Journal of Experimental Psychology: General

Episodic Memory and Sleep Are Involved in the Maintenance of Context-Specific Lexical Information

Matthew H. C. Mak, Adam J. Curtis, Jennifer M. Rodd, and M. Gareth Gaskell

Online First Publication, June 26, 2023. https://dx.doi.org/10.1037/xge0001435

CITATION

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023, June 26). Episodic Memory and Sleep Are Involved in the Maintenance of Context-Specific Lexical Information. *Journal of Experimental Psychology: General*. Advance online publication. https://dx.doi.org/10.1037/xge0001435

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001435

Episodic Memory and Sleep Are Involved in the Maintenance of Context-Specific Lexical Information

Matthew H. C. Mak[1], Adam J. Curtis[1], Jennifer M. Rodd[2], and M. Gareth Gaskell[1]
[1] Department of Psychology, University of York
[2] Department of Experimental Psychology, University College London

Familiar words come with a wealth of associated knowledge about their variety of usage, accumulated over a lifetime. How do we track and adjust this knowledge as new instances of a word are encountered? A recent study (*Cognition*) found that, for homonyms (e.g., *bank*), sleep-associated consolidation facilitates the updating of meaning dominance. Here, we tested the generality of this finding by exposing participants to (Experiment 1; $N = 125$) nonhomonyms (e.g., *bathtub*) in sentences that biased their meanings toward a specific interpretation (e.g., *bathtub-slip* vs. *bathtub-relax*), and (Experiment 2; $N = 128$) word-class ambiguous words (e.g., *loan*) in sentences where the words were used in their dispreferred word class (e.g., "He will loan me money"). Both experiments showed that such sentential experience influenced later interpretation and usage of the words more after a night's sleep than a day awake. We interpret these results as evidence for a general role of episodic memory in language comprehension such that new episodic memories are formed every time a sentence is comprehended, and these memories contribute to lexical processing next time the word is encountered, as well as potentially to the fine-tuning of long-term lexical knowledge.

Public Significance Statement

Our research adds to the growing body of evidence that language and memory, which have historically been studied as distinct cognitive abilities, are heavily intertwined. We tested the notion that episodic memory—memories for specific events—is involved in the maintenance of discourse representations during language comprehension. We showed that these representations, presumably binding words and concepts together, capture context-specific lexical information such as the precise meaning (e.g., relaxing vs. slipping in a bathtub) and their word class (e.g., *loan* as a noun vs. verb). We showed that (a) these representations can prime subsequent lexical processing, and (b) as in other newly acquired episodic memories, these representations may be prone to sleep-related memory consolidation such that their effects on lexical processing was more robust after sleep (vs. wakefulness). Our findings highlighted how episodic memory and sleep may contribute to the updating of lexical memory, providing some degree of malleability to our mental lexicon.

Keywords: episodic memory, sleep, language comprehension, nonhomonym, word class

How a word is interpreted in natural language is almost always context-specific. Consider the English word *stamp*, for example. In "He has five albums of stamps, many of which are very rare," the word refers to collectible stamps, but in "It's best to send your letters first class, so you will need a stamp," it refers to standard stamps for postage. Although the core meaning of *stamp* is identical in these instances, its precise meaning is somewhat different. This means that during comprehension, language users need to make use of the linguistic context to arrive at the precise meaning of any given word (Saussure, 1916). In addition to this, the word class of many English words also varies depending on context. For instance, *stamp* is word-class ambiguous, meaning that it can serve as either a noun or a verb without undergoing any morphological changes. Together, these examples indicate that the linguistic contexts in

Matthew H. C. Mak https://orcid.org/0000-0001-7237-4931

This research was supported by an Economic and Social Research Council Grant (ES/T008571/1) to Gareth Gaskell and Jenni Rodd. We have no conflict of interest to disclose.

All the materials, data, and R scripts are publicly available at https://osf.io/z2ane/?view_only=ac75981a9ac3403b8e45c841a4f31ac5.

Correspondence concerning this article should be addressed to Matthew H. C. Mak and M. Gareth Gaskell, Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom. Email: matthew.mak@york.ac.uk or gareth.gaskell@york.ac.uk

which a word appears provide crucial information about how the word should be processed and interpreted during online comprehension.

How a word is interpreted online is affected by its sentential context (e.g., Borovsky et al., 2012; Kutas et al., 2011). In addition, the specific context in which a word occurs can often affect interpretation next time the word is encountered (see Nation, 2017; Rodd, 2020). One strand of evidence supporting this comes from studies with homonyms, whose meaning is entirely context-dependent (see Rodd, 2020 for a review). These kinds of words have two or more distinct meanings (e.g., *bank*: financial bank vs. river bank), but usually, one of them is more frequent or dominant. Such relative dominance is accrued from experience with natural language and is internalized in the mental lexicon, affecting subsequent comprehension and production. For instance, if English speakers are asked to give the first word that comes to mind upon seeing *bank*, they are more likely to respond with associates related to a financial institute (the dominant meaning) than associates related to rivers (the subordinate meaning; Twilley et al., 1994; see also Gilbert & Rodd, 2022). Importantly, however, this preference for the dominant meaning is not set in stone and is influenced by recent linguistic experience. In Rodd et al. (2013), participants were exposed to homonyms in a disambiguating sentence that primed interpretation toward the homonyms' subordinate meanings (e.g., "The seal came up onto the bank of the river"). In a subsequent associate production task ∼20 min after sentence exposure, participants were more likely to respond to these homonyms by giving associates that were related to their subordinate meanings (as compared with a matched set of homonyms that were unprimed). This finding, referred to as the *word-meaning priming effect*, suggests that recent encounters with a context-specific meaning have consequences for how the word is subsequently interpreted. Notably, this priming effect tends to decay over time (Rodd et al., 2016).

The word-meaning priming effect was originally explained with reference to an *immediate alteration account* (Rodd et al., 2013, 2016), which posits that the word form of a homonym is associated with two or more distinct meanings via weighted connections in long-term semantic memory, with the more dominant meanings having stronger weightings. Recent exposure to a sentential context that primes the subordinate meaning of a homonym will at least temporarily increase the weighting to this context-specific meaning, making it a little easier to access than before. A key tenet of this account is that language exposure can immediately and directly alter established weightings in long-term memory. However, emerging evidence suggests that this is not necessarily the case and that language exposure may result in a new and temporary episodic representation that can influence subsequent interpretations alongside established semantic knowledge.

Using the same kinds of materials and tasks as Rodd et al. (2013, 2016), Gaskell et al. (2019; Experiment 1) tested participants both shortly after sentence exposure (8 min) and after a 2- or 12-hr delay spent awake or asleep. When tested after a brief delay of 8 min, there was a clear word-meaning priming effect across participants, replicating Rodd et al.'s (2013, 2016) original finding. However, after 2 or 12 hr, word-meaning priming was only found for participants who had a sleep opportunity during the delay (vs. those who stayed awake). Potentially, this suggests that sleep may be involved in the maintenance of word-meaning priming via offline consolidation, whereby newly acquired hippocampus-dependent memory progressively integrates into the neocortex for long-term storage (e.g., see Inostroza & Born, 2013; Paller et al., 2021; Rasch & Born, 2013 for reviews). Alternatively, word-meaning priming might be more likely to maintain over sleep than over wakefulness because there is limited interference from sensory and linguistic inputs during sleep (Yonelinas et al., 2019). To address the latter possibility, Gaskell et al. conducted a follow-up experiment with a 24-hr delay. Here, some participants spent the first half of the 24 hr asleep and the second half awake (sleep–wake) while the rest of the participants did the opposite (wake–sleep). This way, both groups would have more or less the same amount of wakeful interference. Interestingly, 24 hr after sentence exposure, participants in the sleep–wake group (vs. wake–sleep) showed a stronger word-meaning priming effect. This was interpreted as indicating that sleep did not simply provide temporary protection of word-meaning priming from interference, but instead, it may have played an active role in consolidating the sentential context, leading to priming that remained robust over subsequent wake periods. These sleep-related findings from Gaskell et al. pose a challenge to the immediate alteration account, which posits direct alterations to established weightings in long-term memory. For this account, there is little reason to suspect that sleep-related consolidation is relevant, because adjustment of long-term lexical knowledge has already taken place in the neocortex prior to sleep, and there is no "new" memory to consolidate (Gaskell et al., 2019). An implication, then, is that word-meaning priming effects in homonyms may involve some kind of new, possibly hippocampal, memory that is susceptible to sleep-associated consolidation.

To accommodate the finding that word-meaning priming is influenced by sleep, Gaskell et al. (2019) proposed an *episodic context account*,[1] which hypothesizes that episodic memory and subsequent sleep-related consolidation contribute to both the online processing and longer-term retention of context-specific information (see also Duff & Brown-Schmidt, 2012, 2017). Regarding online processing, this account proposes that when a homonym is encountered in a sentential context, rather than its form-meaning weightings in long-term memory being altered, a new but temporary episodic trace for the comprehension episode is formed. This representation, binding the homonym and the surrounding words/concepts together, is presumed to facilitate comprehension online. Importantly, however, when the homonym is reencountered (e.g., in associate production), the context-specific representation formed earlier may provide an additional source of information—on top of the established weightings in long-term memory—to guide interpretation of the homonym, potentially biasing language users toward the prior context-specific meaning.

Next, regarding longer-term retention, the episodic context account proposes that the contextually bound episodic representations generated during language comprehension would be better maintained if a sleep opportunity is afforded after comprehension.

[1] This account was originally called the *contextual binding* account (Gaskell et al., 2019). However, to avoid confusion with Yonelinas et al. (2019), it was renamed *episodic context account* in a subsequent paper (Curtis et al., 2022).

A wealth of psychological and neuroscientific evidence suggests that the encoding of episodic memories is supported by the hippocampus; over time, these hippocampal traces may be consolidated into neocortical networks for long-term storage, thereby increasing their accessibility and resistance to decay (e.g., McClelland et al., 1995; Moscovitch et al., 2016). Some theories argue that sleep is critically involved in the consolidation process such that sleep actively facilitates the integration of hippocampus-dependent memories into the neocortex (e.g., Davis & Gaskell, 2009; Klinzing et al., 2019; Lewis & Durrant, 2011; McClelland, 2013; Paller et al., 2021; Rasch & Born, 2013; Stickgold, 2005). Alternatively, sleep may simply represent an optimal time window for consolidation, because the processing of external information is greatly reduced during sleep (e.g., Paller et al., 2021; Siegel, 2021). Regardless of the precise mechanism, if sleep occurs soon after experiencing homonyms in a sentential context, the associated context-specific representations should become partially consolidated into the neocortex; as a result, these representations would be less prone to decay and thus more able to bias subsequent interpretation of the homonyms. In contrast, if a sleep opportunity is not provided soon after exposure, the associated context-specific representations may be more susceptible to hippocampal decay, reducing its likelihood of influencing the subsequent interpretation of the homonyms and hence the emergence of a word-meaning priming effect.

Note that the episodic context account emphasizes a general role of episodic memory in language comprehension, so the predictions that it makes are of a substantially larger scope than the immediate alteration account, which is primarily concerned with the balance between different meanings of homonyms. For instance, the episodic context account hypothesizes that during language comprehension, a context-specific representation is formed regardless of whether a sentence contains homonyms, potentially to play a more general role in supporting the construction and maintenance of situation models and/or the retention of discourse memory (Altmann & Ekves, 2019; Graesser et al., 1997; van Dijk & Kintsch, 1983; Zacks et al., 2007, 2009; Zwaan et al., 1995). One of the key predictions from the episodic context account is that if language comprehension leads to a context-specific representation in episodic memory, then word-meaning priming should not be restricted to homonyms; instead, it should be pervasive and observed for *any* word as long as its surrounding contexts refine its interpretation in some way. Returning to the *stamp* example at the beginning: Depending on the context, the word can refer to collectible stamps or standard postage stamps. The episodic context account, therefore, predicts that such context-specific information is captured by episodic memory during comprehension, and therefore, has the potential to influence processing the next time the word is reencountered. A recent study provides preliminary empirical support for this hypothesis.

Curtis et al. (2022) made use of a comparable design to Rodd et al. (2013), but instead of homonyms, the targets were nonhomonymic nouns, which by dictionary definition, only have one meaning (e.g., *bathtub*). Each of these nonhomonymic targets was paired with a probe word that was semantically related to a specific aspect of the targets' meaning (e.g., *bathtub-slip* vs. *bathtub-relax*). Based on this probe, a prime sentence was created such that it biased the interpretation of the target toward the probe (e.g., *The old man fell while getting out of the bathtub*). Note that the probe word never appeared in the sentence and thus was never experienced together with the nonhomonymic target. To test whether exposure to these sentences subsequently biased participants toward the primed semantic aspect (e.g., *slip*), participants completed two tasks shortly (10–30 min) after sentence exposure: (a) speeded relatedness judgment, where participants decided if a target–probe pair (e.g., *bathtub-slip*) was related in meaning, and (b) associate production, where participants gave the first word that came to mind upon seeing a target in isolation. Across three experiments, both tasks revealed compelling evidence for word-meaning priming for the nonhomonymic targets, such that prior sentential exposure to a nonhomonym biased participants to interpret the word in a way that was consistent with that specific sentential context. This provides support for the notion that during language comprehension, a context-specific representation (for the sentence) is generated, which may, in turn, influence future lexical processing.

The findings from Curtis et al. (2022) are predicted by, and in line with, the episodic context account; however, alternative explanations are possible. For instance, one could argue that since the nonhomonymic targets and their probes are semantically related in the first place (e.g., *bathtub-slip*), the subsequent priming effects may not reflect episodic involvement but changes in weights in long-term semantic memory (akin to the immediate alternation account for homonyms). By this kind of alternative account, the generation of new contextually bound memories is reserved for words that have clear ambiguity (perhaps to help resolve the competition between meanings), but for other less ambiguous words there is no need to recruit a secondary form of memory. Therefore, further tests are necessary to determine the involvement of episodic memory in bringing about the word-meaning priming effects in nonhomonyms (Curtis et al., 2022). One way to do that is to introduce a sleep manipulation to the paradigm in Curtis et al. such that participants are tested after 12 hr spent awake or asleep (as in Gaskell et al., 2019). As mentioned above, while a substantial body of evidence has established a role of sleep in the consolidation of newly acquired episodic memory, there is little reason to suspect that sleep-related consolidation is relevant to crystallized knowledge in the neocortex. Therefore, if word-meaning priming in nonhomonyms survives a 12-hr delay including sleep but not a 12-hr delay of wakefulness, it would provide further supporting evidence for the episodic context account. We tested this prediction in Experiment 1, where we adopted the materials and measures from Curtis et al. (2022) but added a sleep manipulation to the experimental design. In Experiment 2, we tested essentially the same prediction but focused on a different type of context-specific lexical information, namely word class (aka part-of-speech, grammatical class, and syntactic category). This allowed us to test the generality of the episodic context account even further. Below, we set our rationale for this experiment.

English words like *stamp* and *loan* can serve as both a noun and a verb without undergoing any morphological changes. These words are similar to nonhomonyms in that they typically have one core meaning, but their precise interpretation is shaped by its context-specific word class (Gentner & France, 1988). For instance, the word *stamp* in "He got a stamp on his passport" refers to an inked impression, but in "He will stamp his passport," it refers to the act of stamping with a rubber stamp. In these instances, the core meaning of *stamp* is roughly the same, but its precise interpretation inevitably changes when it crosses word class. This semantic difference between the noun and verb version of the same word, at least in

English, may be attributable to verb (vs. noun) meanings being generally more abstract, mutable, and dependent on their surrounding contexts (Bird et al., 2000; Gentner, 1981; Kersten & Earles, 2004; Talmy, 1975; see also Ahrens, 1999; Fausey et al., 2006). Therefore, while words like *stamp* are referred to as being word-class ambiguous, this ambiguity is not restricted to the morphosyntactic level, and it almost always extends to semantics (Chiarello et al., 2002; Vigliocco et al., 2011). This view is supported by the observation that resolving syntactic ambiguity may rely on largely the same cognitive mechanism for resolving semantic ambiguity (MacDonald et al., 1994; Rodd et al., 2010). The episodic context account, therefore, predicts that encountering a word-class ambiguous word in a sentence will trigger a contextually bound episodic representation that has the potential to prime language users to subsequently interpret or use the word the same way as in the representation (i.e., in the same word class). It also predicts that this representation will maintain its ability to prime usage over longer delays if a sleep opportunity is afforded. We tested these predictions in Experiment 2.

To sum up, the two experiments reported in this article evaluated the episodic context account by asking whether contextual priming is stronger after sleep than after wakefulness. The two experiments had several similarities: First, they each comprised two sessions, separated by a 12-hr delay, where half of the participants had a period of overnight sleep (sleep group) and the other half engaged in normal activities in daytime (wake group). Second, participants in Session 1 read some prime sentences before tests of priming 10 min later, and 12 hr later in Session 2. The key difference is that Experiments 1 and 2 made use of nonhomonyms and word-class ambiguous words respectively.

Experiment 1

This experiment, including its exclusionary criteria and analysis plan, was preregistered ahead of data collection (https://aspredicted.org/YGP_4N5). Any deviations from the preregistered plan are explicitly stated. We hypothesized that (a) experiencing nonhomonyms in a sentential context would prime participants to subsequently interpret these words in a way consistent with that specific context (Curtis et al., 2022) and (b) such contextual priming effects would be better maintained over sleep (vs. wakefulness).

Method

Design Overview

Session 1 began with a reading task, where participants read 48 nonhomonymic targets (e.g., *bathtub*) embedded in sentences that primed interpretation toward a specific semantic aspect of the targets (e.g., *slip*). To measure priming shortly after exposure, we used speeded relatedness judgment and associate production (e.g., Cai et al., 2017; Gilbert et al., 2018, 2021). In both tasks, participants were tested on 24 of the primed nonhomonyms and 24 unprimed control items. Twelve hours later, including either daytime wakefulness or overnight sleep, participants completed the same tasks in Session 2. Here, participants were tested on 72 nonhomonyms. Of these, (a) 24 were primed but not tested in Session 1, (b) 24 were primed and already tested in Session 1, and (c) 24 were unprimed control items also already tested in Session 1. The latter two categories (cf. Gaskell et al., 2019) gave us an opportunity to test whether repeated testing interacted with sleep-related effects (e.g., Antony et al., 2017) but they were not intended to address our key research question, and hence, were discarded from the main analysis (interested readers can refer to the online materials at the Open Science Framework page [https://osf.io/z2ane/?view_only=ac75981a9ac3403b8e45c841a4f31ac5]).

Participants

We began by recruiting potential participants from Prolific (www.prolific.co), who filled out a screening survey. Here, they provided basic demographic information, read about the details of the main study, and indicated whether they wanted to take part in it ($N =$ 507). Our inclusionary criteria were: (a) aged between 18 and 25, (b) speak English as (one of) their first language(s), (c) currently reside in the United Kingdom, (d) no history of any psychiatric, developmental, or sleep disorders, and (e) willing and able to take part in both sessions of the study. We screened out 84 respondents who did not meet these criteria. This left us with 423 respondents, who were randomly allocated to the wake or sleep groups. They each received a url link inviting them to take part in the main study at a specific time. Of those who took up the invitation ($N = 199$), 157 completed all sessions. Thirty-two of these were excluded from further analysis for meeting our exclusionary criteria: nine for giving a sleepiness score of six or above on the Stanford Sleepiness Scale (SSS), one in the wake group for having a nap between Sessions 1 and 2, four in the sleep group for sleeping <6 hr in the delay interval, 10 in the sleep group for reporting to have poor sleep quality, and eight for performing below a preregistered threshold in one of the experimental tasks (i.e., accuracy $< 70\%$ in relatedness judgment or mean response time >3 *SD*s from the sample mean).

The final sample size was 125 (83 female, 41 male, one other; $M_{age} = 222.05$, $SD_{age} = 2.18$), with 64 in the sleep group and 61 in the wake group. We note that this fell marginally short of the target sample size, which was 66 per group in our preregistration. The target sample size was selected to be in line with Brysbaert and Stevens' (2018) recommendation of ∼1,600 observations per condition (66 participants × 24 trials per condition = 1,584 observations). All participants were native English speakers, with normal or corrected-to-normal eyesight. None reported history of any language, attentional or sleep disorders.

Materials

Nonhomonymic Targets. We used the 72 nonhomonymic target nouns from Curtis et al. (2022; Experiment 3). Examples include *bathtub, balloon, stamp*, and *infection*. Each has a single entry in the Wordsmyth online dictionary (https://www.wordsmyth.net/). This makes them nonhomonyms in terms of lexicographers' judgments, which tend to agree well with participant responses (Rodd et al., 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., having a single core meaning), 54 of 72 targets had more than one sense according to the Wordsmyth database and the average number of senses across all items was 2.81 (for details see Curtis et al., 2022). Lexical properties of the 72 targets are as follows: Log word frequency ranged from 2.04 to 4.68 per million, with a mean of 3.94 ($SD = 0.52$; subtitle lexicon [SUBTLEX-UK]; van Heuven et al., 2014). Concreteness ranged from 2.14 to 5, with a mean of 4.39

($SD = 0.75$; Balota et al., 2007). Finally, word length ranged from 4 to 13 letters, with a mean of 6.25 ($SD = 1.79$).

Probe Words. Each nonhomonymic target was paired with a probe word, taken from Curtis et al. (2022; Experiment 3). The probe was related to a specific semantic aspect of the nonhomonymic target (e.g., *bathtub-slip*) and shared a weak forward association strength with the target ($M = 0.033$, $SD = 0.048$, range = 0.013–0.324; D. L. Nelson et al., 2004).

Prime Sentences. We used the same prime sentences as Curtis et al. (2022; Experiment 3). These sentences were manually generated such that they biased interpretation of the nonhomonymic targets toward their respective probes (e.g., "The old man fell while getting out the bathtub"). Note that the probe words never appeared in the prime sentences, so they were never experienced together with the targets. Findings from three prior experiments ($N = 196$; Curtis et al., 2022) confirmed that these prime sentences were highly effective in biasing readers' interpretation of the nonhomonyms toward the probes. Given 72 nonhomonym targets, there were a total 72 prime sentences ($M \pm SD$ word count = 12.61 ± 3.17).

We split the prime sentences into three lists of 24, which were matched on target word length, $F(1,70) = 0.41$, $p = .523$, target word frequency, $F(1,70) = 0.008$, $p = .931$, and sentence word count, $F(1,70) = 0.17$, $p = .685$. Participants were randomly assigned to read the prime sentences from two of the lists in Session 1 ($N = 48$) while the remaining list served as the unprimed controls ($N = 24$). Assignment to the primed and unprimed conditions was fully counterbalanced across lists.

Design

The study had a 2 (Group) × 3 (Priming) mixed design. The between-participant variable, Group, had two levels: wake and sleep. The within-participant variable, Priming, had three levels: Primed nonhomonyms tested in Session 1 (PrimedSession1), primed nonhomonyms tested for the first time in Session 2 (PrimedSession2), and unprimed controls tested in Session 1.

Procedure

The procedure of Experiment 1 is visualized in Figure 1. This experiment was programmed using Gorilla experiment builder (https://gorilla.sc/). Participants were restricted to using a desktop or laptop computer and were asked to complete the study at a quiet location of their own choosing. One concern associated with online testing is data quality, as it is impossible to monitor the participants during the experiment or control for potentially important factors like the physical environment of the participants. However, it has been repeatedly demonstrated that as long as appropriate measures, such as preregistration and attention checks, are in place (e.g., Curtis et al., 2022; Mak, 2021; Rodd, 2019), data quality from online experiments is no different from lab-based experiments (e.g., Anwyl-Irvine et al., 2020; Barnhoorn et al., 2015; Mak et al., 2021a; Mak & Twitchell, 2020). Furthermore, three recent online studies using the same sleep manipulation as in the current study (Ashton et al., 2021; Mak et al., 2023; Mak & Gaskell, 2023) found clear evidence of a sleep benefit in the classic paired-associate learning paradigm, replicating well-established evidence from lab-based studies (e.g., Lo et al., 2014; Plihal & Born, 1997; Scullin, 2013). Together, these suggest that it is possible to detect sleep-related memory effects in online experiments, as long as preregistered exclusionary criteria and attention checks are put in place, both of which were implemented in the current (and the next) experiment.

The study was split into two sessions, separated by approximately 12 hr. For the wake group, participants completed Session 1 between 8 and 10 AM and Session 2 between 8 and 10 PM on the same day. Participants in the sleep group completed Session 1 between 8 and 10 PM and Session 2 between 8 and 10 AM the next day.

Session 1. The session began with two measures of alertness. The first is a subjective measure, where participants rated their level of sleepiness on the SSS (Hoddes et al., 1973). The second is an objective measure, namely a simple reaction time (simple RT) task. Then, participants completed a reading task where they read 48 prime sentences plus 32 filler sentences. This was then followed by a 10-min filler task, where participants watched a 10-min video called "Shaun the Sheep," chosen for its minimal linguistic content. To ensure that participants watched the video, it was followed by three simple comprehension questions. We preregistered to exclude participants from all the analyses if they got two or more questions wrong. No participants were excluded on this basis. The session ended with tests of word-meaning priming: speeded relatedness judgment and associate production. The order was fixed across participants, with relatedness judgment always preceding association production. We note that the ordering of the two tasks was important in that the first would likely act as a further source of bias on the outcome of the second task regardless of the order. That said, the potential for bias from the relatedness judgment task is fixed and present in all three priming conditions (i.e., the same probe is used in all conditions). The same, however, 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 priming 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 (Curtis et al., 2022). Session 1 lasted approximately 30 min.

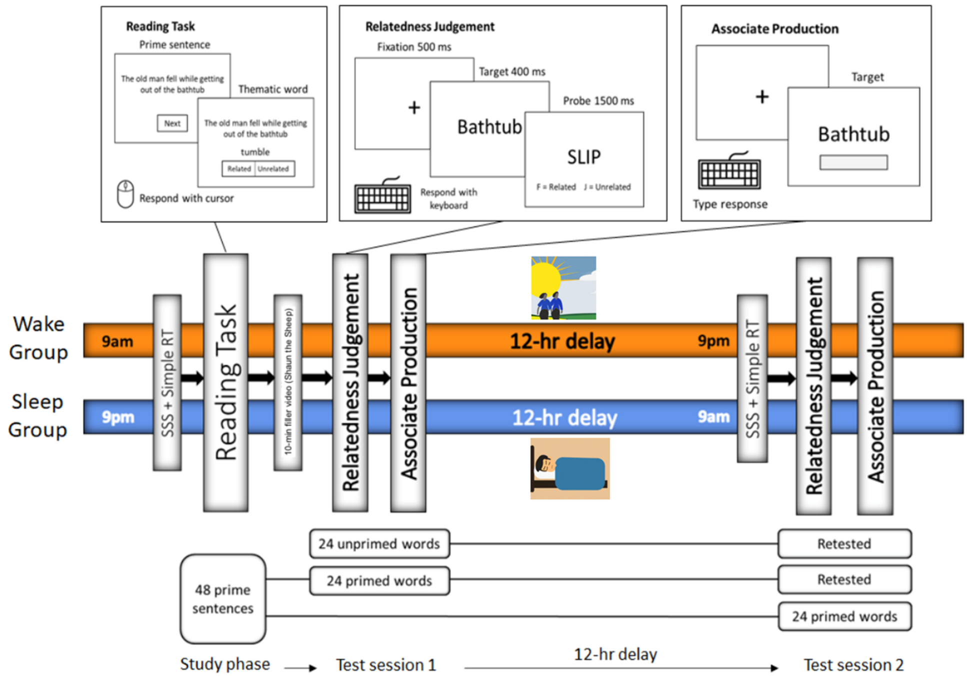
Session 2. It started with participants giving an SSS rating and filling out a reduced version of the Morningness/Eveningness Questionnaire (MEQ; Adan & Almirall, 1991), which assessed circadian preference. This was followed by simple RT, relatedness judgment, and associate production in a fixed order. At the end of this session, participants were asked (a) Have you noticed anything in particular about the sentences and words you just read? (b) Do you have any ideas as to what the experiment was about. Over 70% responded "no" to at least one question, with the remaining hinting that they thought the sentences and test items may have been related in some way or that the study concerned how sleep affects memory. No one indicated that they realized we were trying to prime specific interpretations of words. Session 2 lasted approximately 10 min.

Tasks

Simple RT. This was a simple decision task to assess alertness. In each trial, participants were shown a fixation cross for 500 ms before seeing two digits ordered as "1–0" or "0–1." They pressed the "f" key on their keyboard if "1–0" was shown and the "j" key if "0–1" was shown. No time limit was imposed although participants were asked to respond as quickly as possible. Nine trials showed "1–0," another nine showed "0–1," hence 18 trials in total. Order of presentation was randomized.

Figure 1

Procedure of Experiment 1, Along With Sample Trials



Note. As per our preregistration, the retested items in Session 2 were not included in the main analyses. Interested readers can refer to materials on OSF (https://osf.io/z2ane/?view_only=ac75981a9ac3403b8e45c841a4f31ac5) for details. See the online article for the color version of this figure.

Reading Task. Participants read a sentence in each trial (e.g., "The old man fell while getting out of the bathtub"). No time limit was imposed. A cover task was in place to ensure that participants were paying attention: Once participants finished reading a sentence, they pressed a button to reveal a "thematic word," displayed below the sentence. Here, participants judged whether the word (e.g., *tumble*) was related to the sentence. The thematic words were always related to the prime sentences. This was to ensure that the target words were all processed under similar learning conditions. To create the "unrelated" trials, we included 32 filler sentences from Curtis et al. (2022; $M \pm SD$ word count = 14.28 ± 2.83), which were matched with the prime sentences on word count. They did not contain any of the target words and were always unrelated to their respective thematic words (e.g., "He tried to take a photo of the waterfall as the sun cast a beautiful rainbow through the spray"—*basketball*). In total, participants read 80 sentences, with 48 being the prime sentences and the remaining 32 being fillers. The sentences were split equally between two blocks with materials being counterbalanced between blocks across participants. For each participant, the items that appeared in Block 1 of this reading task always appeared in Block 1 of the subsequent relatedness judgment task and, likewise, items that appeared in Block 2 here always appeared in Block 2 of relatedness judgment. This manipulation was intended to help equalize the amount of passing time between reading and test for any given item within Session 1.

Relatedness Judgment. In each trial, participants judged as quickly and as accurately as possible whether two words were related in meaning. A trial began with a fixation cross for 500 ms, followed by a target word (e.g., *bathtub*) for 400 ms. After a 200-ms blank screen, a probe word, presented in capital letters (e.g., *SLIP*), was shown. Here, participants had 1,500 ms to decide whether the target and probe were semantically related, using the "f" (related) or "j" (unrelated) keys on their keyboards. If participants exceeded 1,500 ms, "Too slow!" appeared on the screen for 1,500 ms. The task began with a practice block of six trials (three related, three unrelated), during which participants received feedback on their performance. No feedback was given in the experimental block. Accuracy and response time were recorded.

In Session 1, participants made judgment to 24 target–probe pairs that were primed in the previous reading task and to 24 unprimed control pairs. Words in these pairs were always related to each other. To create the "unrelated" trials, 48 filler pairs were used (e.g., *mildew–center*). In total, participants were presented with 96 pairs in Session 1. These pairs were then retested in Session 2, which included an additional 24 target–probe pairs. These were the nonhomonyms that were primed but not tested in Session 1.

Associate Production. Each trial began with a 500-ms fixation cross, followed by a target word presented at the center of the screen. Beneath the target was a textbox, where participants typed out the first word that came to mind upon seeing the target. Participants proceeded by pressing the "Return" key on their keyboard but could not proceed without giving a response. No time limit was imposed.

As in relatedness judgment, Session 1 showed 24 nonhomonymic targets primed in the reading task and 24 unprimed controls. In Session 2, these 48 items were retested, in addition to 24 targets that were primed but not already tested in Session 1.

Results

Comparability Between the Wake and Sleep Groups

Before proceeding to the main analysis, we first checked the comparability of the wake and sleep groups on multiple measures (see Table 1).

We first checked whether the two groups differed in their levels of alertness, as reflected by SSS and simple RT. We performed a two-way mixed analysis of variance (ANOVA) on each measure, with Session (1 vs. 2) and Group (wake vs. sleep) as the independent variables. The analysis on SSS showed no main effects of session, $F(1,123) = 3.71$ $p = .056$, or Group, $F(1,123) = 0.6$, $p = .439$, and there was no interaction, $F(1, 123) = 0.23$, $p = .631$. The analysis on simple RT showed a main effect of session, $F(1, 123) = 23.69$, $p < .001$, with participants responding more quickly in Session 2 (vs. 1). Importantly, however, there was no main effect of Group, $F(1, 123) = 0.08$, $p = .777$, and no interaction, $F(1, 123) = 1.89$, $p = .172$. In sum, the wake and sleep groups were highly comparable on both measures of alertness across sessions. Then, an independent t-test showed that the two groups were also well matched on their performance in the reading cover task, $t(121.53) = -0.43$, $p = .666$. Finally, an independent t-test comparing MEQ scores showed a significant difference between groups, $t(122.85) = 2.58$, $p = .011$, with participants in the wake group having greater morningness preference. This was unanticipated, as participants were randomly allocated to the wake and sleep groups. Perhaps, this difference reflects some selective attrition effects such that participants with a morningness preference were more likely to take part in the study if they were assigned to the wake group. In light of this, we explored whether morningness/eveningness preference influenced our outcome measures. We did so by adding MEQ scores and its interaction with Group as fixed effects to all the models reported below (available on OSF, https://osf.io/z2ane/?view_only=ac75981a9ac3403b8e45c841a4f31ac5). Reassuringly, these additional fixed effects were not significant and did not change the interpretation of the findings reported below. This suggests that any group difference in our outcome measures was unlikely to be attributable to differences in time-of-day preference. Notably, the same comparison in Experiment 2 was not significant (and in the opposite direction).

Analysis Approach

All data were analyzed in a mixed-effects environment using the lme4 package (Version 1.1.17; Bates et al., 2015) in R (Version 4.1.1; R Core Team, 2021). The statistical model structures were based on Gaskell et al. (2019). All models had two fixed effects: Group and Priming. Group had two levels (wake, sleep) and was coded using sum contrast. Priming had three levels (PrimedSession1, PrimedSession2, Unprimed control) and was coded using orthogonal Helmert contrasts, giving two comparisons: [i] the Unprimed condition vs. PrimedSession1 and PrimedSession2 combined (Unprimed vs. PrimedS1 + 2), and [ii] PrimedSession1 vs. PrimedSession2 (PrimedS1 vs. PrimedS2). The first of these contrasts assesses the average level of priming across the two sessions. The second assesses the difference in performance in the primed condition across the two sessions (particularly relevant when testing for interactions with Group).

Table 1
Mean (and Standard Deviation) SSS Score, Performance on Simple RT Task, and MEQ Scores Across Groups in Experiment 1

Measures	Wake group	Sleep group
SSS in Session 1	2.42 (1.02)	2.61 (1.06)
SSS in Session 2	2.74 (1.22)	2.80 (1.20)
Simple RT in Session 1 (ms)	404 (68)	400 (60)
Simple RT in Session 2 (ms)	374 (45)	383 (51)
Performance in the reading cover task	93.9% (4.1%)	94.1% (3.9%)
MEQ score	14.57 (3.73)	12.84 (3.77)

Note. (1) SSS stands for Stanford Sleepiness Scale and ranges from 1 to 6, with higher values indicating greater sleepiness. (2) MEQ stands for Morningness/Eveningness Questionnaire; MEQ score ranges from 1 to 25, with higher values indicating greater morningness preference. RT = reaction time.

The random-effect structure of all the models was determined by the *buildmer* package (Voeten, 2020), which automatically finds the maximal model that is capable of converging using backward elimination. This means that we began by providing *buildmer* with the maximal random-effect structure, as justified by the experimental design (i.e., including random slopes and intercepts for both participants and items; Barr et al., 2013). The "bobyqa" optimizer was used to enhance the likelihood of convergence, and the direction of the model was set to "order" so that no elimination of any fixed effects or interactions would take place.

For relatedness judgments, there were two dependent variables: accuracy and response time (RT). Accuracy was binary (correct vs. incorrect) and was analyzed using a generalized linear mixed-effect (GLM) model. RT was numeric and was subject to a log-transformation before being fitted to a linear mixed-effect model (LME). The RT analysis was based only on the "related" trials that received a correct judgment.

For associate production, the dependent variable was binary: whether a participant's response was consistent with the prime sentences read by the participants. Consistency was determined by 12 third-party human raters, recruited via Prolific. They were presented with the associate responses produced by the participants in the main study, one at a time. This was presented beneath two sentences. One of them was a prime sentence (e.g., "The old man fell while getting out of the bathtub"). The other also contained the nonhomonymic target but the sentence referred to a different aspect of the target's meaning (e.g., "After our walk I put the muddy dog in the bathtub and scrubbed him with dog shampoo"). Raters were asked which of the two sentences the associate response was related to or whether it was equally related or unrelated to both. Following Curtis et al. (2022), an associate response was coded as consistent with the prime sentences if at least seven of the 12 raters judged it as being more related to the prime sentences.

Predictions

The overriding prediction in our preregistration was that over a 12-hr delay period including daytime wakeful or overnight sleep, sleep will benefit the retention of word-meaning priming effects for nonhomonymic words. However, the more specific predictions fleshed out from this in the preregistration were not optimal in the sense that they could not be straightforwardly mapped onto our preregistered analysis plan. We decided to adhere to the analysis plan because (a) it was identical to that of a previous study

(Gaskell et al., 2019), and (b) it had been preregistered. The analysis tested these key questions:

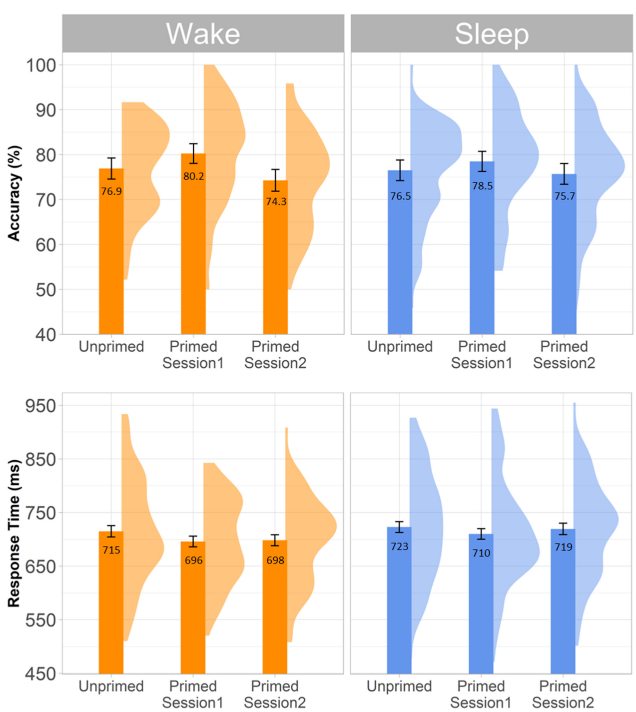
1. Whether there is word-meaning priming (primed > unprimed) when the data are collapsed across groups and sessions. If there is, a Priming [i] main effect (i.e., Unprimed vs. PrimedS1 + 2) should emerge.
2. Whether word-meaning priming changes in magnitude across Sessions 1 and 2. If it does, a Priming [ii] main effect (i.e., PrimedS1 vs. PrimedS2) should emerge.
3. Whether word-meaning priming in Session 2 (vs. Session 1) differs between groups. If it does, an interaction between Group and Priming [ii] (i.e., PrimedS1 vs. PrimedS2) should emerge.

Confirmatory Analyses

Relatedness Judgment/Accuracy. Accuracy across groups and priming conditions is summarized in the top panel of Figure 2. There was no effect of Priming [i] ($\beta = 0.02$, $SE = 0.04$, $z = 0.49$, $p = .627$), suggesting no significant word-meaning priming effect when averaged across all conditions ($M_{(S1+S2)/2-\text{Unprimed}} = +0.48\%$). However, there was a significant effect of Priming [ii] ($\beta = -0.18$, $SE = 0.04$, $z = -4.46$, $p < .001$), suggesting that there was more evidence of priming across groups in Session 1 ($M_{S1-\text{Unprimed}} = +2.7\%$) than Session 2 ($M_{S2-\text{Unprimed}} = -1.7\%$). Importantly, this was qualified by a significant interaction between Group and Priming [ii] ($\beta = -0.07$, $SE = 0.03$, $z = -2.01$, $p = .044$), suggesting that the drop in priming between sessions was more substantial for the wake ($M_{S2-S1} = -5.95\%$) than the sleep ($M_{S2-S1} = -2.78\%$) group. Although this interaction is in line with our predictions, the lack of a main effect for Priming [i] makes interpretation less straightforward. Finally, the interaction between Group and Priming [i] was not significant ($\beta = -0.006$, $SE = 0.02$, $z = -0.31$, $p = .757$).

Figure 2

Accuracy (Top) and RT (Bottom) in Relatedness Judgment, Summarized Across Group and Priming Conditions for Experiment 1



Note. Error bars represent 95% within-subject CI (Morey, 2008). Density functions represent the distribution of participant means in each priming condition. RT = response time. See the online article for the color version of this figure.

Relatedness Judgment/RT. RTs across groups and priming conditions are summarized in the bottom panel of Figure 2. There was a significant effect of Priming [i] ($\beta = -0.23$, $SE = 0.07$, $t = 2.92$, $p = .004$), suggesting that RTs were faster in the primed than in the unprimed conditions ($M_{(S1+S2)/2-\text{Unprimed}} = -13$ ms). However, contrary to our prediction, there was no significant interaction between Group and Priming [ii] ($\beta = 0.06$, $SE = 0.08$, $t = -0.73$, $p = .468$), meaning that the magnitude of word-meaning priming across sessions did not differ between the wake ($M_{S2-S1} =$ +2 ms) and sleep ($M_{S2-S1} = +9$ ms) groups. All other fixed effects and interactions were nonsignificant (see Table 2 for details).

Associate Production. In this analysis, three participants from the wake group were excluded from further analysis, because they responded with the target word in each trial (e.g., target: *bathtub*, response: *bathtub*), suggesting that they misunderstood the task instructions. This analysis, therefore, was based on the remaining 122 participants (64 and 58 participants in the sleep and wake group, respectively).

The percentage of associate responses that were consistent with the prime sentences is summarized across groups and priming conditions in Figure 3. There was a significant effect of Priming [i] ($\beta = -0.13$, $SE = 0.03$, $z = -3.85$, $p < .001$; see Table 2), indicating that there were more consistent associates in the primed conditions than in the unprimed condition ($M_{(S1+S2)/2-\text{Unprimed}} = +6\%$). There was no main effect for Priming [ii] ($\beta = 0.05$, $SE = 0.04$, $z = 1.21$, $p = .227$) and no interaction between Group and Priming [i] ($\beta = 0.009$, $SE = 0.02$, $z = 0.48$, $p = .633$). Most importantly, there was a significant interaction between Group and Priming [ii] ($\beta = 0.11$, $SE = 0.03$, $z = -3.61$, $p < .001$). This suggests that the change in the percentage of consistent response across sessions differed between groups, with the sleep ($M_{S2-S1} =$ +2.4%) group showing a numeric increase but the wake group a numeric reduction ($M_{S2-S1} = -5.7\%$).

Exploratory Analyses

Time-of-Day Effects. Time of day is known to affect performance on some cognitive tasks (e.g., Lorenzetti & Natale, 1996; Oakhill, 1986; Tandoc et al., 2021). Since our AM–PM/PM–AM design is naturally confounded with time of day, it is necessary to consider the extent to which our key finding, relating to the change in performance in the primed condition across periods of sleep and wake, was attributable to time-of-day effects. An interaction attributable to time-of-day effects would be a cross-over interaction, with high levels of priming in Session 1 for the wake group and in Session 2 for the sleep group (both in the morning), and then weak priming in the other two conditions (both in the evening). An interaction that may be attributable to a consolidation effect

Table 2
Summary of the Mixed-Effects Models for Each Dependent Measure in the Confirmatory Analyses, Experiment 1

Fixed effects	Estimates	*SE*	*z/t* values	*p*
Relatedness judgment/accuracy				
Intercept	1.60	0.15	10.68	<.001*
Group	−0.20	0.05	−0.28	.777
Priming [i] (Unprimed vs. PrimedS1 + 2)	0.02	0.04	0.49	.627
Priming [ii] (PrimedS1 vs. S2)	−0.18	0.04	−4.46	<.001*
Group × Priming [i]	−0.006	0.02	−0.31	.757
Group × Priming [ii]	0.07	0.03	−2.01	.044*
Relatedness judgment/RT				
Intercept	2.84	0.006	453.15	<.001*
Group	−0.41	0.44	0.92	.358
Priming [i] (Unprimed vs. PrimedS1+2)	0.23	0.08	2.92	.004*
Priming [ii] (PrimedS1 vs. S2)	0.27	0.14	1.94	.052
Group × Priming [i]	0.06	0.08	−0.73	.468
Group × Priming [ii]	0.08	0.1	0.58	.560
Associate production				
Intercept	−0.58	0.16	−3.71	<.001*
Group	0.05	0.08	0.63	.531
Priming [i] (Unprimed vs. PrimedS1 + 2)	−0.13	0.03	−3.85	<.001*
Priming [ii] (PrimedS1 vs. S2)	0.05	0.04	1.21	.227
Group × Priming [i]	0.009	0.02	0.48	.633
Group × Priming [ii]	0.11	0.02	−3.61	<.001*

Note. Due to the logarithmic scale, the estimate and *SE* of the RT analysis have been multiplied by 100 (excluding the intercept) to improve interpretability. This includes those values reported in the text. RT = reaction time; S1 and S2 = Session 1 and Session 2.
* $p < .05$.

would show high levels of priming for both groups in Session 1, which maintained for the sleep but not the wake group. Both of these interpretations would predict a difference between sleep and wake in terms of priming in Session 2, but only the time-of-day interpretation would predict a difference in Session 1. We, therefore, conducted a set of exploratory analyses comparing the two groups on their performance on relatedness judgment and associate production in Session 1 (which took place in the AM for the wake group and in the PM for the sleep group). We fitted the data from the two experimental tasks into mixed-effect models, which included Group (wake vs. sleep), Priming condition (Unprimed vs. PrimedSession1), and their interaction as the fixed effects. These models (see "Appendix A" for detailed output) consistently revealed no effect of group and no group by priming interaction. These suggest highly comparable priming effects between groups in Session 1, mirroring our alertness measures (i.e., simple RT + SSS). Together, they provide some assurance that our key findings of a more robust contextual priming effect (in associate production) postsleep were unlikely to be driven by time-of-day effects.

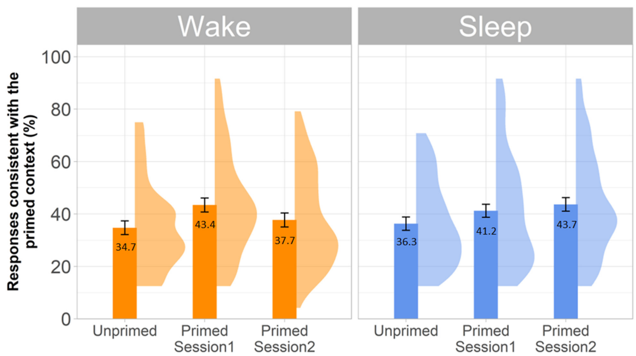
Pairwise Comparisons. As per the request from an anonymous reviewer, we conducted a set of post hoc pairwise comparisons to facilitate interpretation of our findings. The first set compared test performance between Groups within each Session while the second compared test performance between Sessions within each Group. We obtained the estimated marginal means from the mixed-effect models reported above, using the emmeans package (Lenth et al., 2022), with no correction applied for multiple comparisons (see Table 3 for *p* values).

Here, we highlight the key points from these comparisons: In Session 2, the sleep group seemed to produce more probe-consistent responses than the wake group ($p = .0503$), although this was not statistically significant. As for performance across Sessions within each Group, there was a numeric trend suggesting an increase in priming in the sleep group, while in the wake group, there is some evidence suggesting the opposite.

Discussion

A prior study by Curtis et al. (2022) reported clear word-meaning priming effects in nonhomonyms 10–30 min after exposure to these words in a disambiguating sentential context. Here, using the same materials and tests, we investigated whether this priming effect in nonhomonyms is maintained over sleep (vs. wakefulness). Participants first read prime sentences that biased interpretation of a nonhomonymic noun (e.g., *bathtub*) toward a specific semantic aspect of the target's meaning (e.g., *slip*). Shortly after reading these sentences, participants completed two indices of priming: relatedness judgment and associate production. Twelve hours later, which included either daytime wakefulness or a period of overnight sleep, participants attempted the same tasks again for a different set of items. In relatedness judgment, the results are not straightforward. For RT, there was evidence of word-meaning priming overall but this did not vary significantly across groups or sessions. For accuracy, there was no significant priming effect overall but accuracy decreased significantly in the primed conditions from Session 1 to Session 2, consistent with a reduction in priming over time (Rodd et al., 2013, 2016). This drop in accuracy was weaker in the sleep group than the wake group. These findings do not provide clear support for our hypotheses that word-meaning priming is maintained over sleep, although they are broadly consistent with the findings from Curtis et al. (2022) that word-meaning priming extends to nonhomonyms. In associate production, however, the picture was more

Figure 3
Percentage of Associate Responses Consistent With the Prime Sentences in Associate Production, Summarized Across Group and Priming Conditions for Experiment 1



Note. Error bars represent 95% within-subject CI (Morey, 2008). Density functions represent the distribution of participant means in each priming condition. See the online article for the color version of this figure.

clear-cut: There was a clear word-meaning priming effect overall, and more importantly, as predicted, the between-session change in priming magnitude differed between groups, with the sleep group showing a numeric increase but the wake group a numeric reduction. As to why the pattern of priming effects was clearer in this task compared with relatedness judgment, we propose that it may be related to the two tasks requiring different retrieval processes, such that active response generation may be more sensitive to sleep-related effects than yes–no judgment (e.g., Diekelmann et al., 2009). We will revisit this point in General Discussion, where we also consider in detail how episodic memory and sleep are involved in language comprehension.

Before we turn to Experiment 2, we briefly consider the extent to which word-meaning priming for nonhomonyms is driven by changes in meaning/sense availability (see Curtis et al., 2022 for an in-depth discussion). In our previous study (Curtis et al., 2022; Experiment 1), we made use of 72 nonhomonyms, 52 of which are classified by Wordsmyth dictionary as having more than one sense (i.e., polysemes). As in our current study, these nonhomonyms (e.g., *athlete*) were paired with a probe word (e.g., *injury*), but unlike the current study, they were read in either a probe-consistent ("The athlete fell off her skis and sprained her arm, so she was unable to train.") or a probe-inconsistent sentence ("This athlete has only done cross country before so the 800 m will be her first track event."). In both sentences, *athlete* refers to the same sense (i.e., a competitive sportsperson), so if word-meaning priming were a consequence of enhancing the availability of a particular sense, both sentences should be equally (un)able to prime participants toward the probe (i.e., *injury*). However, Curtis et al. (2022; Experiment 1) reported that shortly after sentence exposure (-~20 min), the probe-consistent sentences primed participants toward the probes significantly more than the probe-inconsistent sentences across three experimental measures. These findings make it difficult to explain word-meaning priming for nonhomonyms in terms of an adjustment of the balance between senses of a polysemous word. Notably, of the 72 nonhomonyms in Curtis et al. (2022; Experiment 1), 65 of them were judged to be related to the same sense across the probe-consistent and -inconsistent sentences, providing further support against a sense-adjustment interpretation. Returning to our Experiment 1, we used the same 72 nonhomonyms (75% of which are polysemous) and probe-consistent sentences as Curtis et al. (2022; Experiment 3). These sentences biased interpretation of the nonhomonymic targets toward their respective probe words, which are always a readily comprehensible aspect of the target's meanings (e.g., *athlete-injury; bathroom-slip*). In other words, we never intended to prime participants toward an established sense (e.g., *athlete-sportsperson*). As such, together with the findings from Curtis et al. (2022), we are confident that word-meaning priming for nonhomonyms is *not* a consequence of a particular sense in long-term memory being made more available by the sentential context, but is likely to be driven by an episodic, context-specific representation that is generated during language comprehension.

Table 3
Post Hoc Comparisons for Experiment 1

Post hoc comparisons	Relatedness/Acc	Relatedness/RT	Associate production
Sleep-S1 versus Wake-S1	$p = .249$	$p = .413$	$p = .544$
Sleep-S2 versus Wake-S2	$p = .356$	$p = .248$	$p = .0503$
Sleep-S1 versus Sleep-S2	$p = .037$*	$p = .070$	$p = .188$
Wake-S1 versus Wake-S2	$p \leq .0001$*	$p = .344$	$p = .002$*

Note. RT = reaction time; S1 and S2 = Session 1 and Session 2.
* $p < .05$ (not corrected for multiple comparisons).

Experiment 2

As in prior studies concerned with contextual priming (e.g., Curtis et al., 2022; Gaskell et al., 2019; Rodd et al., 2013, 2016), Experiment 1 only used nouns as the targets, so it remains unclear whether our findings extend to other word classes, such as verbs, which are known to be processed differently than nouns (e.g., nouns are easier to comprehend and produce; Vigliocco et al., 2011). In addition, the word-meaning priming effect we and prior studies have demonstrated is built upon semantic ambiguity within a single-word class (i.e., noun), so it is also unknown whether contextual priming operates when the grammatical class of the word is in itself ambiguous. To test the generalizability of the episodic context account, Experiment 2 made use of word-class ambiguous words (e.g., *stamp, loan, riot*) that were experienced as either a noun or a verb depending on the sentential context. We then tested whether this contextual exposure would influence the likelihood of participants treating the ambiguous word as a noun or a verb at a later point, as well as after a delay including sleep or wake. This provides a further test to the generalizability of the episodic context account, which posits that contextual priming and any subsequent sleep-related effects should apply to *any* words as long as the sentential context modifies their interpretation in some way.

Experiment 2 made use of the same general design as Experiment 1. Participants first read sentences where word-class ambiguous words were used as either a noun or a verb. To index priming, we used sentence generation, instead of relatedness judgment and associate production. This switch was motivated by two reasons: First, the context-free nature of relatedness judgment and associate production made it impossible to tell whether a participant is primed toward a specific word class. Second, in Experiment 1, we found clear sleep-related effects in associate production but not in relatedness judgment, implying that outcome measures requiring active response generation (vs. those requiring a yes–no judgment) may be a better fit for measuring sleep-related effects (we will revisit this point in General Discussion). Third, sentence generation (vs. relatedness judgment and associate production) is arguably a more ecologically valid way of assessing contextual priming, as we produce sentences every day but hardly ever evaluate semantic relatedness or produce a single associate. With these in mind, we switched to sentence generation to test whether participants would be primed toward using a specific word class after exposure to the prime sentences, both shortly afterward and 12 hr later. This experiment, including its exclusionary criteria and analysis plan, were preregistered ahead of data collection (https://aspredicted.org/xe6mv.pdf). Any deviations from the preregistered plan are explicitly stated.

Method

Participants

Recruitment procedures were the same as Experiment 1. A total of 574 adults from the University of York and Prolific filled out a screening survey, out of whom 214 did not meet our inclusion criteria. This left 360 respondents, who were randomly allocated to the wake or sleep groups. Of those who took up the invitation ($N = 194$), 149 completed all sessions. Twenty-one of these were excluded from further analysis for meeting our exclusion criteria: nine for an SSS score of six or above, two in the wake group for napping between Sessions 1 and 2, two in the sleep group for sleeping <6 hr in the delay interval, and eight in the sleep group for reporting to have poor sleep quality.

The final sample size was 128 (102 female, 26 male; $M_{age} = 21.75$; $SD_{age} = 2.21$), evenly split between the wake and sleep groups. We note that again we fell somewhat short of the preregistered target sample size, which was also 66 per group. All participants were native English speakers, with normal or corrected-to-normal eyesight. None reported history of any language, attentional, or sleep disorders.

Materials

Target Words. We chose 66 word-class ambiguous words that have an unbalanced noun–verb preference. Half have an existing noun preference (e.g., *loan, quiz*) while the other half have an existing verb preference (e.g., *rub, divide*). This preference was estimated based on the percentage of times a word serves as a noun or verb in the English Web 2020 corpus (containing 36-billion-word tokens; Jakubı̆ek et al., 2013). A word was said to have a noun preference if over 60% of its occurrences in the corpus were classed as a noun. Likewise, a word was said to have a verb preference if over 60% of its occurrences were classed as a verb. Log frequency of the 66 target words ranged from 2.43 to 5.17 per million, with a mean of 4.0 ($SD = 0.56$; SUBTLEX-UK; van Heuven et al., 2014). Concreteness ranged from 1.68 to 4.93, with a mean of 3.74 ($SD = 0.89$; Balota et al., 2007). Word length ranged from 3 to 9 letters ($M = 5.32$, $SD = 1.49$), and phonemically speaking, their pronunciations do not change across word class.

Prime Sentences. Each sentence primed readers to the dispreferred word class of the target words (see "Appendix B" for the full set). For instance, if a target word had an existing noun preference (e.g., *loan*), it served as a verb in the prime sentence (e.g., "Brentwood Borough Council said it would loan individual projects between £1 m and £20 m to transform the borough"). On the other hand, if a target word had an existing verb preference (e.g., *bother*), it served as a noun in the prime sentence (e.g., "She did the washing without complaining because she didn't want to be a bother"). The use of dispreferred word class was motivated by previous studies (Curtis et al., 2022; Gaskell et al., 2019; Rodd et al., 2013) showing that word-meaning priming effects were larger for interpretations that are more strongly subordinate, indicating that it is easier to prime an aspect of a word's meaning if that aspect is less typical or frequent in the first place. Each target word had two prime sentences in order to reduce any stimulus-specific effects. These sentences were taken and modified from the example sentences under the same definition entry in Oxford English dictionary. Given 66 target words, there were a total of 132 prime sentences. We split them evenly into two sets (Sets A and B) such that a target word appeared once in each set. Participants were randomly allocated to read either Set A or B. Target words in these prime sentences did not require any inflectional suffixes, meaning that each target was always experienced as a singular noun or a bare infinitive in the sentences.

Both sets of prime sentences (A and B) were split into three lists of 22. Within each list, half of the target words had an existing noun preference, while the other half had an existing verb preference. The three lists were matched on target word length, $F(2,63) = 1.24$, $p = .296$, target word frequency, $F(2,63) = 0.22$, $p = .797$, target word concreteness, $F(2, 63) = 0.10$, $p = .903$, and sentence word count [Set A: $F(2, 63) = 0.99$, $p = .377$; Set B: $F(2, 63) = 1.0$, $p = .373$]. Participants were randomly assigned to read the prime sentences from two of the lists ($N = 44$) while the third served as the unprimed control ($N = 22$). Assignment to the primed and unprimed conditions was fully counterbalanced across lists.

Design

The study had a 2 (Group) × 2 (Existing Preference) × 3 (Priming) mixed design. Group was manipulated between-participants and had two levels: wake and sleep. Existing preference and Priming were both manipulated within-participants. The former had two levels: noun preference and verb preference. On the other hand, Priming had three levels: Unprimed, PrimedSession1, and PrimedSession2, as in Experiment 1.

There were minor changes from Experiment 1 in terms of how many times and when an item was tested. In Experiment 1, items tested in Session 1 were retested in Session 2. In Experiment 2, each item was tested just once. This was because we did not intend to test if repeated testing interacted with the effects of sleep here. Second, in Experiment 1, all the unprimed control words were tested in Session 1 and then again in Session 2. In Experiment 2, half of the unprimed control words were tested in Session 1 while the other half in Session 2. This was intended to reduce participants' fatigue in Session 1 and to give us an opportunity to firmly disentangle the effects of passage of time (delay) and priming, which is not possible in Experiment 1 since all the unprimed control items were tested in Session 1.

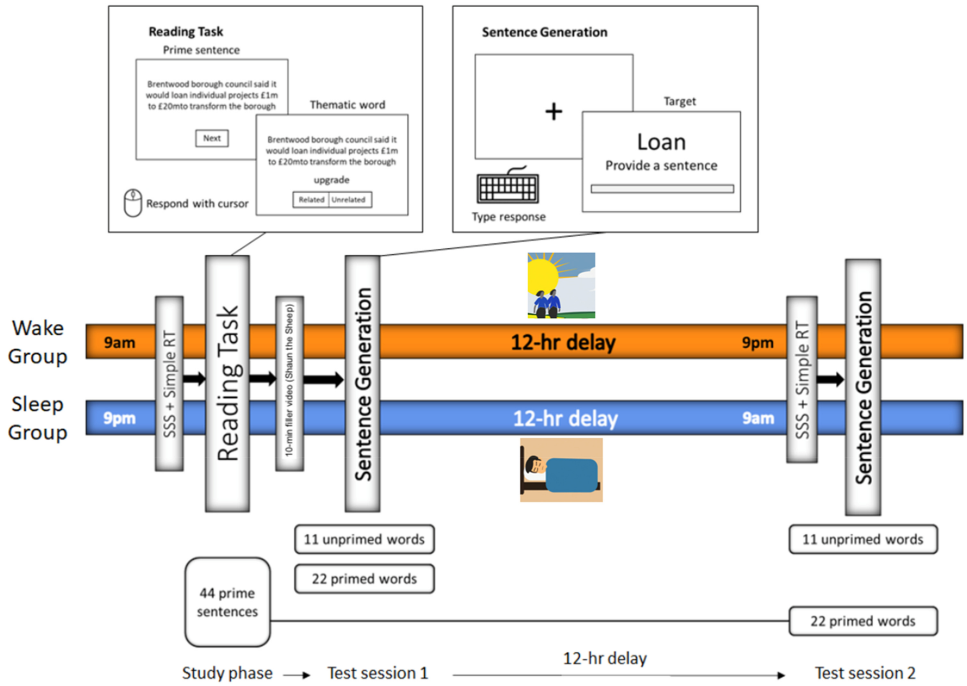
Procedure

The general procedure was largely identical to Experiment 1 and is visualized in Figure 4.

Session 1. This session began with participants rating their level of sleepiness on SSS, followed by a simple RT task. Then,

Figure 4

Procedure of Experiment 2, Along With Sample Trials



Note. See the online article for the color version of this figure.

participants completed the reading task, followed by a filler task, where participants watched a 10-min video of "Shaun the Sheep." This was followed by the same comprehension questions as in Experiment 1, and all participants got at least two of the three questions correct, so no participants were excluded on this basis. Session 1 ended with a sentence generation task, where participants came up with a sentence each for 33 word-class ambiguous words (22 primed + 11 unprimed). Session 1 lasted approximately 35 min.

Session 2. Participants first gave an SSS rating, then completed the MEQ survey. This was followed by a simple RT task and sentence generation, where participants generated sentences to 33 word-class ambiguous words (22 primed + 11 unprimed) that were not tested in Session 1. At the end of this session, participants were asked to guess the purpose of the study. Almost all the participants said the study was interested in whether a word was used as a noun or a verb. No participants showed awareness of the priming phase being there to prime a word toward a specific word class. This session lasted approximately 12 min.

Tasks

Simple RT. This was identical to that of Experiment 1.

Reading Task. The reading procedure, including the cover task, was identical to Experiment 1. Participants read 22 target words with an existing noun preference, each in a prime sentence where they were used as a verb, and 22 target words with an existing verb preference, each in a prime sentence where they were used as a noun. These 44 prime sentences were intermixed randomly with 30 filler sentences that were taken and modified from Experiment 1.

Sentence Generation. The full set of instructions is shown in "Appendix C." Briefly, each trial displayed a word-class ambiguous word at the center of the screen, and participants were required to generate a sentence using the word shown. Participants were encouraged to type the first sentence that came to mind. They were free to use the word as a noun or verb, as long as the sentence contained only one instance of the word and was grammatically correct. If used as a noun, pluralization was acceptable, and as a verb, changing the tense was allowed. Participants could not proceed to the next trial unless their generated sentence contained a minimum of 25 characters (roughly 6–10 words). The next button appeared 12 s after the trial started, forcing participants to spend some time on each trial. Participants generated sentences to a total of 33 word-class ambiguous words (22 primed + 11 unprimed control) in each session. Trial order was randomized.

Results

Comparability Between the Wake and Sleep Groups

As in Experiment 1, we ran a two-way mixed ANOVA on the SSS and simple RT data, with Session (1 vs. 2) and Group (wake vs. sleep) as independent variables (see Table 4 for group means). The analysis of SSS showed no main effects of Session, $F(1, 126) = 0.01$, $p = .913$, or Group, $F(1, 126) = 3.52$, $p = .063$, and there was no interaction, $F(1, 126) = 0.96$, $p = .328$. The analysis of the simple RT data showed a similar pattern as Experiment 1: There was a main effect of Session, $F(1, 126) = 44.51$, $p < .001$, with participants responding more quickly in Session 2 than in Session 1. However, importantly, there was no main effect of

Table 4
Mean (and Standard Deviation) SSS Score, Performance on Simple RT Task, and MEQ Scores Across Groups in Experiment 2

Measures	Wake group	Sleep group
SSS in Session 1	2.80 (1.01)	2.69 (0.92)
SSS in Session 2	2.95 (1.28)	2.56 (1.18)
Simple RT in Session 1 (ms)	409 (56)	396 (58)
Simple RT in Session 2 (ms)	376 (47)	368 (47)
Performance in the reading cover task	91.3% (3.5%)	91.5% (3.9%)
MEQ score	12.84 (4.53)	14.70 (3.53)

Note. (1) SSS stands for Stanford Sleepiness Scale; it ranges from 1 to 6, with higher values indicating greater sleepiness. (2) MEQ stands for Morningness/Eveningness Questionnaire; score ranges from 1 to 25, with higher values indicating greater morningness preference. RT = reaction time.

Group, $F(1,126) = 1.74$, $p = .19$, and no interaction, $F(1, 126) = 0.38$, $p = .54$. Together, these suggest that the two groups did not differ in terms of their alertness across sessions. Next, an independent *t*-test showed that the two groups were also well matched on their performance in the reading cover task, $t(124.69) = -0.189$, $p = .851$. Finally, an independent *t*-test comparing the morningness/evening scores found no significant differences between groups, $t(121) = 1.76$, $p = .08$, suggesting that the two groups were comparable in terms of time-of-day preference (cf. Experiment 2). All in all, these measures increase our confidence that any interactions between group and session in the main analyses below are unlikely to be caused by time-of-day effects on sleepiness or vigilance.

Data Preprocessing

A total of 80 (or 0.94%) trials were excluded from further analysis because: (a) participants added a derivational suffix to a target word (e.g., *loan → loanable; poison → poisonous*), (b) participants used the target words twice in one sentence and the two words were of different part-of-speech (e.g., *All of my friends have parents who are divorced or are getting a divorce*), (c) the target word was used as a proper noun (e.g., *There is a pub in a town near me called The Anchor*), or (d) participants misread the target (e.g., *brother* for *bother*). These exclusions were not preregistered as they were not anticipated.

In our preregistration, we planned to use two independent human raters to determine the word class of the target words in the generated sentences. However, since we had a very large dataset (8,448 sentences), we decided to improve efficiency by automating the process with a part-of-speech tagger (Schweinberger, 2016), in addition to a human rater (Adam J. Curtis). Both determined whether the target word in each generated sentence was a noun or verb. The agreement rate was high, at 98.8%. Trials where the two disagreed were usually due to spelling or grammatical errors in the generated sentences and were resolved by a third rater (Matthew H.C. Mak).

Analysis Approach

The dependent variable was binary: whether participants used a target word in its dispreferred (i.e., primed) word class. For instance, the word *loan* has an existing noun preference; if a participant generated a sentence that used *loan* as a verb, this trial received a score of 1. However, if used in the preferred word class (i.e., noun), it was scored as 0. This dependent variable was directly analogous to the dependent variable in associate production of Experiment 1, where an associate response was classed as either consistent or inconsistent with the prime sentence.

Following the analysis approach in Experiment 1, we fitted the data to a GLM model, using the lme4 package in R. The model had four fixed effects: Group, Existing Preference, Priming, and their interactions. Group and Existing Preference each had two levels and were effect coded. Priming had three levels (Unprimed, PrimedSession1, and PrimedSession2), which were coded using orthogonal Helmert contrasts. As in Experiment 1, this compared [i] the Unprimed condition with the combination of PrimedSession1 and 2 (Unprimed vs. PrimedS1 + 2), and [ii] PrimedSession1 with PrimedSession2 (PrimingS1 vs. PrimingS2). The random-effect structure was determined by the *buildmer* package, as in Experiment 1.

Predictions

Our preregistered predictions and how each of them corresponds to our analysis are as follows:

1. A word-class ambiguous word would be more likely to be used in its dispreferred word class if it was primed in the reading task. If correct, this would correspond to a main effect of Priming [i].
2. This priming effect would decay from Sessions 1 to 2. If correct, this would correspond to a main effect of Priming [ii].
3. The change in the magnitude of priming between sessions would be different between the sleep and wake groups. If correct, this would correspond to a significant interaction between Group and Priming [ii].

Confirmatory Analyses

Table 5 shows the percentage of generated sentences where the target words were used in their dispreferred word class, summarized across groups, priming conditions, and existing word-class preference.

Table 5
Mean Percentage (and Standard Deviation) of Generated Sentences Where the Target Words Were Used in Their Dispreferred Word Class, Summarized Across Groups, Existing Word-Class Preference, and Priming Conditions

Groups	Wake			Sleep		
Priming conditions	Unprimed	Primed S1	Primed S2	Unprimed	Primed S1	Primed S2
Words with a noun preference	19.7% (14.7)	24.7% (14.1)	22.8% (15.5)	20.9% (13.2)	26.7% (14.6)	27.4% (16.1)
Words with a verb preference	26.7% (13.1)	35.1% (18.0)	31.4% (16.5)	25.1% (14.2)	33.0% (16.5)	37.8% (19.2)

Note. S1 = Session 1; S2 = Session 2.

As revealed by the main analysis below, existing word-class preference did not interact with the other two variables, so for simplicity's sake, we collapsed the data across existing preference for visualization (see Figure 5).

Consider the unprimed control words first. On average, these were used in their dispreferred word class in only 23% of the generated sentences, validating the use of corpus statistics as a proxy to existing word-class preference. Next, to test for priming, we fitted the data to a GLM model. In terms of the random-effect structure, the reported model contained a random slope for existing preference on the participant level and random intercepts for both participants and items (see Table 6).

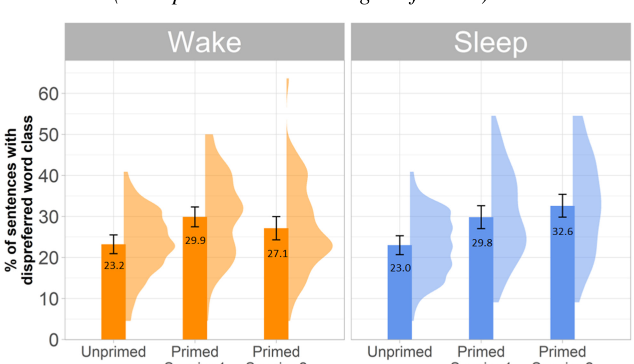
First, there was a main effect of existing word-class preference ($\beta = -0.034$, $SE = 0.159$, $z = -2.14$, $p = .032$), suggesting that words with an existing verb preference were more likely to be used in their dispreferred word class (i.e., as a noun) than those with an existing noun preference. This fits with the established finding that nouns (vs. verbs) are easier to produce (e.g., Dockrell et al., 2007; Gentner, 2006). Existing preference did not interact with the other two fixed effects.

Next, there was a main effect of Priming [i], which compared the unprimed control items with the primed items in Sessions 1 and 2 ($\beta = -0.151$, $SE = 0.021$, $z = -7.33$, $p < .001$). This provides support for our first prediction, indicating a clear priming effect overall such that reading a word-class ambiguous word in its dispreferred word class increased its likelihood of being used in that specific word class ($M_{(S1+S2)/2-\text{Unprimed}} = +6.76\%$). There was no main effect of Priming [ii] ($\beta = -0.001$, $SE = 0.034$, $z = -0.02$, $p = .983$), so our second prediction regarding a temporal decay in priming is not supported.

There was also a main effect of Group ($\beta = 0.063$, $SE = 0.031$, $z = 2.05$, $p = .040$), but this was qualified by an interaction with Priming [i] ($\beta = -0.050$, $SE = 0.021$, $z = -2.37$, $p = .018$) and [ii] ($\beta = 0.073$, $SE = 0.034$, $z = 2.16$, $p = .031$). The first interaction indicates that the priming effect was larger in the sleep ($M_{(S1+S2)/2-\text{Unprimed}} = +8.05\%$) than the wake ($M_{(S1+S2)/2-\text{Unprimed}} = +5.3\%$) group when collapsed across sessions; however, this needs to be interpreted alongside the second interaction (i.e., Group × Priming [ii]), which suggests that the change in priming magnitude between sessions was different between the two groups, with the sleep group showing an increase ($M_{S2-S1} = +2.75\%$) and the wake group a reduction ($M_{S2-S1} = -2.80\%$). In other words, the former interaction (Group × Priming [i]) was at least partly driven by group differences between sessions. To sum up, mirroring the associate production data from Experiment 1, we found clear evidence for our key prediction that an overnight sleep is involved in the maintenance of context-specific lexical information.

Figure 5

Mean Percentage of Dispreferred Word Class Being Used in the Generated Sentences, Summarized Across Group and Priming Conditions (Collapsed Across Existing Preference)



Note. Error bars represent 95% within-subject CI (Morey, 2008). Density functions represent the distribution of participant means in each priming condition. See the online article for the color version of this figure.

Exploratory Analyses

Time-of-Day Effects. As in Experiment 1, we explored whether our finding was attributable to time-of-day effects by comparing the wake and sleep groups on their test performance in Session 1. As in the previous exploratory analysis, we fitted a generalized mixed-effect model to the sentence recognition data from Session 1, with Group (wake vs. sleep), Priming (Unprimed vs. PrimedSession1), and their interaction as the fixed effects. There was a main effect of Priming ($\beta = -0.26$, $SE = 0.06$, $z = -4.47$, $p < .001$), but no effect of Group ($\beta = 0.03$, $SE = 0.05$, $z = 0.73$, $p = .463$) and no Group by Priming interaction ($\beta = 0.01192$, $SE = 0.04$, $z = 0.27$, $p = .784$). The absence of a Group effect and a Group by Priming interaction mirrors the alertness measures (i.e., simple RT and SSS) and the exploratory analysis of Experiment 1, providing further assurance that our findings of greater contextual priming postsleep were unlikely to be attributable to time-of-day effects.

As a further test to potential time-of-day effects, we followed a reviewer's suggestion by splitting the unprimed control items between sessions. If time of day influenced the overall tendency to use a word's dispreferred word class, then this should be seen most clearly in the unprimed baseline condition. Therefore, we fitted a 2 (Session) × 2 (Group) Generalised Linear Mixed Effects (GLME) model to the unprimed condition, as in the procedure as before. It showed no main effects of Groups or Session ($ps > .40$) but a nonsignificant tendency toward an interaction ($p = .0543$), potentially reflecting weak time-of-day effects. Breaking this interaction down with emmeans showed that it was driven by a marginally significant difference between the two groups in Session 2 ($p = .0567$), with more dispreferred responses in the evening (wake) than the morning (sleep). However, importantly, this tendency in Session 2 was reversed in the primed conditions (see Figure 5), which generally showed more dispreferred responses in the morning (sleep) than the evening (wake). This means that if there was any circadian effect on performance, it acted to reduce the strength of the effect in the primed conditions rather than accentuate it. This interpretation was confirmed in a further exploratory analysis where we fitted the full data set to a 2 (Session: 1 vs. 2) × 2 (Priming: Primed vs. Unprimed) × 2 (Group: wake vs. sleep) GLME model. It revealed a main effect of Priming ($z = -7.27$, $p < .001$), a Priming by Group interaction ($z = -2.51$, $p = .012$), and a three-way interaction ($z = 2.84$, $p = .004$; see OSF [https://osf.io/z2ane/?view_only=ac75981a9ac3403b8e45c841a4f31ac5] for model output). These findings align perfectly with those from our preregistered analyses and provide further assurance that time-of-day effects are unlikely to underlie our findings.

Pairwise Comparisons. As per Experiment 1, we conducted a set of post hoc pairwise comparisons to facilitate interpretation of our findings. First, for the between-group comparison within

Table 6

Summary of the Generalized Linear Mixed-Effects Model Examining the Effects of Group, Existing Word Class Preference, and Priming Categories in Sentence Generation, Experiment 2

Fixed effects	Estimates	*SE*	*z* values	*p*
Intercept	−1.29	0.154	−8.39	<.001*
Existing word class preference (noun vs. verb preference)	−0.340	0.159	−2.14	.032*
Priming [i] (Unprimed vs. PrimedS1 + 2)	−0.151	0.021	−7.33	<.001*
Priming [ii] (PrimingS1 vs. PrimingS2)	−0.001	0.034	−0.02	.983
Group (wake vs. sleep)	0.063	0.031	2.05	.040*
Existing Preference × Group	0.040	0.049	0.80	.423
Priming [i] × Group	−0.050	0.021	−2.37	.018*
Priming [ii] × Group	0.073	0.034	2.16	.031*
Existing Preference × Priming [i]	−0.001	0.021	−0.03	.978
Existing Preference × Priming [ii]	−0.007	0.034	−0.20	.845
Existing Preference × Priming [i] × Group	0.016	0.021	0.77	.441
Existing Preference × Priming [ii] × Group	−0.039	0.034	−1.18	.237

Note. S1 = Session 1; S2 = Session 2.
* $p < .05$.

sessions, there was no significant difference in Session 1 ($p = .428$); however, in Session 2, the sleep group outperformed the wake group ($p < .001$). Finally, for performance across sessions within each Group, there was no significant difference in either the wake ($p = .125$) or the sleep group ($p = .128$).

Discussion

Experiment 2 investigated whether exposure to a word-class ambiguous word in its dispreferred word class would prime language users to subsequently use the word in that specific word class, and if it would, whether sleep is involved in its longer-term retention. Following the general design as Experiment 1, participants first read word-class ambiguous words (e.g., *loan*) in prime sentences where they were used in their dispreferred word class (e.g., "Brentwood Borough Council said it would loan individual projects between £1 m and £20 m to transform the borough"). Shortly afterward, participants completed a sentence generation task, where they came up with a sentence using a word-class ambiguous word. They were free to use the word as either a noun or verb. Twelve hours later, which included either daytime wakefulness or a period of overnight sleep, participants attempted sentence generation again to a different subset of items. In both the immediate and delayed sessions, we found clear evidence that previously experiencing a word-class ambiguous word in its dispreferred word class increased the likelihood of the word being used in that same word class. From here onward, we refer to this as the "word-class priming effect." Importantly, between Sessions 1 and 2, the change in priming magnitude was different between the wake and sleep participants, with those in the sleep group showing an increase but the wake group showing a reduction. In sum, findings from Experiment 2 aligned with those from associate production in Experiment 1, providing further evidence for the episodic context account. In the "General Discussion, " we consider this account in length, focusing on the contribution of episodic memory and sleep to day-to-day discourse comprehension, complemented by neuropsychological and computational findings.

General Discussion

Built upon our prior studies (Curtis et al., 2022; Gaskell et al., 2019), the two experiments presented in this article were intended to shed further light on the involvement of episodic memory and subsequent periods of sleep and wake in the processing and retention of context-specific lexical information. The current study goes some way beyond the two prior publications from our group and provides, for the first time, a critical mass of evidence that allows us to argue that episodic memory plays a general role in day-to-day language comprehension.

The starting point for our investigation was Gaskell et al. (2019), who reported that word-meaning priming for homonyms was prone to sleep-related effects. This provided the basis for an episodic context account, which argues that language comprehension leads to a contextually bound episodic representation that is prone to subsequent sleep-related memory effects. A key limitation of Gaskell et al. (2019), however, is that their study was based exclusively on homonyms (e.g., *bank*), leaving open the question of whether the effects of episodic updating and sleep are restricted to these relatively rare words, which make up no more than 7% of all the lexical items in the English language (Rodd et al., 2002). To fill this research gap, Curtis et al. (2022) examined whether word-meaning priming extends to nonhomonymic nouns, and they found preliminary evidence that these words are also susceptible to episodic updating effects. However, crucially, Curtis et al. only examined the short-term consequence of this effect (10–20 min postexposure); without the longer delays and the sleep manipulations that we used in the current article, Curtis et al. (2022) on its own does not provide strong evidence for the involvement of episodic memory; instead, their findings of short-term priming effects could be interpreted with reference to an alternative account such that priming reflects immediate alteration in lexical-semantic weights in long-term memory.

The current study tests the more discriminating prediction from the episodic account that word-meaning priming for nonhomonyms should remain across periods including sleep but decay across periods including wake. Moreover, we also tested the generality of word-meaning priming further, by assessing whether word-class usage is also susceptible to contextual priming, both in terms of the effect of priming in the short term (across 20-min delays) and across wake and sleep. Data from Experiments 1 and 2 revealed a coherent picture: We found evidence that encountering nonhomonyms (Experiment 1) and word-class ambiguous words (Experiment 2) in a sentential context can prime the subsequent interpretation and usage of the words in a way that is consistent with that specific context. Crucially, in associate production (Experiment 1) and sentence generation (Experiment 2), the contextual priming effects (i.e., word-meaning and word-class priming, respectively) were more robust and resistant to time-dependent decay over a period of sleep than an equivalent amount of daytime wakefulness. Although data from these two tasks were highly consistent with the episodic context account, findings from relatedness judgment (Experiment 1) were less so, in which no clear sleep-related effects were observed. Therefore, overall, our experiments demonstrated some clear, but not complete, evidence for the episodic context account.

Our study extends prior studies (Curtis et al., 2022; Gaskell et al., 2019) in several important ways, and so establishes a firm empirical basis for our episodic account of language comprehension. The two experiments provide a critical mass of evidence that allows us to argue that *all* content words encountered in naturalistic texts are prone to some degree of episodic updating and sleep-related effects. Of the two experiments, Experiment 2 is particularly revealing: First, the observation of word-class priming, which is in itself is a novel finding in the psycholinguistic literature, extends prior work by showing, for the first time, that contextual priming applies to not only nouns but also verbs, both in terms of short-term effects and in terms of delay periods including sleep but not wake. Second, it raises the possibility that contextual priming may not be restricted to the semantic level, but potentially extend to the morphosyntactic level, highlighting the possibility that episodic memory may capture a range of context-specific information. Third and finally, Experiment 2 departed from prior priming studies by using sentence generation as the outcome measure, which is arguably a more ecologically valid way (e.g., vs. associate production) of showing that contextual experience of a word can influence subsequent usage of that word. All in all, Experiment 2 provided compelling and further evidence that episodic updating is perhaps a general feature of language comprehension. Below, we first consider more thoroughly the overall relationship between the findings from the two experiments and the episodic context account before turning to the implications o f each individual experiment.

Episodic Memory and Contextual Priming

In this section, we put aside the sleep-related findings and focus on how episodic memory may contribute to contextual priming. The notion of episodic memory having a role to play in language processing is not new and has been formalized in prior theories concerned with, for example, single-word processing (e.g., Goldinger, 1998; Jiang & Forster, 2001; Hutchison, 2003; Nation, 2008; Tenpenny, 1995) and novel word learning (e.g., A. B. Nelson & Shiffrin, 2013; Bakker et al., 2014; Brown & Gaskell, 2014; Dumay & Gaskell, 2007; Mak et al., 2021b). The episodic context account, in contrast, posits the routine involvement of episodic memory in *day-to-day discourse comprehension* (Gaskell et al., 2019). Specifically, it hypothesizes that during comprehension, episodic memory contributes to the generation and maintenance of a new and temporary context-specific representation that binds together the words and concepts in the sentence/discourse. This representation is assumed to play a general role in guiding online comprehension, perhaps by supporting the retention of discourse memory and the construction of situation models (e.g., Altmann & Ekves, 2019; Graesser et al., 1997). In addition to these functions, the context-specific representation may also have the potential to bias future lexical processing—a prediction that was explicitly tested here: Using naturalistic and familiar language, we showed that exposure to words whose meanings and/or syntactic roles are modified, however subtly, by their sentential contexts could subsequently bias interpretation and usage of those words in a way consistent with the prior contexts. We interpreted these priming effects as being driven by the context-specific representations formed earlier during comprehension. Notably, this is not to say that these representations would supplant the use of long-term semantic knowledge in lexical processing; instead, they may serve as an additional source of information alongside long-term knowledge. Whether a context-specific representation would exert an influence, we believe, may depend on factors such as its quality and the nature of the lexical tasks. Regarding the first factor, we suggest that reliance on context-specific episodic representations should be stronger if these representations are of a higher quality (e.g., better-specified; F. Ferreira & Patson, 2007; Gilbert et al., 2021). In turn, this may lead to stronger contextual priming effects. Future studies evaluating this possibility are needed, perhaps by varying discourse quality with different frequencies of exposure (e.g., Betts et al., 2018). Regarding the second factor on task nature, we will revisit this point below where we consider the inconsistent findings from relatedness judgment and associate production in Experiment 1. To sum up, we attribute the contextual priming effects observed in our experiments to influences from the context-specific representations generated “on the fly” by episodic memory during comprehension (see also Jamieson et al., 2018, 2022 for computational evidence).

Now, turning to the nature of these context-specific representations: Are they a word-for-word copy of the discourse experienced? Or, are they an abstract, high-level representation of the gist? We believe it is unlikely to be the former because memories for surface details (e.g., exact wordings) tend to be rapidly forgotten after initial exposure (e.g., |Sachs, 1967, 1974). At present, we favor the latter view that a context-specific representation is likely to be an abstract representation of the gist, in line with the evidence that enduring memory for text tends to be gist-like (Fisher & Radvansky, 2018; Kintsch et al., 1990; Mak et al., 2023; Trabasso & van den Broek, 1985) and that such gist-like memory can subsequently lead to semantic priming (Woltz et al., 2015; see also Curtis et al., 2022). What is unclear, though, is the composition of this representation, or to phrase differently, what aspects of a discourse get captured in this representation (see also F. Ferreira & Patson, 2007)? A prior study with homonyms (Rodd et al., 2013) showed that the size of the word-meaning priming effect was not affected by whether the voice at comprehension matched the voice at test, suggesting that episodic information like an interlocuter’s voice may not be captured in the context-specific representation. However, a recent study (Ryskin et al., 2020) reported that if an interlocuter’s voice contributes crucially to the meaning of the discourse, the voice would be retained, subsequently influencing lexical processing. Together, these suggest that what is captured by a context-specific representation may well depend on how meaningful or central an element is to the discourse but on the whole will tend to abstract from surface detail. Thus, we are not arguing that contextual priming is underpinned by the formation of highly detailed, near-veridical representations of the language input (cf., episodic views of mental lexicon, Goldinger, 1998). Rather, the representations that we argued for are episodic in the sense that they are knowledge-driven representations of the linguistic event or episode, and make use of our conceptual knowledge relating to the words in the sentence (see also Renoult et al., 2019) alongside other relevant contextual aspects of the interaction.

Although the general contextual priming we see in our experiments is consistent with (and was predicted by) the episodic context account, we should return to another possible explanation. Earlier findings of word-meaning priming for homonyms were originally explained in terms of an immediate alteration account (e.g., Gilbert et al., 2018). By this explanation, encountering a sentential context that was consistent with one meaning of the homonym would lead to direct and immediate changes in the balance between those meaning representations in long-term lexical-semantic memory (Rodd et al., 2013, 2016), making the primed meaning more readily available in the future. Such an account could easily be extended to a wider range of circumstances to fit the priming we see here for nonhomonyms and noun–verb ambiguities. For instance, a word-class ambiguous word may be associated with two sets of word-class information in long-term memory, one for noun and one for verb (e.g., Caramazza & Miozzo, 1997; Roelofs et al., 1998). When these words are experienced in a specific word class, the corresponding information may be activated, making it more accessible than before and hence increasing the subsequent likelihood of that word class being used. As discussed in Curtis et al. (2022), a similar argument can be made for aspects of meaning in nonhomonyms. Therefore, the existence of word-meaning priming in nonhomonyms and word-class priming is *not* evidence against an immediate alteration account. To further understand the mechanisms with which contextual priming arises (e.g., via a contextually specific episodic representation or via direct adjustments to long-term semantic memory), we need to assess the time course of contextual priming over wake and sleep.

Sleep and the Consolidation of Context-Specific Representations

The episodic context account predicts that if a context-specific representation is of an episodic nature, it should be more likely to be consolidated over sleep, just like any other newly acquired episodic memories. For instance, randomly paired words (e.g., *friend–palace*) have no preexisting associations, so encoding them in paired-associate learning will likely result in memories of an episodic nature (as opposed to semantic). It is well-documented that these episodic memories are typically better remembered after sleep than after wakefulness (e.g., Mak & Gaskell, 2023; Payne et al., 2012; Plihal & Born, 1997), potentially because of these new memories being stabilized and strengthened over sleep-related consolidation (see Inostroza & Born, 2013; Paller et al., 2021; Rasch & Born, 2013 for reviews). A consequence of this is enhanced accessibility and greater resistance to decay (e.g., Diekelmann et al., 2009; Squire et al., 2015). Our findings of more robust contextual priming postsleep (vs. postwake) suggest that the context-specific representations underpinning these effects have been stabilized over sleep, maintaining their ability to bias lexical processing 12 hr later. This might reflect that sleep actively consolidated these episodic representations (Gaskell et al., 2019) or passively protected these representations from wakeful interference, reducing their rate of decay (Jenkins & Dallenbach, 1924; Yonelinas et al., 2019). While our experiments were neither designed nor intended to tease apart an active and a passive account of sleep, the latter cannot explain the finding from Gaskell et al. (2019; Experiment 2), where participants showed stronger word-meaning priming in homonyms 24 hr after sentence exposure if the first (vs. second) half of the 24 hr was filled with sleep. In other words, a sleep opportunity soon after exposure appeared to make contextual priming robust, such that there was little or no decay in priming in the subsequent period awake. Given this, we *favor* the interpretation that sleep actively consolidated the context-specific representations, leading to more robust contextual priming. However, as the 24-hr finding from Gaskell et al. (2019) was based on homonyms, it remains to be seen if it generalizes to nonhomonyms or word-class ambiguous words. Future work is needed to shed light on the contribution of active and passive mechanisms to the current findings of stronger contextual priming effects postsleep. Nonetheless, both types of mechanisms would be compatible with the central tenet of our episodic account that episodic memories are routinely formed during comprehension and act alongside established lexical knowledge to guide future comprehension.

Returning to the immediate alteration account, there are several aspects of the current results that are hard to accommodate by such an account. First, contextual priming typically decays over time awake. This temporal decay was seen in prior studies (e.g., Rodd et al., 2013, 2016) and in our wake groups in both experiments [e.g., word-class priming decreased from Sessions 1 ($M = +6.7\%$) to 2 ($M = +3.7\%$)]. Such a decay is hard to explain by an immediate alteration account, which posits that language exposure can lead to an instant update of existing lexical-semantic knowledge: If such knowledge has already been updated, then why should that knowledge show preferential decay, as opposed to less recent updates to long-term memory? Relatedly, if the new information contained in a word's contextual interpretation has already been integrated into long-term memory networks, then why would sleep lead to stabilization of this knowledge? A model that explains declarative memory in terms of complementary learning systems (McClelland et al., 1995) has no obvious mechanism by which neocortical memories can decay or be consolidated, given that the intended end state has already been achieved. Therefore, an immediate alteration account does not predict sleep–wake differences, and nor can it easily accommodate them. In contrast, an episodic context account can easily explain why contextual priming may decay over time awake and why it is prone to sleep-related effects. As in other hippocampus-dependent episodic memories, the context-specific representations that underlie contextual priming are susceptible to time-dependent decay (Hardt et al., 2013); however, if a sleep opportunity soon follows encoding, the episodic representations can be (at least partially) consolidated into the neocortical store via offline replay (Bendor & Wilson, 2012), increasing the subsequent likelihood of these representations exerting an influence on lexical processing. In sum, our finding that contextual priming was more robust after sleep aligns well with an episodic context account but poses a challenge to the immediate alteration model originally proposed by Rodd et al. (2013, 2016). Notably though, this is not to say the episodic context and immediate alteration accounts are mutually exclusive; perhaps, under certain circumstances, some immediate alteration mechanisms may be involved in shaping lexical knowledge (e.g., Mak, 2019; Tse et al., 2007); however, the current sleep-related findings sit more comfortably with the episodic context account.

Neural Underpinnings

Although our experiments were strictly behavioral, it is important to consider how our findings may be explained on a neurocognitive level. As proposed by Gaskell et al. (2019), the context-specific representations generated during comprehension may be underpinned by the hippocampus. Two strands of evidence support this view. First, while patients with hippocampal amnesia can perform on par with healthy controls on simple comprehension tasks involving encapsulated sentences (e.g., predicting the upcoming referents; Brown-Schmidt et al., 2021), clear deficits emerge when context-specific information needs to be integrated across sentences (see Duff & Brown-Schmidt, 2012, 2017 for reviews). For instance, during listening comprehension, these patients struggle with linking pronouns (e.g., *she, that man*) with the intended referents that appeared in the preceding sentence (Covington et al., 2020; Duff et al., 2011; Kurczek et al., 2013). Potentially, this deficit arises because the hippocampus is crucial to associative memory (e.g., Brasted et al., 2003; Mayes et al., 2007; Warren & Duff, 2014), enabling arbitrary associations between a pronoun and an intended referent to be encoded rapidly. In the case of a context-specific representation, it can be conceptualized as a kind of associative memory, as it presumably binds together words and concepts in a discourse (akin to paired-associate learning; Gaskell et al., 2019). Given its associative nature, it seems reasonable to infer that the hippocampus plays a crucial role in the generation and maintenance of these context-specific representations during comprehension (see also Blank et al., 2016; Pu et al., 2020). A second strand of evidence implicating a role of hippocampus is that sleep is known to support the consolidation of hippocampus-dependent memory (e.g., Marshall & Born, 2007; Wixted & Cai, 2013). In both of our experiments, contextual priming was more robust and resistant to decay after sleep than after wakefulness, suggesting that the context-specific representations were prone to sleep-related consolidation effects. A reasonable assumption, then, is that these representations are at least partially supported by the hippocampus.

Computational Models of Semantics

Findings from our experiments have significant implications for computational models of semantics, which can be broadly divided into “prototype” and “instance” models. The former, such as word2-vec (Mikolov et al., 2013) and Latent Semantic Analysis (Landauer & Dumais, 1997), attempt to create a single representation for a word’s meaning, averaged across the linguistic contexts in which they appear (e.g., Beekhuizen et al., 2021; see Kumar, 2021 for a review). Instance models, on the other hand, store a unique representation of each language encounter, allowing the models to subsequently construct a words meaning on the fly when supplied with a retrieval cue (Hintzman, 1984; Jamieson et al., 2018, 2022). These two types of computational models are highly relevant here, such that long-term semantic knowledge may resemble a prototype while the episodic context-specific representations formed during language comprehension may be unique instances. As we have argued above, both may influence lexical processing/usage. Less clear, though, is how prototype and instance models are related. Potentially, over time, instances may sum to a prototype that maintains some degree of malleability; in other words, representations with some characteristics of both instances and prototypes might emerge. This view is similar to the “memory system reorganization” framework such that memory traces are never set in stone but are instead malleable and change as a result of new input and instances across the lifetime (Moscovitch & Gilboa, 2022). This suggests that prototype and instance models may not be mutually exclusive, and some kind of integration of the two may provide us with a powerful theoretical framework to understand the human language system (see Reid & Jamieson, 2023). Now, turning to the role of sleep: Instance models such as Jamieson et al.’s (2018) are grounded on episodic memory, so they can be easily extended to capture the finding of stronger contextual priming after sleep such that an instance, presumably stored in the hippocampus, is strengthened over sleep but less so over wake (Moscovitch et al., 2005). On the other hand, some prototype models such as word2vec are grounded on learning mechanisms based on prediction errors (Mandera et al., 2017), so as they stand, they are less readily extendable to capture the effect of sleep in maintaining contextual priming. Future work is needed to examine how such models may explain sleep-related effects and how prediction errors may relate to sleep-related consolidation.

Contextual Priming: Nature of Lexical Task

Experiment 1 was modeled upon Gaskell et al. (2019; Experiment 1); the key differences are that they used homonyms as the targets (instead of nonhomonyms), and associate production was their sole outcome measure. Data from our associate production task were highly comparable to theirs, showing that word-meaning priming was more resistant to decay after sleep (vs. after wakefulness). However, in our relatedness judgment data, there was no evidence of sleep affecting word-meaning priming. At present, it is unclear why such a discrepancy arose. This was complicated by the fact that no other study, to the best of our knowledge, has used relatedness judgment to index word-meaning priming over longer delays (for short-term evidence see Cai et al., 2017; Curtis et al., 2022; Gilbert et al., 2018), so we can only speculate at this point. First, the nature of the two tasks is different—relatedness judgment only required a binary decision while associate production required active generation of a response. Evidence from other lines of memory research suggests that procedures requiring active response generation, such as free and cued recall, are more likely to reveal sleep-related benefits in declarative memories than procedures requiring a forced-choice judgment (e.g., recognition; Berres & Erdfelder, 2021; Diekelmann et al., 2009). This difference has been observed in sleep studies concerned with eyewitness (Morgan et al., 2019; Tamminen, 2021), false (Newbury & Monaghan, 2019), emotional (Lipinska et al., 2019; Schäfer et al., 2020), and discourse memory (Mak et al., 2023). Potentially, this is related to active generation being supported by overlapping neurocognitive underpinnings as sleep-related consolidation (i.e., hippocampus; Diekelmann et al., 2009). Second, our relatedness judgment task required speeded responses while our associate production (as well as sentence generation in Experiment 2) had no time pressure. It is possible that reliance on context-specific representations (vs. long-term semantic memory) to aid task performance may vary due to time pressure (see also Hill et al., 2002;

Hoedemaker & Gordon, 2017; Wilding & Herron, 2006). Based on our findings, it seems that speeded tasks might have limited the likelihood of participants (implicitly and/or explicitly) retrieving the context-specific representation to aid task performance. In turn, this might have reduced the magnitude of contextual priming and hence the likelihood of any sleep-related benefit from arising. In line with this, data from our untimed tasks (i.e., associate production and sentence generation) showed clear contextual priming overall, while in our timed task (i.e., relatedness judgment), evidence for priming was only found in RT but not in accuracy. This hints at the possibility that the context-specific representations were more influential in lexical processing when participants had sufficient time to (implicitly and/or explicitly) retrieve those representations.[2] Furthermore, we note that it is not necessarily the case that word-meaning priming, as indexed by relatedness judgment and associate production, must be attributable to one single mechanism (Gaskell et al., 2019). While performance on these tasks may be influenced by a context-specific representation generated at comprehension, the precise mechanism with which the representation exerts an effect may vary across tasks, depending on factors such as the retrieval process. Future studies are needed to explore the time course with which a context-specific representation exerts an influence on lexical processing, and how contextual priming may be affected by factors such as task demands and time pressure.

Word-Class Priming: Semantic or Morphosyntactic?

Experiment 1 and all prior studies testing the episodic context account (e.g., Curtis et al., 2022; Gaskell et al., 2019) used exclusively nouns as the targets. Experiment 2 is the first of this type to make use of verbs—whose processing on the neurocognitive levels may be different from nouns due to inherent semantic differences (Alyahya et al., 2018; Vigliocco et al., 2011) and differences in the required argument structures (Friederici & Weissenborn, 2007). It is worth noting that existing word-class preference (noun preference vs. verb preference) did not interact with the other independent variables (priming conditions and group) in our data. This suggests that a word-class ambiguous word, regardless of its existing preference, is prone to word-class priming and subsequent sleep-related effects. This is important because the episodic context account posits that *all* kinds of words are subject to contextual priming as long as the prior context modifies their interpretation in some way. Although we found clear evidence for word-class priming, it is unclear the extent to which such priming was semantic and/or morphosyntactic in nature. As fleshed out in the introduction, the meaning of a word-class ambiguous word is typically different, albeit subtly, when it crosses word class. Therefore, it is very possible that word-class priming was driven by such semantic subtlety being maintained in the episodic representation, as in the case of nonhomonyms in Experiment 1. Nevertheless, we cannot rule out the possibility that word-class priming was driven by morphosyntactic information being captured by episodic memory; it is possible that such information constitutes a part of the context-specific representation generated at comprehension, which may, in turn, bias language users toward that specific word class. In some way, this is similar to the well-established phenomenon known as *syntactic priming*, where exposure to a specific syntactic construction will prime language users to reuse that specific construction (Bock, 1986; Estival, 1985; Giles & Powesland, 1975; Levelt & Kelter, 1982; Schenkein, 1980; Tannen, 1989). However, some research suggested that *word-specific* syntactic priming is a relatively short-lived phenomenon, typically lasting no more than half an hour (Branigan et al., 1999; Hartsuiker et al., 2008; Wheeldon & Smith, 2003; cf. Kaschak et al., 2011). The fact that we observed word-class priming 12 hr after sentence exposure casts some doubt over the possibility of this effect being purely morphosyntactic in nature. Furthermore, a few neuropsychological studies have shown that syntactic priming was *intact* in patients with hippocampal damage (Knowlton et al., 1992; Schmolck et al., 2000; V. S. Ferreira et al., 2008), hinting that priming on the syntactic level is unlikely to be supported by the hippocampus, and hence, affected by sleep-related effects. In light of these different strands of evidence, we favor the interpretation that word-class priming is likely to be primarily driven by semantics. However, before any firm conclusion can be drawn, future work is needed to further explore the relative contribution of semantics and morphosyntax to word-class priming in both the short- and long-term.

What Do We Gain From Context-Specific Representations?

We have provided evidence that contextually bound episodic representations may be generated during language comprehension that can, in turn, bias subsequent lexical processing. A key question then arises: What purpose might these context-specific representations fulfill? We propose that they serve at least three functions. First, these representations ensure that the mental lexicon can adapt to changes. For example, when familiar words (e.g., *tweet*) acquire new meanings (e.g., a short message on Twitter; Rodd et al., 2012), the context-specific representations could store such information temporarily so that it can be later consolidated into long-term lexical knowledge. This guarantees a certain degree of plasticity in the mental lexicon, while protecting it against the possibility of existing lexical-semantic knowledge (e.g., *tweet* = bird chirp) or distributional statistics being overwritten (i.e., catastrophic interference; McClelland et al., 1995). Second, these context-specific representations might contribute to comprehension by facilitating (a) integration across sentences (e.g., linking pronouns with referents) and (b) discourse retention over time. The former enables comprehenders to establish coherence within a discourse, providing a key foundation for online comprehension (Kintsch & van Dijk, 1978). Similarly, the latter helps maintain discourse information, allowing comprehension to operate across time, from seconds to hours, and possibly days and weeks (Fisher & Radvansky, 2018). Third and finally, these context-specific representations may allow language users to better deal with repetition, which is a ubiquitous feature in naturalistic texts. For example, in an article about river systems, all instances of *bank* will likely refer to river banks; if readers revert to its dominant interpretation (i.e., financial bank) after each

[2] As suggested by an anonymous reviewer, we conducted a set of exploratory analyses where we categorized the relatedness judgment data into "fast" and "slow" responses via various splitting methods (e.g., median split on a participant level). These analyses, available on OSF (https://osf.io/z2ane/?view_only=ac75981a9ac3403b8e45c841a4f31ac5), found no significant main effect of fast versus slow responses and this did not interact with Group and Priming [ii], in line with the preregistered analyses. These null findings neither support nor contradict our suggestion that speeded tasks may have reduced the influence from episodic memory and sleep.

instance, reading efficiency would be greatly compromised. In contrast, with a context-specific representation in place, comprehenders could adhere to the context-appropriate interpretation, allowing comprehension to proceed smoothly. In other words, a context-specific representation may help comprehenders to deal with the clustering inherent in language (Myslín & Levy, 2016) and to overcome the laborious task of constantly updating expectations during comprehension. To sum up, a context-specific representation enables language users to cope with the highly dynamic and versatile nature of language inputs.

Conclusion

When we encounter a word in discourse, we make use of the context to determine its precise meaning and to update our expectations so that subsequent processing of the same word can be facilitated. The episodic context account (Gaskell et al., 2019) predicts that at the point of comprehension, episodic memory should come into play by generating a context-specific representation that binds different elements (e.g., words and concepts) in the discourse together, perhaps at a somewhat abstract level of meaning. This representation serves to guide comprehension online and provide an additional source of information (on top of long-term semantic memory) for future lexical processing. Evidence from the two experiments generally supports this view: Exposure to words whose meanings were modified by their surrounding contexts biased subsequent interpretation and usage of those words in a way consistent with the prior contexts. Furthermore, in the active production tasks, such contextual priming was found to be more resistant to decay after sleep (vs. wakefulness), highlighting the episodic nature of these representations and the role of sleep in updating lexical knowledge. Importantly, these findings were observed in both nonhomonymic nouns (e.g., *bathtub*; Experiment 1) and word-class ambiguous words (e.g., *loan*; Experiment 2), providing evidence for the proposal that each and every encounter of a word in naturalistic language constitutes a new hippocampal learning episode that, if consolidated into long-term memory, can fine-tune our lexical knowledge and improve our predictive model of language use. Our research represents an important step in understanding the mechanisms that underlie effective communication across the lifespan in the face of highly variable and changing linguistic inputs, providing for the first time a sound empirical basis to underpin our theoretical framework in which we can understand the flexibility of the human language system that is seen across a range of populations and psycholinguistic paradigms.

References

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness–eveningness questionnaire: A reduced scale. *Personality and Individual Differences*, *12*(3), 241–253. https://doi.org/10.1016/0191-8869(91)90110-W

Ahrens, K. (1999). The mutability of noun and verb meaning. In Y. Yin (Ed.), *Chinese languages and linguistics V: Interactions in language* (pp. 335–548). Academia Sinica.

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

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018). Noun and verb processing in aphasia: Behavioural profiles and neural correlates. *NeuroImage: Clinical*, *18*, 215–230. https://doi.org/10.1016/j.nicl.2018.01.023

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, *21*(8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ashton, J. E., Cairney, S. A., & King, B. R. (2021). Future-relevant memories are not selectively strengthened during sleep. *PLoS One*, *16*(11), Article e0258110. https://doi.org/10.1371/journal.pone.0258110

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116–130. https://doi.org/10.1016/j.jml.2014.03.002

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using qualtrics. *Behavior Research Methods*, *47*(4), 918–929. https://doi.org/10.3758/s13428-014-0530-7

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), 1–48. https://doi.org/10.18637/jss.v067.i01

Beekhuizen, B., Armstrong, B. C., & Stevenson, S. (2021). Probing lexical ambiguity: Word vectors encode number and relatedness of senses. *Cognitive Science*, *45*(5), Article e12943. https://doi.org/10.1111/cogs.12943

Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, *15*(10), 1439–1444. https://doi.org/10.1038/nn.3203

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin*, *147*(12), 1309–1353. https://doi.org/10.1037/bul0000350

Betts, H. N., Gilbert, R. A., Cai, Z. G., Okedara, Z. B., & Rodd, J. M. (2018). Retuning of lexical-semantic representations: Repetition and spacing effects in word-meaning priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1130–1150. https://doi.org/10.1037/xlm0000507

Bird, H., Howard, D., & Franklin, S. (2000). Why is a verb like an inanimate object? Grammatical category and semantic category deficits. *Brain and Language*, *72*(3), 246–309. https://doi.org/10.1006/brln.2000.2292

Blank, I., Duff, M., Brown-Schmidt, S., & Fedorenko, E. (2016). *Expanding the language network: Domain-specific hippocampal recruitment during high-level linguistic processing*. bioRxiv. https://doi.org/10.1101/091900

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

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

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review*, *6*(4), 635–640. https://doi.org/10.3758/BF03212972

Brasted, P. J., Bussey, T. J., Murray, E. A., & Wise, S. P. (2003). Role of the hippocampal system in associative learning beyond the spatial domain.

Brain: A Journal of Neurology, *126*(5), 1202–1223. https://doi.org/10.1093/brain/awg103

Brown, H., & Gaskell, M. G. (2014). The time-course of talker-specificity and lexical competition effects during word learning. *Language, Cognition and Neuroscience*, *29*(9), 1163–1179. https://doi.org/10.1080/23273798.2014.916409

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), Article 9. https://doi.org/10.5334/joc.10

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73–101. https://doi.org/10.1016/j.cogpsych.2017.08.003

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, *64*(3), 309–343. https://doi.org/10.1016/S0010-0277(97)00031-0

Chiarello, C., Liu, S., Shears, C., & Kacinik, N. (2002). Differential asymmetries for recognizing nouns and verbs: Where are they? *Neuropsychology*, *16*(1), 35–48. https://doi.org/10.1037/0894-4105.16.1.35

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

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

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

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*(5), 309–321. https://doi.org/10.1016/j.smrv.2008.08.002

Dockrell, J. E., Lindsay, G., Connelly, V., & Mackie, C. (2007). Constraints in the production of written text in children with specific language impairments. *Exceptional Children*, *73*(2), 147–164. https://doi.org/10.1177/001440290707300202

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

Duff, M. C., & Brown-Schmidt, S. (2017). Hippocampal contributions to language use and processing. In D. Hannula, & M. C. Duff (Eds.), *The hippocampus from cells to systems: Structure, connectivity, and functional contributions to memory and flexible cognition* (pp. 503–536). Springer.

Duff, M. C., Hengst, J. A., Gupta, R., Tranel, D., & Cohen, N. J. (2011). Distributed impact of cognitive-communication impairment: Disruptions in the use of definite references when speaking to individuals with amnesia. *Aphasiology*, *25*(6–7), 675–687. https://doi.org/10.1080/02687038.2010.536841

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

Estival, D. (1985). Syntactic priming of the passive in English. *Text—Interdisciplinary Journal for the Study of Discourse*, *5*(1–2), 7–22. https://doi.org/10.1515/text.1.1985.5.1-2.7

Fausey, C. M., Yoshida, H., Asmuth, J., & Gentner, D. (2006). The verb mutability effect: Noun and verb semantics in English and Japanese. In R. Sun, & N. Miyake (Eds.), *Proceedings of the twenty-eighth annual meeting of the cognitive science society* (pp. 214–219). The Cognitive Science Society.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Ferreira, V. S., Bock, K., Wilson, M. P., & Cohen, N. J. (2008). Memory for syntax despite amnesia. *Psychological Science*, *19*(9), 940–946. https://doi.org/10.1111/j.1467-9280.2008.02180.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

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain Research*, *1146*, 50–58. https://doi.org/10.1016/j.brainres.2006.08.038

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

Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, *4*(2), 161–178.

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek, & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). Oxford University Press.

Gentner, D., & France, I. M. (1988). The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution* (pp. 343–382). Morgan Kaufmann.

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*(January), Article 104188. https://doi.org/10.1016/j.jml.2020.104188

Gilbert, R. A., & Rodd, J. M. (2022). Dominance norms and data for spoken ambiguous words in British English. *Journal of Cognition*, *5*(1), Article 4. https://doi.org/10.5334/joc.194

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. Academic Press.

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*(1), 163–189. https://doi.org/10.1146/annurev.psych.48.1.163

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences*, *17*(3), 111–120. https://doi.org/10.1016/j.tics.2013.01.001

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, *58*(2), 214–238. https://doi.org/10.1016/j.jml.2007.07.003

Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming—Differentiation by event-related potentials. *International Journal of Psychophysiology*, *44*(3), 197–218. https://doi.org/10.1016/S0167-8760(01)00202-1

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods Instruments and Computers*, *16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology*, *10*(4), 431–436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

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

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544

Inostroza, M., & Born, J. (2013). Sleep for preserving and transforming episodic memory. *Annual Review of Neuroscience*, *36*(1), 79–102. https://doi.org/10.1146/annurev-neuro-062012-170429

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V..2013). The TenTen corpus family. In *Proceedings of the 7th International Corpus Linguistics Conference*. (pp. 125–127). Lancaster University.

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

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology*, *1*(3), 174–183. https://doi.org/10.1038/s44159-022-00025-3

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology*, *35*(4), 605–612. https://doi.org/10.2307/1414040

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

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*(3), 381–388. https://doi.org/10.3758/s13421-010-0042-3

Kersten, A. W., & Earles, J. L. (2004). Semantic context influences memory for verbs more than memory for nouns. *Memory and Cognition*, *32*(2), 198–211. https://doi.org/10.3758/BF03196852

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

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

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, *3*(3), 172–179. https://doi.org/10.1111/j.1467-9280.1992.tb00021.x

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin and Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Kurczek, J., Brown-Schmidt, S., & Duff, M. C. (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

Kutas, M., DeLong, K., & Smith, N. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.

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

Lenth, R. V., Bolker, B., Buerkner, P., Gine-Vazquez, I, Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means (R package Version 1.8.6). [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78–106. https://doi.org/10.1016/0010-0285(82)90005-6

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

Lipinska, G., Stuart, B., Thomas, K. G. F., Baldwin, D. S., & Bolinger, E. (2019). Preferential consolidation of emotional memory during sleep: A meta-analysis. *Frontiers in Psychology*, *10*, Article 1014. https://doi.org/10.3389/fpsyg.2019.01014

Lo, J. C., Dijk, D. J., & Groeger, J. A. (2014). Comparing the effects of nocturnal sleep and daytime napping on declarative memory consolidation. *PLoS One*, *9*(9), Article e108100. https://doi.org/10.1371/journal.pone.0108100

Lorenzetti, R., & Natale, V. (1996). Time of day and processing strategies in narrative comprehension. *British Journal of Psychology*, *87*(2), 209–221. https://doi.org/10.1111/j.2044-8295.1996.tb02586.x

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

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

Mak, M. H. C. (2021). Children's motivation to learn at home during the COVID-19 pandemic: Insights from Indian parents. *Frontiers in Education*, *6*(October), Article 744686. https://doi.org/10.3389/feduc.2021.744686

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023). *Recall and recognition of discourse memory across sleep and wake*. https://doi.org/10.31234/osf.io/6vqh9

Mak, M. H. C., & Gaskell, M. G. (2023). *Effects of sleep and retrieval practice on verbal paired-associate learning across 12 and 24 hr*. https://doi.org/10.31234/osf.io/phe5j

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

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021b). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language*, *118*. 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

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences*, *11*(10), 442–450. https://doi.org/10.1016/j.tics.2007.09.001

Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, *11*(3), 126–135. https://doi.org/10.1016/j.tics.2006.12.003

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190–1210. https://doi.org/10.1037/a0033812

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex:

Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. https://doi.org/10.48550/arXiv.1301.3781

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

Morgan, D. P., Tamminen, J., Seale-Carlisle, T. M., & Mickes, L. (2019). The impact of sleep on eyewitness identifications. *Royal Society Open Science*, *6*(12), Article 170501. https://doi.org/10.1098/rsos.170501

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annual Review of Psychology*, *67*(1), 105–134. https://doi.org/10.1146/annurev-psych-113011-143733

Moscovitch, M., & Gilboa, A. (2022). Has the concept of systems consolidation outlived its usefulness? Identification and evaluation of premises underlying systems consolidation. *Faculty Reviews*, *11*. https://doi.org/10.12703/r/11-33

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G., & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, *207*(1), 35–66. https://doi.org/10.1111/j.1469-7580.2005.00421.x

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

Nation, K. (2008). Learning to read words. *Quarterly Journal of Experimental Psychology*, *61*(8), 1121–1133. https://doi.org/10.1080/17470210802034603

Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill. *NPJ Science of Learning*, *2*(1), 1–3. https://doi.org/10.1038/s41539-017-0004-7

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, *120*(2), 356–394. https://doi.org/10.1037/a0032020

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Newbury, C., & Monaghan, P. J. (2019). When does sleep affect veridical and false memory consolidation?: A meta-analysis. *Psychonomic Bulletin and Review*, *26*(2), 387–400. https://doi.org/10.3758/s13423-018-1528-4

Oakhill, J. (1986). Effects of time of day and information importance on adults' memory for a short story. *Quarterly Journal of Experimental Psychology*, *38*(3), 419–430. https://doi.org/10.1080/14640748608401606

Paller, K. A., Creery, J. D., & Schechtman, E. (2021). Memory and sleep: How sleep cognition can change the waking mind for the better. *Annual Review of Psychology*, *72*(1), 123–150. https://doi.org/10.1146/annurev-psych-010419-050815

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

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, *9*(4), 534–547. https://doi.org/10.1162/jocn.1997.9.4.534

Pu, Y., Cheyne, D., Sun, Y., & Johnson, B. W. (2020). Theta oscillations support the interface between language and memory. *NeuroImage*, *215*, Article 116782. https://doi.org/10.1016/j.neuroimage.2020.116782

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

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

Reid, J. N., & Jamieson, R. K. (2023). True and false recognition in MINERVA 2: Extension to sentences and metaphors. *Journal of Memory and Language*, *129*, Article 104397. https://doi.org/10.1016/j.jml.2022.104397

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic–episodic distinction. *Trends in Cognitive Sciences*, *23*(12), 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008

Rodd, J. M. (2019, February 27). *How to maintain data quality when you can't see your participants*. Observer. https://www.psychologicalscience.org/observer/how-to-maintain-data-quality-when-you-cant-see-your-participants

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., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory and Cognition*, *40*(7), 1095–1108. https://doi.org/10.3758/s13421-012-0209-1

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, 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, Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, *115*(3), 182–188. https://doi.org/10.1016/j.bandl.2010.07.005

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

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition*, *69*(2), 219–230. https://doi.org/10.1016/S0010-0277(98)00056-0

Ryskin, R., Ng, S., Mimnaugh, K., Brown-Schmidt, S., & Federmeier, K. D. (2020). Talker-specific predictions during language processing. *Language, Cognition and Neuroscience*, *35*(6), 797–812. https://doi.org/10.1080/23273798.2019.1630654

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, *2*(9), 437–442. https://doi.org/10.3758/BF03208784

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory and Cognition*, *2*(1), 95–100. https://doi.org/10.3758/BF03197498

Saussure, F. (1916). *Course in general linguistics*. Duckworth.

Schäfer, S. K., Wirth, B. E., Staginnus, M., Becker, N., Michael, T., & Sopp, M. R. (2020). Sleep's impact on emotional recognition memory: A meta-analysis of whole-night, nap, and REM sleep effects. *Sleep Medicine Reviews*, *51*, Article 101280. https://doi.org/10.1016/j.smrv.2020.101280

Schenkein, J. A. (1980). Taxonomy for repeating action sequences in natural conversation. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 21–47). Academic Press.

Schmolck, H., Stefanacci, L., & Squire, L. R. (2000). Detection and explanation of sentence ambiguity are unaffected by hippocampal lesions but are impaired by larger temporal lobe lesions. *Hippocampus*, *10*(6), 759–770.

https://doi.org/10.1002/1098-1063(2000)10:6<759::AID-HIPO1013>3.0.CO;2-A

Schweinberger, M. (2016). *Part-of-speech tagging with R*. http://martinschweinberger.de/docs/articles/PosTagR.pdf

Scullin, M. K. (2013). Sleep, memory, and aging: The link between slow-wave sleep and episodic memory changes from younger to older adults. *Psychology and Aging*, *28*(1), 105–114. https://doi.org/10.1037/a0028830

Siegel, J. M. (2021). Memory consolidation is similar in waking and sleep. *Current Sleep Medicine Reports*, *7*(1), 15–18. https://doi.org/10.1007/s40675-020-00199-3

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), Article a021766. https://doi.org/10.1101/cshperspect.a021766

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272–1278. https://doi.org/10.1038/nature04286

Talmy, L. (1975). Semantics and syntax of motion. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 4, pp. 181–238). Academic Press. https://doi.org/10.1163/9789004368828_008

Tamminen, J. (2021, November 4–7). *Sleep, not time of day, benefits eyewitness memory* [Conference presentation]. Psychonomic Annual Meeting, Online. https://adobeindd.com/view/publications/1d6c8884-a05a-4c81-9a4d-b5248f90174b/b39L/publication-web-resources/pdf/PS21_Abstract_Book_Combo.pdf

Tandoc, M. C., Bayda, M., Poskanzer, C., Cho, E., Cox, R., Stickgold, R., Schapiro, A. C., & King, B. R. (2021). Examining the effects of time of day and sleep on generalization. *PLoS One*, *16*(8), Article e0255423. https://doi.org/10.1371/journal.pone.0255423

Tannen, D. (1989). *Talking voices. Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review*, *2*(3), 339–363. https://doi.org/10.3758/BF03210972

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory and Cognition*, *22*(1), 111–126. https://doi.org/10.3758/BF03202766

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press. https://doi.org/10.2307/415483

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

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Voeten, C. C. (2020). *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. (R package Version 2.1) [Computer software]. https://CRAN.R-project.org/package=buildmer

Warren, D. E., & Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus*, *24*(8), 920–933. https://doi.org/10.1002/hipo.22279

Wheeldon, L., & Smith, M. (2003). Phrase structure priming: A short-lived effect. *Language and Cognitive Processes*, *18*(4), 431–442. https://doi.org/10.1080/01690960244000063

Wilding, E. L., & Herron, J. E. (2006). Electrophysiological measures of episodic memory control and memory retrieval. *Clinical EEG and Neuroscience*, *37*(4), 315–321. https://doi.org/10.1177/155005940603700409

Wixted, J. T., & Cai, D. J. (2013). Memory consolidation. In S. Kosslyn, & K. Ochsner (Eds.), *Oxford handbook of cognitive neuroscience* (Vol. 2, pp. 436–455). Oxford University Press.

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

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

(*Appendices follow*)

Appendix A

Mixed-Effects Models Examining Time-of-Day Effects in Session 1 (Experiment 1)

Fixed effects	Estimates	*SE*	*z*/*t* values	*p*
Relatedness judgment/accuracy				
Intercept	1.72	0.16	10.74	<.001*
Group (wake vs. sleep)	−0.004	0.07	−0.06	.951
Priming (Unprimed vs. PrimedS1)	−0.06	0.06	−0.99	.321
Group × Priming	0.022	0.04	0.61	.543
Relatedness judgment/RT				
Intercept	283.8	0.16	1,749.9	<.001*
Group (wake vs. sleep)	0.277	0.16	1.71	.087
Priming (Unprimed vs. PrimedS1)	0.384	0.16	2.37	.018*
Group × Priming	−0.005	0.16	−0.03	.974
Associate production				
Intercept	−0.62	0.17	−3.73	<.001*
Group (wake vs. sleep)	−0.004	0.09	−0.05	.963
Priming (Unprimed vs. PrimedS1)	−0.22	0.05	−4.31	<.001*
Group × Priming	0.05	0.03	1.46	.145

Note. [1] Both Group and Priming were coded using sum contrast. [2] Due to the logarithmic scale, the estimate and standard error of the RT analysis have been multiplied by 100 (excluding the intercept) to improve interpretability. RT = reaction time.
* $p < .05$.

(*Appendices continue*)

Appendix B

Word-Class Ambiguous Words and Their Priming Sentences in Experiment 2

List	Target word	Preferred word class	% of occurrences as a noun in corpus	Priming sentence Batch A	Priming sentence Batch B
1	Bait	Noun	83	My government actually has to bait people with catchy songs in order to do the thing that should be a civic duty.	In soccer it has nearly become acceptable to bait opposing fans, to chant and jeer at the other team's followers.
1	Bias	Noun	73	Critics of the statements argue that they unfairly bias the judge at the conclusion of the trial.	The man's lawyer, Pete Theodocion, contends that admitting the messages will bias the jury against his client.
1	Dust	Noun	90	You should dust my bookshelves, wash the windows and clean the carpets twice a week.	He fired her maid because she forgot to dust the master bedroom.
1	Flash	Noun	71	A shriek echoed around the hill as bright lights began to flash under the tree.	We have had cars sound their horn and flash their lights, just because we were keeping within the speed limit.
1	Joke	Noun	77	Even my friend, a Mexican, was stunned by the condition of the road; those 6 miles were so memorable that we still joke about it.	Even to the very end, he would joke about his languid body in an effort to ease the suffering of his loved ones.
1	Label	Noun	65	I wanted to label the cups to help everyone keep track of their own water cup throughout the day.	If a manufacturer wants to label their product gluten-free, they are required to run tests in a certified laboratory.
1	Loan	Noun	97	Brentwood Borough Council said it would loan individual projects between £1 m and £20 m to help transform the borough.	93% of elderly Americans would loan money to a family member in financial need, according to a recent online poll.
1	Punch	Noun	63	To fight off a great white shark, you need to punch it vehemently in the head; otherwise, your chance of survival is slim to none.	After the dough has risen once, punch it down with your fist; then, you can shape the loaves in a round shape.
1	Rebel	Noun	82	We must settle for nothing less than leaders who maturity as they rebel against the beliefs that no longer work.	Studying the population of adolescents revealed that many do not rebel against authority but maintain good relationships with parents and teachers.
1	Stamp	Noun	69	He says that it's a ridiculous requirement to have to sign and stamp documents for just about every process.	Lay the fabric over the design, and stamp the fabric to complement the embroidery.
1	Trick	Noun	86	His real purpose is to trick his way into your home to see what he can steal.	A dark pattern is a user interface carefully crafted to trick users into buying things they might not otherwise do.
1	Bend	Verb	35	I wasn't going too fast but probably faster than conditions warrant, for you never really know what's round the next bend in the road.	My father did not turn again as the car drove quickly off up the hill and disappeared round a bend in the road.
1	Divide	Verb	13	A society where the social divide between haves and have-nots has become a chasm is a society that breeds violence and brutality.	The combined effects of disenfranchisement laws, inmate population trends and economic realities perpetuate a racial divide in society.
1	Frown	Verb	26	His usually unguarded expression was marred by the deep frown of meditation he now wore.	She turned her face down while fiddling with her purse to hide her frown of disappointment.
1	Grasp	Verb	31	Horse riding and golf are no longer elitist in the United Kingdom, but within the grasp of the middle class.	Representatives from national governments will gather for the UN Environment Assembly next week; this means that a global plastics treaty is nearly within our grasp.
1	Grimace	Verb	39	For a moment, he hesitated, lips pulling down into a vaguely uncomfortable grimace as he shifted, legs shuffling against the floor.	Regan attempted a smirk, but had to settle for a grimace when the effort made his face flare in pain again.
1	Hail	Verb	37	Wind hit an estimated 140 mph in Tennessee and the storms carried torrential rain and golf-ball-sized hail.	A nest was considered storm-destroyed if it was flattened by wind or badly damaged by hail.
1	Laugh	Verb	29	Steve always liked a laugh at the right times but he was always very professional in his approach.	The communications officer reported, a slight tremor in his voice, as if he were stifling a laugh.

(*table continues*)

(*Appendices continue*)

Table (*continued*)

List	Target word	Preferred word class	% of occurrences as a noun in corpus	Priming sentence Batch A	Priming sentence Batch B
1	Plunge	Verb	33	IT firms were left with huge inventories and massive amounts of excess capacity, which triggered a plunge in IT-sector growth.	He was under water for at least 2 min, causing him to take in a lot of water and making his temperature plunge.
1	Rub	Verb	17	The pain startled him out of his thoughts, but a quick rub of the injury relieved the throbbing.	Some trees renew their attractive bark by peeling off the old; you can help the process by giving them a friendly rub as you pass.
1	Squeeze	Verb	22	"No, it's a great idea," he reached for her hand, enclosing it in his and giving it a little squeeze.	I am surprised when a long arm gives my shoulders a gentle squeeze and a quick kiss is placed upon my cheek.
1	Suspect	Verb	31	John Keating is the prime suspect in a crime he claimed he didn't commit, though most believe he did.	Police said they questioned three people in connection with the kidnapping so far and had identified a key suspect.
2	Consent	Noun	84	The notion that individuals suffering from mental illness must consent to treatment ignores how brain diseases like schizophrenia work.	To keep their jobs, Amazon's delivery drivers are required to consent to biometric monitoring, according to multiple reports.
2	Divorce	Noun	68	He was a notorious womanizer and his wife wanted to divorce him after one liaison too many.	His wife, Sheryl, always wanted to divorce him because he was too fat, didn't work out and would not stop eating junk food.
2	Dock	Noun	70	On Tuesday, a cruise liner attempted to dock in Bangkok but was denied permission; the liner had to go to Vietnam instead.	An unpiloted spacecraft carrying supplies and an experimental robot attempted to dock with the International Space Station Monday morning.
2	Favor	Noun	63	The ministry also approved a relaxation of some of the conditions attached to its initial approval, and this will for sure favor the developer.	Some politicians favor a graduate tax, which students pay back once they start earning a set amount following completion of their course.
2	Jail	Noun	83	An Italian man asks police to jail him because life at home with his wife is "unbearable."	Jessica stood up at her daughter's rapist's sentencing and asked Judge Pamela Brooks not to jail him, but help him seek treatment instead.
2	Quiz	Noun	89	Cops will quiz more than 50 people who are believed to have been at rule-busting Downing Street parties, the Met has confirmed.	Scotland Yard will quiz more than 50 people over no. 10 Partygate breaches this week.
2	Riot	Noun	95	The umpire correctly gave him "out," but 90,000 spectators disagreed and proceeded to riot until the game was stopped.	Indiana's legislators took to Twitter Wednesday to call for peace after Trump supporters began to riot and broke into the Capitol.
2	Sanction	Noun	76	More and more medical societies have begun to sanction members with penalties like suspension or revocation of their society membership.	Biden was asked if he planned to sanction or block the Russian imports, but he did not answer the question.
2	Toast	Noun	64	The grill has dual gas controls with a full range of temperatures, so you can sear burgers on one side and toast buns on the other.	Most mornings, we're lucky if we have time to eat a bowl of cereal or toast a slice of bread.
2	Transport	Noun	72	At the same time, companies which transport goods by ship need to reassure concerned investors that they are taking green issues seriously.	The vehicles were also used to transport illegal goods, such as alcohol and slot machines for gambling.
2	Volunteer	Noun	78	Occasionally, an employee may volunteer to work off the clock to complete a project or help other workers finish their jobs.	Today, the committee has 30 employees who volunteer their help for the meals-on-wheels service.
2	Burn	Verb	21	Electrical connections in the warming components in the jackets and vest can overheat, posing a burn hazard.	The severity of a burn depends on the intensity of the heat and time in contact with the heat.
2	Feed	Verb	31	The first things to try are offering your baby a feed and checking whether he/she needs a nappy change.	It is an extremely sad society if a mother has to dash off to sit in the toilet every time her baby needs a feed.
2	Guarantee	Verb	33	Trade unions are demanding a guarantee of the right to retire at 50 without conditions.	While I have tried my best, there is no guarantee that the data shown on this booking site are accurate.
2	Handle	Verb	16	It shows excellent research on his part and flawless writing so everyone can get a handle on the situation.	So no new troops yet requested to be sent to Iraq, but the United States will hold on to some of the troops until they get a handle on this situation.
2	Neglect	Verb	39	I think the council could do more to prevent this kind of vandalism and dispel the air of neglect which pervades this allotment.	The majority of the houses obtained by private landlords were now exhibiting signs of neglect and dilapidation.
2	Slip	Verb	25	Shaun began the event cautiously knowing that one slip on the loose surface could lose him the whole thing.	One slip and you could fall to your death, so walk carefully.
2	Spit	Verb	27	I brought up all this phlegm and spit into my mouth, and at first it was so, so foul I nearly choked.	Old Bruce is not happy to be reminded that he was once a porky loser who talks as if his mouth is full of spit and looks like a living doughnut.

(*table continues*)

(*Appendices continue*)

Table (*continued*)

List	Target word	Preferred word class	% of occurrences as a noun in corpus	Priming sentence Batch A	Priming sentence Batch B
2	Stir	Verb	12	The rare find of a 200-pound octopus created a stir of interest in the isolated fishing community.	For the first time since she'd arrived in Ethiopia, she felt a stir of pity for this young woman who was only a few months older than herself.
2	Tiptoe	Verb	31	She leaned her bike against the stone wall and stood on tiptoe to see what's happening behind the wall.	She stepped closer and raised herself on her tiptoe to kiss his cheek.
2	Vomit	Verb	34	The recovery position ensures that an unconscious person maintains an open airway so that any vomit or fluid will not cause choking.	There are between 70 and 150 deaths per year in the United Kingdom caused by suffocation, heart failure or choking on vomit.
2	Yawn	Verb	37	I spent this period of instruction trying to stifle a yawn and resisting saying how old-hat this all seemed.	I tried not to show my boredom, but my yawn was coming quicker and quicker.
3	Anchor	Noun	63	One needs to find a space to anchor the ship before applying for a license at the office next to the pier.	I know there's a cavern over there on that island, along with a spot to anchor the ship.
3	Auction	Noun	91	Arnold has announced to auction two watches he wore in his most recent film, with proceeds going to aid needy children.	The £20,000 that he was able to auction his art for will make a massive difference to disabled children lives.
3	Bail	Noun	61	It will be interesting to see how the Fed and other central banks will attempt to bail the world economy out of this COVID-19 storm.	Peter was not charged, but his father was so angry that he refused to bail him out, leaving him in the police station for 8 days.
3	Bandage	Noun	76	If a blister opens up, you might need to bandage it with a nonstick dressing.	If you notice any of the following signs, seek medical attention rather than trying to bandage the wound yourself.
3	Bargain	Noun	82	Kate Middleton regularly visits shops at Bicester Village and loves to bargain, it has been revealed.	A study has revealed that in Malaysia, men love to bargain more, but women get bigger discount.
3	Bomb	Noun	84	Ahmed Ressam subsequently admitted that he planned to bomb Los Angeles International Airport on 31st December 1999.	Terrorists planned to bomb United Kingdom's largest mosque "to get justice" for the Manchester attack.
3	Doubt	Noun	69	The awful truth is that I doubt the relaxation of the licensing laws will make much difference.	Though they have fine words to say about democracy, I think deep down they doubt the ability of the people to act wisely.
3	Dye	Noun	72	I found it hilarious that she wanted to cut her hair short and dye it red for the Marilyn role.	Max decided to dye his blond hair green because he was sick of people mixing him up with his twin brother.
3	Floss	Noun	78	Erupted wisdom teeth are more likely to develop plaque and cavities because they are harder to brush and floss properly.	This is the main reason why proper dental care, such as regular visits to the dentist and your effort to floss your teeth, is crucial.
3	Poison	Noun	66	A scientist who served 7 years in prison for trying to poison his wife has secured a job teaching ethics, university officials said today.	He tried to poison us like lower animals, like the mice that pester storybook villages.
3	Shelter	Noun	88	The wall is paved by bricks and filled with earth, and during war time it served to shelter people in the town from disaster.	Lying alone in a tiny cave barely large enough to shelter one person, I listen as the storm rages on.
3	Arrest	Verb	38	The internal report will provide valuable ammunition for the Hamiltons who have said they intend to sue Scotland Yard for unlawful arrest.	Stephen Maddox, 32, appeared before Judge George Cannon for an initial appearance after his arrest on the charges of drug possession.
3	Blame	Verb	19	Cynthia had made false accusations against him, made him go on the run and set him up to take the blame for her frauds.	The company's spindoctors are now working overtime to put the blame on everyone but themselves.
3	Blow	Verb	32	The wind started out as just a gentle breeze, but soon evolved into a strong blow.	Its weight massed in the crowns makes trees prone to toppling in a strong blow.
3	Bother	Verb	5	A woman who turned 104 last Thursday had just one wish for her birthday—she didn't want any fuss or bother.	She did the washing without complaining because she didn't want to be a bother.
3	Crawl	Verb	25	The prime minister is counting on the budget to pull the country out of the COVID-19-induced doldrums, with the economy beginning with a crawl this year.	She slowed from a run to something of a crawl as rain began to pour down on her and splash her in the face.
3	Growl	Verb	29	Make sure your children understand your pet's boundaries, such as not going near them when they are eating or understanding what a growl means.	My dog never bites anyone, as far as I can remember, but he is a master of the threatening growl.
3	Highlight	Verb	31	The festival has been a great success and we now hope that it will now become an annual event and a highlight in summer.	Outdoor screenings of classic movies are always the highlight of this annual event in Durham.

(*table continues*)

(*Appendices continue*)

Table (*continued*)

List	Target word	Preferred word class	% of occurrences as a noun in corpus	Priming sentence Batch A	Priming sentence Batch B
3	Howl	Verb	36	His mouth was bleeding and he let out a howl like a wolf that had just been shot by a hunter, before finally collapsing.	The only sound was the sound of a lone wolf's howl into the hopeless night.
3	Permit	Verb	34	Under the Marine Mammal Protection Act, it is illegal to "take" any marine mammal without a permit.	Every construction activity in the city should obtain a building permit released by the agency.
3	Snore	Verb	16	His head was leaning back, and every time he breathed a snore would erupt from his nostrils.	Matt let out an abnormally large snore and I imagined myself smothering him with his own pillow.
3	Wail	Verb	27	His substitution prompted a wail of anguish from the midfielder and tears to sting his eyes.	The end of the working day in the tea garden is marked by the wail of an air-raid siren.

Appendix C

Instructions for Sentence Generation (Experiment 2)

Page 1

Instruction 1: In each trial, you will be given an English word (e.g., swim). You job is to write down the first meaningful sentence that comes to your mind. The sentence MUST contain the word shown.

Instruction 2: The sentence you come up with must be grammatically correct and have a minimum of 25 characters—that is about 6–10 words. However, you are very welcome to write longer sentences.

Page 2

Your task is to write down the first sentence that comes to mind upon seeing the word shown. The sentence MUST contain that word (e.g., swim). Some acceptable sentences include "John and I went for a swim in the lake" and "We swam for 10 hr on Friday."

You can use the word either as a noun or as a verb in your sentence, as long as it is grammatically correct. For nouns, you are welcome to pluralize it if it is appropriate. For verbs, you are welcome to change the tense (e.g., swims, swam, swimming) to fit your sentence.

Finally, you MUST NOT start a sentence using the word shown. For instance, a sentence like "Swimming is good for you" is NOT acceptable. If you do, you will risk your reimbursement being reduced.

Page 3

Final points

- There are a total of 33 trials (i.e., 33 words). This task takes ∼12–16 min.
- In each trial, the Next button will only appear 12 s after the trial started. It's fine if you need more time.
- Do NOT write more than one sentence in each trial.
- Please do not leave full-screen mode during the task, or try to copy-and-paste (we have disabled copy-and-pasting).
- Finally, we encourage you NOT to overthink but to go with your first instinct. The first sentence that comes to mind is what we want.

The task begins on the next page.

Received August 11, 2022
Revision received March 10, 2023
Accepted April 25, 2023 ■