank, I. A., & Fedorenko, E. (2020). No evidence for differences among language regions in their temporal receptive windows. *NeuroImage, 219*, Article 116925. https://doi.org/10.1016/j.neuroimage.2020.116925

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development, 8*(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Brown-Schmidt, S., Cho, S.-J., Nozari, N., Klooster, N., & Duff, M. (2021). The limited role of hippocampal declarative memory in transient semantic activation during online language processing. *Neuropsychologia, 152*, Article 107730. https://doi.org/10.1016/j.neuropsychologia.2020.107730

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 9. https://doi.org/10.5334/joc.10

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

Covington, N. V., Kurczek, J., Duff, M. C., & Brown-Schmidt, S. (2020). The effect of repetition on pronoun resolution in patients with memory impairment. *Journal of Clinical and Experimental Neuropsychology, 42*(2), 171–184. https://doi.org/10.1080/13803395.2019.1699503

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Duff, M., & 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., & Brown-Schmidt, S. (2017). Hippocampal contributions to language use and processing. In *The Hippocampus from cells to systems* (pp. 503–536). Springer.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language, 102*, 130–141. https://doi.org/10.1016/j.jml.2018.05.008

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

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

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., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2021). The relationship between sentence comprehension and lexical-semantic retuning. *Journal of Memory and Language, 116*, Article 104188. https://doi.org/10.1016/j.jml.2020.104188

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163–189. https://doi.org/10.1146/annurev.psych.48.1.163

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one Person's mistake is Another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

van Heuven, W. J. B., 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. https://doi.org/10.1080/17470218.2013.850521

Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *The Journal of Neuroscience, 32*(17), 5795–5803. https://doi.org/10.1523/JNEUROSCI.6294-11.2012

Hoedemaker, R. S., & Gordon, P. C. (2017). The onset and time course of semantic priming during rapid recognition of visual words. *Journal of Experimental Psychology: Human Perception and Performance, 43*(5), 881–902. https://doi.org/10.1037/xhp0000377

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2021). Semantic diversity is best measured with unscaled vectors: Reply to Cevoli, Watkins and Rastle (2020). *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01693-4

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language, 103*, 114–126. https://doi.org/10.1016/j.jml.2018.08.005

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language, 44*(1), 32–51. https://doi.org/10.1006/jmla.2000.2737

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin and Review, 23*(4), 1214–1220. https://doi.org/10.3758/s13423-015-0980-7

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology, 66*(2), 115–124. https://doi.org/10.1037/a0026727

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1051–1062. https://doi.org/10.1037/0278-7393.20.5.1051

Kapnoula, E. C., & Samuel, A. G. (2019). Voices in the mental lexicon: Words carry indexical information that can affect access to their meaning. *Journal of Memory and Language, 107*, 111–127.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295x.95.2.163

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*(2), 133–159. https://doi.org/10.1016/0749-596X(90)90069-C

Kurczek, J., Brown-Schmidt, S., & Duff, M. (2013). Hippocampal contributions to language: Evidence of referential processing deficits in amnesia. *Journal of Experimental Psychology: General, 142*(4), 1346–1354. https://doi.org/10.1037/a0034026

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2022). *The retention and transformation of discourse memory across sleep and wake*.

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

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 and Review, 27*(5), 1059–1069. https://doi.org/10.3758/s13423-020-01773-0

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*(3), 295–322. https://doi.org/10.1177/00238309010440030101

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York, NY: Psychology Press.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation/.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238. https://doi.org/10.1016/s0010-0285(03)00006-9

Parks, R., Ray, J., & Bland, S. (1998). Wordsmyth English dictionary—Thesaurus [ONLINE]. Available: http://www.wordsmyth.net/.

Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology, 70*, 1–30. https://doi.org/10.1016/j.cogpsych.2014.01.001

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7*(5), 187–189. https://doi.org/10.1016/s1364-6613(03)00092-5

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

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

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*(2), 245–266. https://doi.org/10.1006/jmla.2001.2810

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., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89–104. https://doi.org/10.1207/s15516709cog2801_4

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

Sharkey, N. E. (1990). A connectionist model of text comprehension. In *Comprehension Processes in Reading*. Routledge.

Shtyrov, Y., Nikulin, V. V., & Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *Journal of Neuroscience, 30*(50), 16864–16867. https://doi.org/10.1523/JNEUROSCI.1376-10.2010

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27*, 324–340. https://doi.org/10.1016/0749-596X(88)90058-7

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. https://doi.org/10.3758/BF03202766

Wang, H. C., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review, 24*(4), 1186–1193. https://doi.org/10.3758/s13423-016-1182-7

Was, C., Woltz, D., & Hirsch, D. (2019). Memory processes underlying long-term semantic priming. *Memory & Cognition, 47*(2), 313–325. https://doi.org/10.3758/s13421-018-0867-8

Witzel, N. O., & Forster, K. I. (2012). How L2 words are stored: The episodic L2 hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1608–1621. https://doi.org/10.1037/a0028072

Woltz, D. J., Sorensen, L. J., Indahl, T. C., & Splinter, A. F. (2015). Long-term semantic priming of propositions representing general knowledge. *Journal of Memory and Language, 79–80*, 30–52. https://doi.org/10.1016/j.jml.2014.11.002

Yonelinas, A., Ranganath, C., Ekstrom, A., & Wiltgen, B. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews. Neuroscience, 20*(6), 364–375. https://doi.org/10.1038/s41583-019-0150-4

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*(2), 307–327. https://doi.org/10.1037/a0015305

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273