



## The relationship between sentence comprehension and lexical-semantic retuning

Rebecca A. Gilbert<sup>a,b,\*</sup>, Matthew H. Davis<sup>b</sup>, M. Gareth Gaskell<sup>c</sup>, Jennifer M. Rodd<sup>a</sup>

<sup>a</sup> University College London, London, UK

<sup>b</sup> MRC Cognition and Brain Sciences Unit, Cambridge, UK

<sup>c</sup> University of York, York, UK

### ARTICLE INFO

#### Keywords:

Language comprehension  
Lexical ambiguity  
Semantic ambiguity  
Priming  
Learning  
Sentence processing

### ABSTRACT

A single encounter with an ambiguous word (e.g. bark, ball) in the context of a less-frequent meaning (e.g. “Sally worried about how crowded the ball would be.”) can shift the later interpretation of the word toward the same subordinate meaning. This lexical-semantic retuning functions to improve future comprehension of ambiguous words. The present paper investigates the relationship between this form of learning and the specific processes that occur during sentence comprehension. One possibility is that lexical-semantic retuning occurs immediately upon hearing the ambiguous word, during initial meaning activation and selection, so priming should be strongest when the disambiguating context is provided *before* the ambiguous word (prior disambiguation). Alternatively, priming may relate to the degree of reinterpretation needed, which would predict maximal learning when the word is initially misunderstood because the critical context is given *after* the word (subsequent disambiguation, e.g. “Sally worried that the ball would be too crowded.”). In four experiments, adults listened to prior and subsequent disambiguation sentences, and were later tested on their interpretations of primed and unprimed ambiguous words. The results showed that lexical-semantic retuning can occur for *both* sentence types. Importantly, however, the emergence of priming for subsequent disambiguation sentences was sensitive to the prime conditions: when the task could potentially be performed without needing to re-analyse the ambiguity, then no significant priming was observed. This is consistent with the ‘good enough’ view of language processing which states that representations can remain as (im)precise as mandated by the situation, and that lexical-semantic retuning operates on the output of good-enough interpretation. More generally, our findings suggest that lexical-semantic retuning is driven by participants’ final interpretation of the word meanings during the prime encounter, regardless of initial meaning activation or misinterpretation.

### Introduction

The ability to access word meanings rapidly and accurately is a key component of language comprehension. This task is made difficult by the inherent ambiguity of most words: more than 80% of common English words have multiple dictionary definitions (Rodd, Gaskell, & Marslen-Wilson, 2002). Successful language comprehension requires that, for each ambiguous word, the listener/reader selects the correct interpretation from the set of familiar meanings. This semantic disambiguation process is facilitated by comprehenders’ knowledge of distributional properties of word meanings. In particular, listeners and readers are highly sensitive to the relative usage frequencies (dominance) of meanings for a given ambiguous word. There is abundant

evidence that the meanings that are used more frequently for a given word are more easily accessed. For example, when presented in a neutral context, a word like “pen” will most likely be interpreted as referring to its most frequent (dominant) ‘writing implement’ meaning, rather than its less frequent (subordinate) ‘animal enclosure’ meaning. In addition, when ambiguous words are presented in a clearly disambiguating context, they are processed more easily when that prior context points towards the dominant meaning (e.g. “the fountain pen”) compared to a subordinate meaning (e.g. “the sheep pen”). Similarly, when the word is presented in a neutral context (e.g. “Sally worried that the ball would be too...”), processing slows when subsequent context is encountered which is only consistent with a subordinate meaning (e.g. “crowded”), compared to when this context is compatible with the dominant

\* Corresponding author at: MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, England.

E-mail address: [becky.gilbert@mrc-cbu.cam.ac.uk](mailto:becky.gilbert@mrc-cbu.cam.ac.uk) (R.A. Gilbert).

<https://doi.org/10.1016/j.jml.2020.104188>

Received 22 May 2019; Received in revised form 21 October 2020; Accepted 24 October 2020

Available online 8 November 2020

0749-596X/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

meaning (e.g. “expensive”; see [Rodd, 2020](#); [Vitello & Rodd, 2015](#) for reviews).

Much previous research on the availability of meanings for ambiguous words has focused on the effects of (1) meaning frequency in the language as a whole and (2) immediate context. However, what remains unclear is exactly how individual encounters with ambiguous words in particular contexts lead to the accumulation of long-term knowledge about relative word-meaning frequencies. Recently, a nuanced picture has emerged in which the availability of a word’s meanings at any point in time is driven by a complex combination of the individual comprehender’s shorter- and longer-term experience with the use of that particular word ([Rodd et al., 2016](#)). These findings show that learning mechanisms play a key role in allowing individuals to keep track of word use in their linguistic environment to facilitate access to the most likely meanings of words. This type of lexical-semantic retuning seems consistent with the broader tendency for comprehenders to continuously adapt to any fluctuating aspects of their linguistic environments, presumably to refine predictions and maximise communicative efficiency. For instance, there is analogous evidence surrounding adaptation that follows from exposure to particular syntactic structures ([Fine & Jaeger, 2013](#); [Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vandereelst, 2008](#); [Jaeger & Snider, 2013](#); [Ryskin, Qi, Duff, & Brown-Schmidt, 2017](#)) and to perceptually ambiguous/degraded speech ([Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005](#); [Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008](#); [Jesse & McQueen, 2011](#); [Norris, McQueen, & Cutler, 2003](#)), leading to similar conclusions about the role of learning mechanisms.

Regarding the comprehension of ambiguous words, there are two main findings that show that it is affected by experience with word usage over a range of timescales. First, individuals more readily access word meanings that are frequently used in their particular environments, for instance through hobbies and jobs, but are subordinate in the language more generally (e.g. the rowing-related meanings of words like “square” and “feather” gradually become more dominant with increasing years of rowing experience; [Rodd et al., 2016](#); see also [Wiley, George, & Rayner, 2016](#)). Second, evidence from a paradigm known as ‘word-meaning priming’ has shown that the availability of word meanings changes following even a single encounter with an ambiguous word in a disambiguating context ([Rodd, Lopez Cutrin, Kirsch, Millar, & Davis, 2013](#)). In particular, lower frequency meanings become more readily available when they have been encountered within the preceding minutes or hours. The present paper explores the kinds of sentences that promote or prevent word-meaning priming, with the goal of revealing precisely when and how this form of learning occurs.

In word-meaning priming experiments, participants initially encounter an ambiguous word within a sentence context that disambiguates toward a subordinate meaning (e.g., “Sally worried about how crowded the ball would be.”). Then, after a delay of anything from two minutes to 24 h, they perform a task that provides a measure of the availability of the primed word’s different meanings. The most commonly used test of word-meaning priming has been word association, which is a measure of meaning preference ([2013](#); [Betts, Gilbert, Cai, Okedara, & Rodd, 2018](#); [Gaskell, Cairney, & Rodd, 2019](#); [Gilbert, Davis, Gaskell, & Rodd, 2018](#); [Rodd et al., 2016](#)). In this test, participants encounter each ambiguous word in isolation (e.g., “ball”) and are asked to generate the first associated word that comes to mind. Following exposure to a prime sentence containing the ambiguous word, participants are more likely to generate responses that correspond to the meaning that was used in the prime sentence (e.g., more responses such as “gown” or “dance” rather than “toy” or “tennis” for the word “ball”) compared with an unprimed baseline condition in which there was no prime sentence containing the word. An alternative approach has been to measure the effect of priming on access to the primed meaning, for instance via a speeded semantic relatedness task in which participants have to make a judgement as to whether the ambiguous word is related to a probe word that is only associated with the primed meaning (e.g.

“ball” - DANCE; [Betts et al., 2018](#); [Gilbert et al., 2018](#)). Both the meaning preference and meaning access tasks have consistently shown that just one encounter with an ambiguous word in a strongly disambiguating context can produce a significant boost in the availability of that meaning when the word is next encountered. This effect is particularly strong at short delays (2–3 min; [Rodd et al., 2016](#)), but remains robust and stable at 20–40 min between prime and test ([Rodd et al., 2013](#)). Remarkably, this alteration in the availability of the meanings of an ambiguous word has also been observed after 24 h in the case where participants are able to sleep soon after the sentence exposure ([Gaskell et al., 2019](#)). Word-meaning priming has not only been shown in carefully controlled lab-based conditions, but also in more naturalistic conditions, such as if prime sentences are embedded within short vignettes on a radio programme, and the test of priming is conducted minutes or hours later via a web-based word association task ([Rodd et al., 2016](#)). The relatively long time-course of word-meaning priming has led to the view that it is best thought of as reflecting the operation of learning mechanisms that maintain knowledge of word meanings (for similar arguments regarding syntactic priming, see e.g. [Bock & Griffin, 2000](#); [Chang, Dell, Bock, & Griffin, 2000](#); [Hartsuiker et al., 2008](#); [Ryskin et al., 2017](#)).

Importantly, the longevity of word-meaning priming distinguishes it from more general semantic priming, and therefore it cannot be explained in terms of residual activation in the semantic system. [Rodd et al. \(2013\)](#) adapted the word-meaning priming paradigm to investigate whether the same priming effect could be produced by exposing participants to sentences in which the critical ambiguous word (e.g. ‘post’ in the sentence “The girl wondered when the post would be advertised”) was replaced with a synonym (e.g. “The girl wondered when the job would be advertised”). This later sentence type (a ‘semantic prime’ condition), carries the same semantic information as the word-meaning prime sentence, but critically did not contain the ambiguous word itself. After exposure to these two sentence types, participants were later tested on their interpretations of the ambiguous words (e.g. ‘post’). If word-meaning priming is simply the result of residual activation of semantic features related to one particular meaning of an ambiguous word, then the semantic prime and word-meaning prime conditions should produce similar effects on word association responses. However that was not the authors’ finding; the semantic prime condition had a biasing effect on ambiguous word interpretations after a three minute delay between prime and test, but the impact of this sentence type was abolished after 20 min, whereas the word-meaning prime effect was robust at both shorter and longer delays. Thus word-meaning priming is specific to encountering the ambiguous word itself in a context favouring its subordinate meaning, and this leads to a relatively lasting change in the way that the word is subsequently interpreted (see [Ryskin et al., 2017](#) for similar evidence regarding the effect of recent experience on word-specific updating of preferred interpretations for ambiguous syntactic structures).

In this work we report four experiments with the goal of specifying the cognitive mechanisms that support long-term word-meaning priming in more detail. In particular, we aim to understand how this form of lexical-semantic learning relates to the cognitive processes that support the comprehension of ambiguous words within their sentence contexts. The fact that long-term word-meaning priming requires exposure to the ambiguous word itself in the prime sentence suggests that lexical-semantic updating may occur at or around the point of the primed word’s recognition and meaning access ([Rodd et al., 2013](#)). We will therefore briefly describe key cognitive processes involved in the comprehension of ambiguous words in sentences.

Based on evidence from both cross-modal semantic priming studies (e.g. [Onifer & Swinney, 1981](#); [Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982](#); [Swinney, 1979](#)) and eye-movement research (e.g. [Duffy, Kambe, & Rayner, 2001](#)), the field has converged on a view exemplified by the reordered access model ([Duffy et al., 2001](#); [Duffy, Morris, & Rayner, 1988](#)). According to this view, when readers/listeners

encounter an ambiguous word, they initially activate multiple possible meanings in parallel but then rapidly select one of these meanings to be integrated with the ongoing discourse representation. The level of initial meaning activation is modulated by both sentence context and meaning frequency such that more frequent or more contextually-appropriate meanings are more likely to be selected and are selected more rapidly. In most cases, these constraints will support the selection of the correct meaning. However, occasionally the listener/reader will select the wrong meaning of a word and correct comprehension will depend on subsequently detecting that error and engaging in a cognitively demanding reinterpretation process (Macgregor et al., 2020; Rodd, Johnsrude, & Davis, 2010).

Based on this cognitive distinction between (i) initial meaning access plus selection, and (ii) subsequent error detection and reinterpretation, we can explore two plausible mechanisms by which the relative availability of word meanings might be adjusted as a consequence of sentence comprehension. One possibility is that learning occurs primarily at the time of initial meaning access following the encounter with the primed word, for instance, via an associative process in which the connections between an ambiguous word's form and the initially-selected meaning are automatically strengthened. Another possibility is that changes to word-meaning preferences occur primarily in response to expectancy violations, and will therefore be greatest following comprehension failures during the prime encounter. These two learning mechanisms make opposite predictions concerning the presence/strength of word-meaning priming effects with two different sentence constructions: prior and subsequent disambiguation. We will examine these two possibilities in turn.

Under the first view, learning will be optimal when comprehenders are able to rapidly select the correct subordinate meaning upon encountering the ambiguous word. Prior disambiguation sentences will provide the best opportunity for this to occur, because the subordinate-meaning context will already be available to guide initial meaning selection. For instance, in the prior disambiguation sentence "Sally worried about how crowded the ball would be", the disambiguating word "crowded" is presented before the ambiguous word "ball" and thus allows the comprehender to quickly rule out the dominant 'round object' meaning while facilitating the selection of the appropriate 'dancing party' meaning. Previous word-meaning priming experiments have used prior disambiguation sentences, so we expect to see priming with this context positioning. Furthermore, this account would parallel results showing greater perceptual learning for perceptually-ambiguous or degraded stimuli when disambiguating information is available before, rather than after, the ambiguity (see Davis et al., 2005; Hervais-Adelman et al., 2008 for evidence from degraded speech; see Jesse & McQueen, 2011 for evidence from ambiguous speech sounds).

This view predicts that word-meaning priming would be reduced, or even absent, when the critical disambiguating context is given after the ambiguous word. For such subsequent disambiguation sentences (e.g., "Sally worried that the ball would be too crowded."), the comprehender is likely to select the *inappropriate* dominant meaning (e.g. 'round object' meaning of 'ball') around 200 to 1500 ms after encountering the ambiguous word (Onifer & Swinney, 1981; Seidenberg et al., 1982; Swinney, 1979). Any learning that results from initially selecting the dominant meaning would need to be at least partially corrected when the contextually-appropriate subordinate meaning is selected during later reinterpretation. Under this account, it is therefore unlikely that learning from subsequent disambiguation sentences would be as efficient as for prior disambiguation sentences because of interference from the initial selection of the dominant meaning. While exposure to the dominant meaning of an ambiguous word will almost certainly produce a weaker effect than exposure to a subordinate meaning, previous word-meaning priming studies suggest that dominant meaning priming is nonetheless possible (Betts et al., 2018) and will depend on the word's baseline dominance (Rodd et al., 2013). Also, studies of garden-path sentences show that there are lasting effects of initial

misinterpretations, for instance because the misparsed version of the sentence remains active, even after it has been corrected (Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013; van Gompel, Pickering, Pearson, & Jacob, 2006), or because the listener is able to maintain multiple incompatible interpretations in parallel until more evidence is acquired (Levy, Bicknell, Slattery, & Rayner, 2009). Regardless of the specific mechanism, there may be a similar type of lingering effect that occurs with initial lexical-semantic misinterpretations. Finally, word-meaning priming in response to subsequent disambiguation sentences may be weak or absent because the ambiguous spoken word form is no longer present during reinterpretation (though the word may be covertly recalled).

A second possibility is that error detection and repair processes support learning. By this view, we would expect the opposite pattern of results: learning would be maximal for subsequent disambiguation sentences (e.g. "Sally worried that the ball would be too crowded.") in which the inappropriate meaning of the ambiguous word is initially selected, and the later disambiguating context triggers a reinterpretation or repair process. This account would be consistent with the more general view that expectancy violations enhance learning because learning seeks to minimise future prediction errors (Brod, Hasselhorn, & Bunge, 2018; Greve, Cooper, Kaula, Anderson, & Henson, 2017; Henson & Gagnepain, 2010; Jaeger & Snider, 2013). The repair processes that occur during subsequent disambiguation sentences may similarly signal that the current representations are not optimally supporting comprehension. Indeed, brain imaging data suggest that there is a large processing cost when these types of sentences resolve to a subordinate meaning, compared to when they resolve to a dominant meaning or remain unresolved, which is likely due to an effortful search for alternative word meanings and sentence re-evaluation (Macgregor et al., 2020; Rodd, Johnsrude, & Davis, 2012). Given that word-meaning priming has already been observed using prior disambiguation sentences (Betts et al., 2018; Gaskell et al., 2019; Gilbert et al., 2018; Rodd et al., 2016, 2013), this account must predict that some degree of learning occurs automatically in response to *all* ambiguous words. However, if learning is linked to error detection or cognitively-demanding repair processes, then sentences that lead to transient interpretation errors should result in even greater changes to preferred meanings.

Our aim was to investigate whether word-meaning priming is primarily driven by initial meaning access/selection, or if it is boosted by more effortful reinterpretation. We examined the processing of ambiguous words that were previously heard in one of two synonymous sentence types: one in which the subordinate-meaning context was presented before the ambiguous word (prior disambiguation, e.g. "Sally worried about how *crowded* the ball would be.") and another in which the context was presented after the ambiguous word (subsequent disambiguation, e.g. "Sally worried that the ball would be too *crowded*."). First, if word-meaning priming does not depend critically on the immediate selection of the appropriate subordinate meaning during the prime phase, then we should find that, like prior disambiguation sentences, subsequent disambiguation sentences also produce a reliable priming effect. Second, if word-meaning priming occurs primarily through one of these two learning processes, then we expect to see significantly more priming for one or other of the two sentence types.

In a set of four experiments, we investigated whether having to revise an initial incorrect meaning selection affects the magnitude of retuning towards the primed meaning. All experiments used the long-term word-meaning priming paradigm, with average delays of around 20 min between the prime and test encounters. In Experiment 1, we tested participants' meaning preferences using the word association task, in which participants heard each ambiguous word in isolation and generated an associated word. Word association provides a measure of meaning preference in the absence of any biasing context or probe, and it allows for a direct comparison in priming magnitude with many previous word-meaning priming experiments (Betts et al., 2018; Gaskell et al., 2019;

Gilbert et al., 2018; Rodd et al., 2016, 2013). Experiment 2 was an attempt to replicate the results of Experiment 1 using a test of meaning access, which was a speeded semantic relatedness task. The results of Experiment 2 presented an interesting follow-up question: does the presence of additional disambiguating context after the prime sentence also influence the magnitude of word-meaning priming? Experiments 3 and 4 were designed to address this question using the same meaning preference and meaning access test tasks used in Experiments 1 and 2.

### Experiment 1

#### Method

##### Participants

Thirty volunteers (14 women; mean age 24.3 years, SD = 9.0 years) were recruited through the University College London participant pool and were paid for participation. In order to be eligible, participants were required to be native British English speakers with no reported hearing or reading impairments.

##### Design

Prime Type (Prior Disambiguation, Subsequent Disambiguation, Unprimed) was manipulated within-subjects and within-items. For each participant, each item appeared in only one Prime Type. Therefore three lists of 19 words each were created to cross items with Prime Type so that, across participants, each word appeared in each condition equally often but each participant encountered each word in only one of the three conditions.

##### Materials and procedure

Participants were randomly assigned to one of the three versions at the start of the session and were tested individually in a quiet room. After completing a brief language background questionnaire to ensure that they met the inclusion criteria, participants completed the following: (i) sentence exposure task (prime phase), (ii) digit span task (filler), (iii) meaning preference task (test phase), (iv) clarification of meaning preference responses. The average duration between the prime and test presentations of an ambiguous word was approximately 25 min (range 10–42 min). All tasks were programmed using Cogent/Matlab software, and audio stimuli were presented via headphones. Each task

began with three practice trials. See Fig. 1 for an overview of the procedure.

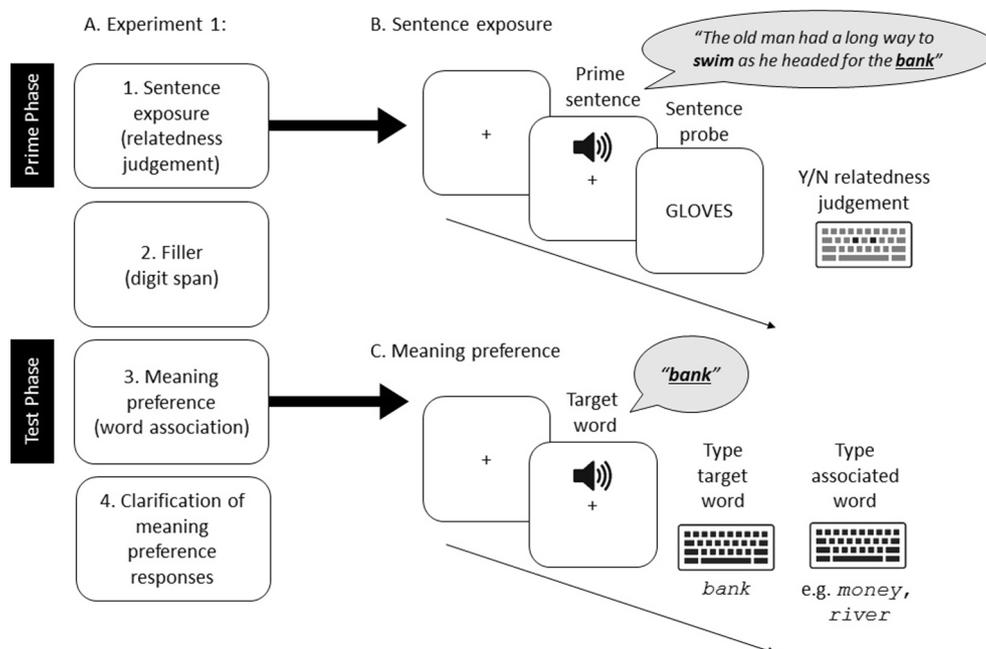
**Sentence exposure task.** The purpose of this task was to expose participants to the subordinate-resolved sentences for ambiguous words in the Prior and Subsequent Disambiguation Prime Types. Participants heard 19 Prior Disambiguation sentences, 19 Subsequent Disambiguation sentences, and 41 filler sentences in random order. After each auditory sentence ended, a visual probe word appeared and participants made a yes/no judgement about the semantic relatedness of the sentence and probe word. Responses were made via a computer key press. Experimental items that were assigned to the Unprimed condition were not presented in this phase. See Table 1 for example stimuli.

The experimental items were 57 ambiguous words, which were either homonyms (same pronunciation and spelling, e.g. BANK) or homophones (e.g. same pronunciation and different spelling, e.g. NIGHT/KNIGHT; see Table S1 in the Supplementary Material for a list of experimental words and meanings). For each word there were two prime sentences; one where the disambiguating context was given *before* the ambiguous word (Prior Disambiguation) and one where this context was

**Table 1**

Experiment 1 stimuli for a single item (bank) and all three Prime Type conditions. For each participant, each experimental item was assigned to only one of the three Prime Type conditions. Ambiguous words are underlined and disambiguation words are in italics in the prime sentences. In the sentence exposure task, sentence probe words were always unrelated to the sentences for experimental items, and semantically-related to the sentences for filler items.

Prime type	Sentence exposure task		Meaning preference task Target word (auditory)
	Prime sentence (auditory)	Unrelated sentence probe (visual)	
Prior	The old man had a long way to <i>swim</i> as he headed for the <u>bank</u> .	GLOVES	bank
Subsequent	The old man headed to the <u>bank</u> but he had a long way to <i>swim</i> .	GLOVES	bank
Unprimed	–	–	bank



**Fig. 1.** Overview of Experiment 1 task order (A), and procedures for the sentence exposure task (B) and the meaning preference task (C).

given *after* the ambiguous word (Subsequent Disambiguation; see Table 1 for examples, and see Table S2 in the Supplementary Material for the full set of sentences). In order to reduce any differences between the two sentence types other than the position of the disambiguation, the sentence pairs were constructed such that they used as many of the same words as possible.

The mean sentence length was similar for the two types of experimental sentences (Prior: 13.1 words,  $SD = 1.9$ , range = 9–17; Subsequent: 13.0 words,  $SD = 1.4$ , range = 11–16). In the Subsequent sentences, the context word was always the last word in the sentence, and the ambiguous word position was in the middle (mean = 6.1 words in,  $SD = 0.8$ , range = 5–8). These positions were more variable in the Prior sentences; the context word was usually in the middle of the sentence (mean = 7.3 words in,  $SD = 2.1$ , range = 2–12) and the ambiguous word was at the end of the sentence, usually the final word (mean = 12.6 words in,  $SD = 2.1$ , range = 8–17). The mean distance between the ambiguous word and disambiguating context was greater for Subsequent sentences (6.9 words,  $SD = 1.4$ , range = 5–10) compared to Prior sentences (5.4 words,  $SD = 2.1$ , range = 1–10).

All sentences containing ambiguous words were biased towards the subordinate meaning (mean dominance of primed meanings for words in isolation = 0.20,  $SD = 0.08$ , range = 0.08–0.38). A pre-test with a separate group of participants ( $N = 40$ ) was conducted on the Subsequent Disambiguation sentences to confirm that they were not unintentionally biased toward the subordinate meaning before the disambiguating context word was presented (for full details, see Blott, Rodd, Ferreira, & Warren, 2019, Supplementary Methods: <https://osf.io/786pu/>). Participants were presented with the sentence lead-ins up to the sentence-final disambiguating word (e.g. “Sally worried that the ball would be too...”) and asked to generate another word that could be substituted for the ambiguous word as it was used in the sentence lead-in. These responses were scored according to whether the word referred to the meaning that would have been indicated by the missing disambiguation word (e.g. responses such as “dance” and “gala” in the “ball” sentence lead-in) or to any other meaning (e.g. responses such as “toy” and “object”). The prime-consistent subordinate meanings remained in the minority in these responses; the mean proportion of responses related to the primed meaning in lead-in sentences was 0.26 ( $SD = 0.16$ , range = 0.00–0.60). This showed that the Subsequent Disambiguation sentences were, on average, not unintentionally biased toward the subordinate meanings before the sentence-final disambiguating word.

Each prime sentence was paired with a sentence probe word (see Table 1 for an example, and see Tables S2 and S3 in the Supplementary Material for the full list of probe words that followed the experimental and filler sentences). For the sentences containing experimental items, the sentence probe words were always semantically unrelated and never related to an inappropriate (e.g. dominant) meaning of the ambiguous word. The use of unrelated sentence probes ensured that any priming was only due to exposure to the ambiguous word in the subordinate-meaning sentence context. This also prevented the need to remove word association responses that matched the sentence probe words, as these could reflect associative pairings between the ambiguous word and sentence probe. For each item, the same sentence probe word was paired with Prior and Subsequent Disambiguation sentence versions.

There were 41 low-ambiguity filler sentences in the sentence exposure task (see Table S3 in the Supplementary Material for a list of filler words and sentences). These filler sentences were included to reduce the salience of the ambiguity and to balance the proportion of related and unrelated responses in the sentence exposure task. Since the experimental sentences were always followed by unrelated probe words, the filler sentences were always paired with semantically-related probes. Filler sentences were constructed by taking Prior and Subsequent Disambiguation experimental sentences and replacing key words to change the meaning, in order to mimic the structures of these two sentence types.

**Filler task.** A digit span task was included to increase the delay between the prime and test phases, and to create a switch in task context between these phases. Visually-presented number strings (500 ms per digit) were presented for immediate serial recall. There were 75 sequences ranging from three to nine digits in length. Following a cue at the end of each sequence, participants entered their response followed by a specific key press, which ended their response and triggered the start of the next trial. The sequences were presented in blocks of 25 trials, with 15 s breaks between blocks.

**Meaning preference task.** The purpose of this task was to measure meaning preferences for ambiguous words in the three Prime Type conditions. All 57 experimental items (19 primed with Prior Disambiguation sentences, 19 primed with Subsequent Disambiguation sentences, and 19 Unprimed) were presented one at a time in random order. After each spoken word presentation, participants typed the spoken target word into a text box (in order to determine whether they had heard the word correctly) and then typed the first associated word that came to mind.

**Meaning clarification task.** This task was included to clarify which meaning was intended when the participant made the word association response. For some word association responses, it may not be obvious to an experimenter how the ambiguous word and response are related. Also, some responses could relate to more than one meaning of the ambiguous word (e.g. the associate response “foot” could relate to either the walking action or raised platform meaning of the ambiguous word “step”). We therefore asked participants to indicate which meaning of the ambiguous word was intended when they made each of their responses. In this task the spoken ambiguous words were presented again, one at a time, along with the participant’s associated response for each. Participants then selected from multiple choice options which included definitions for the dominant and subordinate (primed) meanings of the ambiguous word as the first two multiple choice options (presented in random order), and “other meaning”, which was always the third option. For instance, if the ambiguous word was “chest” and the participant’s response was “heart”, they would select the “body part” definition rather than the “storage container” definition. If the participant’s response related to a meaning of the word that was not listed in the options, they would select “other meaning” and write down this meaning on a sheet of paper provided.

All of the auditory materials (experimental and filler prime sentences, single experimental words) were recorded by a female British English speaker (JMR) in a sound-proof booth. The audio files were matched for RMS amplitude.

## Results

Data from this and subsequent experiments is available at <https://osf.io/t9q6v/>. Accuracy in the sentence exposure task was good (mean correct = 95.4%, minimum = 81.0%), indicating that participants were attending to the prime sentences and processing their meanings. Two items were removed from the analysis of the meaning preference task responses: “landing” was excluded due to technical problems during the experiment, and “club” was excluded because preliminary analysis showed that the prime sentences did not fully disambiguate the word.

### Response coding

The meaning preference responses for the remaining 55 items were coded using the responses from the meaning clarification task. A response was coded as consistent with the primed meaning if it referred to the same meaning used in the prime sentence, and inconsistent if it referred to any other meaning of the word. In order to ensure that participants had accurately coded their meaning preference responses,

the responses from the two tasks were cross-referenced by the experimenter. The participants' meaning clarification responses were overruled in certain cases. First, when a participant selected the "other meaning" option but their response clearly related to one of the two definitions provided (0.06% of responses). Second, when a participant's response clearly related to a meaning other than the one selected (e.g. for "ace", if the response was "spades" but the definition selected was "serve in tennis" rather than "card suit"), then we assumed that this selection was an error (0.18% of responses). Third, when two key press responses were made during the meaning clarification task, we assumed that this was a correction to the first response so the second response was used.

#### Meaning preference analysis

Of the 1650 trials (30 participants  $\times$  55 items), 17 trials were excluded because no response was recorded (seven trials), the response was uninterpretable (one trial), or the participant's typed response to the spoken word indicated that they misheard the word (nine trials). Fig. 2 shows the proportions of the remaining 1633 responses (participant grand means) that were consistent with the primed subordinate meaning within each Prime Type. The data were analysed with a logistic mixed effects model using R (R Core Team, 2013) and the "lme4" (version 1-1.10; Bates, Mächler, Bolker, & Walker, 2015) and "phia" (version 0.2-1; Rosario-Martinez, 2015) packages for R. The fixed factor was Prime Type with three levels (primed with Prior Disambiguation sentence, primed with Subsequent Disambiguation sentence, Unprimed), which was assigned Helmert contrasts to provide separate estimates for primed vs unprimed conditions (Prior:  $-1/3$ , Subsequent:  $-1/3$ , Unprimed:  $2/3$ ) and for Prior vs Subsequent Disambiguation sentence types (Prior:  $-1/2$ , Subsequent:  $1/2$ , Unprimed: 0). Here and in subsequent experiments, we started with the maximal random effects structure in order to prevent an inflated probability of Type I error (Barr, Levy, Scheepers, & Tily, 2013). Then, if the model showed evidence of being over-fitted (e.g. did not converge, perfect correlations between random effects), we reduced the model by removing the random effect that accounted for the least variance. In the present experiment, the maximal model did not require any reductions, which resulted in by-item and by-subject random intercepts and slopes for Prime Type

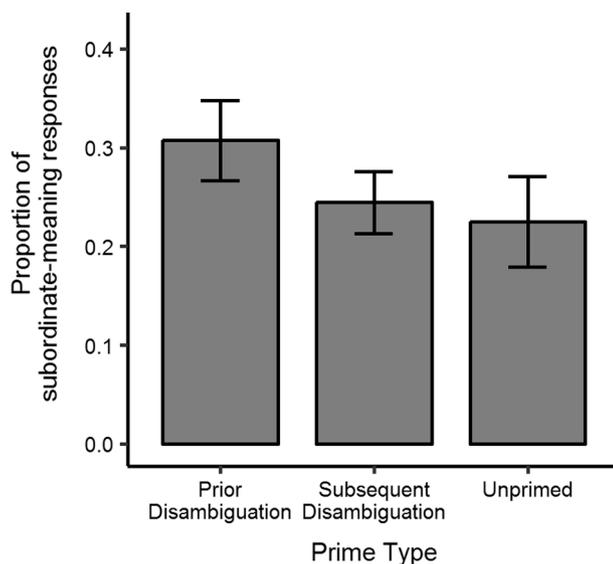


Fig. 2. Experiment 1 proportions of meaning preference task responses that were consistent with the primed subordinate meaning for ambiguous words in the Prior (left) and Subsequent (middle) Disambiguation prime conditions, and in the Unprimed condition (right). Bars show the participant grand means, and error bars show 95% confidence intervals, adjusted to remove between-subject variance (Morey, 2008).

contrasts.

The model revealed a significant effect of the Prime Type factor,  $\chi^2(2) = 8.92, p = .012$ . The primed vs unprimed contrast was significant,  $\beta = -0.42, SE = 0.19, z = -2.24, p = .025$ . There was also a significant effect of the contrast for Prior vs Subsequent sentence types,  $\beta = -0.36, SE = 0.16, z = -2.22, p = .027$ . Pairwise comparisons with Holm adjustment for multiple comparisons revealed that the increase in proportions of consistent meaning preference responses in the primed vs unprimed conditions was driven by the Prior Disambiguation condition. There were significantly more consistent meaning preference responses in the Prior Disambiguation compared to the Unprimed condition,  $\chi^2(1) = 8.65, p = .010$ , but proportions of consistent responses in the Subsequent Disambiguation and Unprimed conditions were not significantly different,  $\chi^2(1) = 1.43, p = .232$ .

#### Discussion

We observed a significantly greater preference for interpreting single ambiguous words with their primed subordinate meanings when the words were primed with Prior Disambiguation sentences, compared to both the Unprimed and Subsequent Disambiguation conditions. The average increase in subordinate-meaning responses for words primed with Prior Disambiguation sentences was of a similar magnitude (around 8%) as seen in previous word-meaning priming studies using similar methods (Rodd et al., 2016, 2013). While there were, on average, more subordinate-meaning responses for words primed with Subsequent Disambiguation sentences compared to the Unprimed condition, this difference was smaller (around 2%) and not statistically significant. It is not clear whether the lack of significant priming for words in the Subsequent Disambiguation condition was due to a true absence of the effect or insufficient power to detect it. Taken together, these results provide preliminary support for the view that word-meaning priming is the result of a process that occurs during initial meaning access.

However, given that word association is not a speeded task, it was possible that the effect of priming was due to an off-line strategy during this task, rather than a change in meaning preferences for the mapping from word form to meaning. That is, there may not have been any difference in the *initial* processing of primed and unprimed ambiguous words during the test phase. One possibility is that participants were explicitly recalling the prime sentences for ambiguous words during the word association task, and then deciding to respond with a consistent meaning. For this reason it was important to obtain convergent evidence from a speeded test task (see Gilbert et al., 2018 for a similar approach).

#### Experiment 2

Experiment 2 investigated whether the pattern of results from Experiment 1 would replicate using a speeded test of word-meaning access, instead of the meaning preference task. In the test phase, each trial consisted of a spoken ambiguous word followed immediately by a written subordinate-related probe word (e.g. "ball" – DANCE; see Fig. 3 in the Method section). Participants were asked to decide, as quickly as possible, whether or not the meanings of the two words are related. If recent experience with an ambiguous word in a subordinate-biased sentential context facilitates access to this meaning during the comprehension of the word at its next encounter, then we expect to find that semantic relatedness responses to primed word pairs will be faster and/or more accurate compared to responses for unprimed pairs. In addition, based on the results of Experiment 1, we predicted that any priming effects would be greater for ambiguous words that have been primed with Prior compared to Subsequent Disambiguation sentences.

A number of changes were made to the sentence exposure task and stimuli composition as a result of the change to the test task. Because the trials in the meaning access task required a yes/no decision, and because all of the experimental items were paired with related probe words in the meaning access task (in order to assess the speed of access to

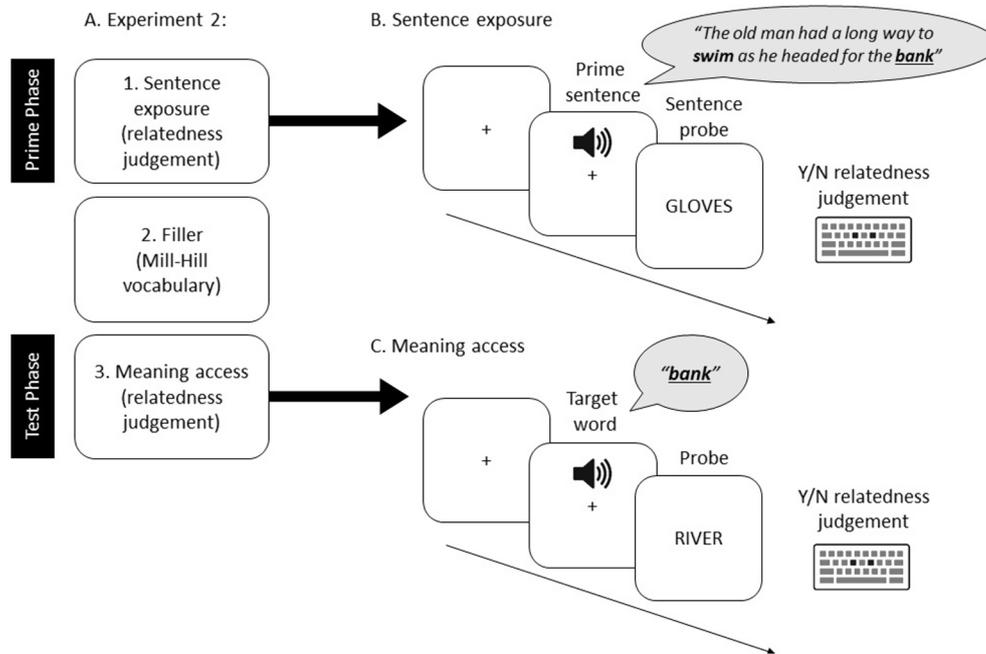


Fig. 3. Overview of Experiment 2 task order (A), and procedures for the sentence exposure task (B) and the meaning access task (C).

subordinate meanings), an equal number of unrelated target-probe pairs were needed as fillers to balance the related/unrelated responses. To decouple the meaning access task response type (related or unrelated) from the target word ambiguity and familiarity (primed or unprimed), equal numbers of experimental and filler Prior and Subsequent Disambiguation sentences were presented in the prime phase so that the ambiguous words from filler sentences could be paired with *unrelated* probe words in the test phase. Also, because one third of the experimental items (paired with related probe words) in the meaning access task were unprimed, we also included an equal proportion of unprimed ambiguous filler words (paired with unrelated probe words) to balance the familiarity of target words in the related and unrelated trials. These changes to the design were necessary to prevent participants from learning that, whenever the target word in the test phase was familiar (i. e. remembered from the prime phase) and/or ambiguous, the probe word that followed would be semantically-related.

We also made a few changes to the study procedure. One change in the current experiment was the addition of a disambiguation post-test, which allowed us to determine, on an individual participant level, which primed word meanings each participant was familiar with and could successfully access based on the prime sentence contexts. This allowed us to obtain a more accurate measure of priming by excluding individual trials from the test task analysis in which the participant likely misunderstood the prime sentence for that word. Another change was that, for this and subsequent experiments, data collection was web-based. Recent advances in browser-based stimulus presentation methods and the increased access to computers with fast internet connections has made web-based experiments a viable alternative to lab-based studies. There are several advantages to collecting data over the internet, including the ability to collect large numbers of data sets quickly and efficiently, and the ability to access a broader population than is typical in experimental psychology (Woods, Velasco, Levitan, Wan, & Spence, 2015). Recent studies have shown that the quality of data collected online in general, and for reaction time experiments in particular, is similar to that collected in a traditional laboratory setting (Barnhoorn, Haasnoot, Bocanegra, & van Steenbergen, 2015; de Leeuw & Motz, 2016; Hilbig, 2015; Reimers & Stewart, 2015; Schubert, Murteira, Collins, & Lopes, 2013).

## Method

### Participants

We collected data over the internet from 124 individuals. Data from 28 participants was discarded because of technical problems. Seven participants completed the experiment but were excluded because the session ran longer than the maximum allowed duration of one hour (two participants), accuracy in the sentence exposure and/or meaning access tasks was close to chance rate (three participants), or they did not follow instructions in the post-test (three participants). The remaining 89 participants were included in the analysis (30 in versions 1 and 2, 29 in version 3; 50 women; mean age 30.0 years, SD = 8.2 years, range = 18–48 years). Participants were paid £4 to complete the session, which lasted around 35–45 min. The excluded participants were paid for their time (pro-rata for those who stopped part-way through due to technical problems).

In order to be eligible, participants had to be 18–50 year old native British English speakers currently living the UK. Participants were recruited online via Prolific Academic ([www.prolific.ac](http://www.prolific.ac); Palan & Schitter, 2017; Peer, Samat, Brandimarte, & Acquisti, 2015). Eligibility was verified using Prolific's pre-screening filters and a brief questionnaire at the start of the experiment, and current location was verified based on the IP address. We also used a vocabulary test as the filler task (in place of the digit span task in Experiment 1) to verify English language proficiency.

### Design

The design was the same as that used in Experiment 1. Prime Type (Prior Disambiguation, Subsequent Disambiguation, Unprimed) was a within-subject manipulation. The experimental items were split into three lists (matched for mean dominance), and word lists were rotated across Prime Types, resulting in three versions of the experiment. Thus each participant encountered each ambiguous word in only once in the prime and test phases, and ambiguous words were represented in each condition equally often across participants. The dependent variables were the response times and accuracy of semantic relatedness decisions for target and subordinate-related probe word pairs in the meaning access task.

### Materials and procedure

The experiment was conducted online using Qualtrics survey software and QRTE (Barnhoorn et al., 2015), a JavaScript-based add-on to Qualtrics that allows precise timing for response time experiments. In QRTE, the trial information and stimuli are loaded prior to each trial, which means that stimulus presentation timing is never disrupted by internet communication delays *during* a trial. However, internet communication problems can affect timing *between* trials, as each trial does not begin until all communication is complete between the participant's computer and the server. In order to reduce the likelihood of inter-trial intervals becoming much longer than the intended duration, we conducted a connection speed test at the start of the experiment to estimate the inter-trial loading time (see [www.qrtengine.com](http://www.qrtengine.com) for more details). If the estimated inter-trial interval was longer than the threshold, the participant was shown an error message and prevented from starting the experiment.

Participants were instructed to ensure that they were in a quiet, distraction-free environment and to wear headphones if possible. The experiment began with a sound test to verify that the audio could be heard clearly and to allow any adjustment to the volume. Participants then completed the following: (i) sentence exposure task, (ii) Mill-Hill vocabulary (filler), (iii) meaning access task<sup>1</sup>. See Fig. 3 for an overview of the procedure.

**Sentence exposure task.** In each trial in this task, a spoken sentence was presented, followed by a visual probe word, and the participant made a decision as to whether the sentence and probe word are semantically related (see Fig. 3B). Each participant was exposed to 44 of the 66 experimental items; 22 of these were presented in Prior Disambiguation sentences and 22 were presented in Subsequent Disambiguation sentences. The remaining 22 experimental items were unprimed and thus were not presented in this task. There were 44 ambiguous filler sentence-probe pairs, as well as 24 low-ambiguity filler sentence-probe pairs, resulting in 112 total trials in this task. Half of the trials contained probe words related to the sentence meaning, and the other half contained unrelated probe words. See Table 2 for example stimuli from each condition. The trials were separated into four blocks of 28 trials with self-paced breaks between blocks.

Each item/sentence was paired with either a semantically-related or unrelated probe word for the sentence exposure task. Unlike Experiment 1, which used word association in the test phase, here there was no need to pair the experimental sentences with only unrelated probe words. We therefore decoupled item type (experimental or filler) from the sentence exposure task response type (related or unrelated) by pairing sentences in all conditions with half related and half unrelated probe words. This also allowed us to assess comprehension for experimental sentences that were followed by semantically-related probe words, which made up half of those trials.

Whenever possible, the sentence probe word related specifically to the intended meaning of the ambiguous word, and not to any other words in the sentence. This was to prevent the ability to make a correct 'related' response in the sentence exposure task without successful disambiguation. The semantically-related probe words that followed experimental sentences were selected from a set of words associated with the ambiguous word, which was generated to produce potential probe words for the meaning access task. If there was only one clear option for a subordinate-related probe word for a given experimental

<sup>1</sup> After the meaning access task, we presented a recognition memory test in which participants were asked to select the ambiguous words that they remember from the prime phase. The set of words was presented on a single screen, and participants ticked a box next to any word that they remembered from the first task. However this data was not analysed in detail as preliminary checking revealed very inconsistent performance, which indicated that it was not a useful measure.

**Table 2**

Experiment 2 stimuli for two items (bank, ball) and all three Prime Type conditions. For each participant, each experimental item was assigned to only one of the three Prime Type conditions. Ambiguous words are underlined and disambiguation words are in italics in the prime sentences. In the sentence exposure task, half of the experimental items (e.g. bank) were always paired with unrelated sentence probe words, and the other half of items (e.g. ball) were always paired with semantically-related sentence probe words. In the meaning access task, experimental target words were always paired with a probe word related to the primed subordinate meaning.

Prime type	Sentence exposure task			Meaning access task	
	Prime sentence (auditory)	Related sentence probe (visual)	Unrelated sentence probe (visual)	Target word (auditory)	Probe word (visual)
Prior	The old man had a long way to <i>swim</i> as he headed for the <u>bank</u> .	–	HOROSCOPE	bank	RIVER
Subsequent	The old man headed to the <u>bank</u> but he had a long way to <i>swim</i> .	–	HOROSCOPE	bank	RIVER
Unprimed Prior	–	–	–	bank	RIVER
	Sally worried about how <i>crowded</i> the <u>ball</u> would be.	FORMAL	–	ball	DANCE
Subsequent	Sally worried that the <u>ball</u> would be too <i>crowded</i> .	FORMAL	–	ball	DANCE
Unprimed	–	–	–	ball	DANCE

item, then this word was always used as the probe word in the meaning access task. In cases where it was not possible for the sentence probe word to relate specifically to the ambiguous word (e.g. any probe word for "ace" would also relate to the word "tennis" in the sentence "The man knew that one more ace might be enough to win the tennis."), then an unrelated sentence probe word was chosen for that sentence.

Because of the additional ambiguous filler sentences that were needed in this experiment, the full set of stimuli was recorded again by a different female speaker of southern British English. We took this opportunity to improve the stimuli by starting with the 168 ambiguous words and Subsequent Disambiguation sentences that were originally pre-tested. As in Experiment 1, ambiguous words were excluded if the meaning used in the sentence was not subordinate when meaning preferences were tested for the word in isolation (via word association). We also used additional pre-test measures to form exclusion criteria based on the comprehensibility of the full Subsequent Disambiguation sentence and compatibility of the sentence with the word's dominant meaning (see Blott et al., 2019 Supplementary Methods for details on all pre-test measures). Specifically, items were excluded if: (1) listeners showed a preference for interpreting the ambiguous word with the primed subordinate meaning in the Subsequent Disambiguation sentence lead-in, where the final disambiguating word was omitted (e.g. "Sally worried that the ball would be too..."); (2) comprehension ratings for the full Subsequent Disambiguation sentence were low; and (3) comprehension ratings for the full Subsequent Disambiguation sentence were high when participants were asked to interpret the sentence with

the dominant meaning of the ambiguous word (i.e. the context did not fully disambiguate the word). Finally, some items were excluded for practical reasons, specifically if: (1) the dominant and subordinate word meanings were too similar, making it difficult to select a single subordinate-related probe word for the meaning access task that is also clearly unrelated to the dominant meaning; or (2) it was not possible to form a Prior Disambiguation sentence by re-ordering all or most of the words in the pre-tested Subsequent Disambiguation sentence.

After these exclusions there were 66 experimental items remaining. The meaning preference pre-test for words in isolation showed that the meanings used in the prime sentences were subordinate (mean dominance = 0.21, SD = 0.14, range = 0.00–0.50). The meaning preference pre-test for the Subsequent Disambiguation sentence lead-ins showed that these sentences did not bias the interpretation of ambiguous words toward the subordinate meaning before the sentence-final disambiguation word; when the sentence-final disambiguation word was omitted, participants tended not to interpret the ambiguous word with the primed meaning (mean proportion of subordinate meaning responses = 0.24, SD = 0.15, range = 0.00–0.50). The comprehension pre-tests showed that participants rated the full Subsequent Disambiguation sentences as being fairly easy to understand (mean rating = 1.35 on a 1–4 scale from “Makes sense instantly” to “Does not make sense”, range = 1.00–2.11). Finally, the meaning compatibility pre-tests showed that participants rated the Subsequent Disambiguation sentences as being compatible with the subordinate definitions of the ambiguous words (mean rating = 1.06 on a 1–4 scale from “Makes perfect sense” to “Does not make sense at all”, range = 1.00–1.57) but not compatible with their dominant definitions (mean = 3.43, range = 2.63–4.00).

The items were split into three lists of 22 words matched for mean single-word dominance. There were 66 ambiguous fillers selected from words that were excluded as experimental items based on the criteria specified above (see [Table S4 in the Supplementary Material](#) for filler words, sentences and probe words). Of these filler words, 44 were presented during the sentence exposure task (22 as Prior Disambiguation sentences, 22 as Subsequent Disambiguation sentences) and 22 filler words were unprimed. There were also 24 low-ambiguity filler words, which were included to reduce the salience of the ambiguity. All low-ambiguity fillers were presented during the sentence exposure task, which were constructed by replacing key words in either Prior or Subsequent Disambiguation sentences in order to form filler sentences that mimicked the structure of the experimental sentences.

On average, the Subsequent Disambiguation sentences were slightly longer than Prior Disambiguation sentences (Subsequent: mean length = 12.9 words, SD = 1.3, range = 11–15; Prior: M = 12.7, SD = 2.0, range = 9–17). The position of the ambiguous word was earlier in the sentence for Subsequent compared to Prior Disambiguation (Subsequent: mean position = 6.2 words in, SD = 0.9, range = 4–8; Prior: M = 11.9, SD = 2.4, range = 6–17). In the Subsequent Disambiguation sentences, the disambiguating context was always the final word in the sentence, so the distribution of context position is the same as that for sentence length. The disambiguation context in the Prior Disambiguation sentences was 7.4 words from the start of the sentence on average (SD = 2.0, range = 3–12). In the Subsequent Disambiguation sentences, the ambiguous word was an average of 6.7 words (SD = 1.2, range = 5–9) before the disambiguating context, whereas in the Prior Disambiguation sentences, the ambiguous word was an average of 4.5 words (SD = 2.4, range = 1–10) after the disambiguating context.

**Filler task.** After the prime phase, participants completed the Mill-Hill multiple choice vocabulary test (Raven, Raven, & Court, 1998). In this test, a target word is shown with six single-word multiple choice options, and the participant must choose the option that most closely matches the meaning of the target word. The original test has 34 items, but one item was removed due to semantic overlap between the correct response and an experimental ambiguous word. We also changed eight of the

incorrect (foil) response options due to lexical or semantic overlap with experimental ambiguous words. There were 33 trials in total, presented in order of increasing difficulty.

**Meaning access task.** In the test phase, speed of access to the primed subordinate meaning was tested using a semantic relatedness task. Participants were told they would make judgments about whether or not the meanings of two words are related. There were two examples given with correct responses and brief explanations, followed by seven practice trials with feedback after each. Each trial began with a 300 ms fixation cross, followed by the auditory target word (with the fixation cross still on the screen). Immediately after the offset of the target word, the fixation cross was replaced with a probe word, written in upper case, along with response key reminders to the lower left- and right-hand sides of the probe word. This display remained until the participant made a valid response. The inter-trial interval was 700 ms. In order to encourage participants to respond as quickly as possible, if the participant’s response was longer than 1500 ms then a screen appeared with a warning that the response was too slow. There were 156 trials separated into blocks of 39 trials, and the first four trials of the task were filler items. Breaks were self-paced up to one minute, at which point a warning cue was presented and the next block would begin automatically. In order to minimise large discrepancies in prime-to-test durations across items, the items were presented in the same halves of the sentence exposure and meaning access tasks (unprimed items were divided equally between the first and second halves of the meaning access task).

The main task consisted of 156 word pairs (see [Tables S8 and S9 in the Supplementary Material](#) for the experimental and filler word pairs, respectively, used in the test phase). The 66 experimental words were paired with subordinate-related probe words (e.g. “bark” – TREE), and the 66 ambiguous filler words were paired with unrelated probe words (e.g. “pen” – YEAR). Half of the 24 low-ambiguity filler words were paired with related probe words at test (e.g. “mop” – CLEAN) and half with unrelated probe words (e.g. “oil” – HOTEL). All participants were presented with the same target-probe pairs in the test phase and the same filler sentences during the prime phase, with the only difference across participants being the assignment of experimental items to Prime Type (Prior Disambiguation, Subsequent Disambiguation, Unprimed).

Subordinate-related probe words for the experimental items were selected from common word association responses and from the Edinburgh Associative Thesaurus. Criteria for probe word selection were that it did not also relate to the dominant meaning of the target word, did not occur in the prime sentence, and that its pairing with the target word did not form a common two-word phrase. The related and unrelated probe words for filler items were selected from a list of low-ambiguity words.

The target-probe word pairs were matched for frequency and length, and were pre-tested for semantic relatedness in a separate group of participants (N = 16). However, we note that it was not necessary to closely match these variables across experimental items or conditions because we were interested in the effect of Prime Type (Prior Disambiguation, Subsequent Disambiguation, or Unprimed) on meaning access responses to the same word pairs. The semantic relatedness pre-test participants were shown each word pair, one at a time. For each pair, they were asked to say how related the meanings of the two words are to one another using a scale from 1 (“Strongly related, easy to see the association”) to 7 (“Completely unrelated”). Because we were interested in the strength of association specifically with the subordinate primed meanings of the ambiguous words, we presented the prime sentence on each trial, with the target word shown in bold (e.g. “Karen knew that she had to push one last **button**.”) and the meaning access task probe word (e.g. PRESS) written below the sentence. We asked participants to rate the relatedness of the probe word and target word as it is used in the sentence. For experimental items, only the Prior Disambiguation sentences were presented. The filler item pairs were included to encourage the full use of the scale, and the 22 unprimed high-ambiguity filler items

were not presented in this rating task because we did not have prime sentences for these items. Thus there were a total of 134 trials (66 experimental items, 44 high-ambiguity fillers, 24 low-ambiguity fillers). In order to help participants understand the use of the scale, at the start of the session they were presented with four example trials (two low-ambiguity target words, two high-ambiguity target words in subordinate meaning contexts). The results showed that all experimental target-probe pairs were rated above the midpoint of the relatedness scale (median = 2.0, range = 1.2–3.9). The filler pairs were similarly rated as expected (low-ambiguity, related: median = 2.3, range = 1.4–3.4; low-ambiguity, unrelated: median = 6.9, range = 6.1–7.0; high-ambiguity, unrelated: median = 6.9, range = 5.1–7.0).

**Disambiguation post-test.** Participants were invited to complete an unexpected post-test 28 days after they completed the initial session. The purpose of this test was to determine, on an individual and item level, whether participants could successfully disambiguate each experimental prime sentence in the absence of time pressure. In this post-test, participants heard each of the 66 Subsequent Disambiguation sentences, followed by a one-second silence and the critical ambiguous word (e.g. “Sally worried that the ball was going to be too crowded” ... “ball”). In each trial they were asked to generate a synonym or short description of the critical word as it was used in the preceding sentence (appropriate responses for “ball” are e.g. “event”, “dancing party”, “gala”). Responses relating to any inappropriate (e.g. dominant) meaning of the critical word (e.g. a response of “toy” for the word “ball”) indicate that the participant was unable to comprehend the correct meaning of the word in the Subsequent Disambiguation sentence, for instance because s/he is unfamiliar with the subordinate meaning used for that word or unable to infer the appropriate meaning from the context. The results of this test allowed us to conduct a more sensitive test of word-meaning priming, because we were able to exclude individual trials from the test task in which the participant likely misunderstood the prime sentence for that word. We asked participants to complete the synonym generation task for all 66 experimental items, regardless of what condition the word appeared in for that participant, as a strict measure of meaning familiarity for each participant and item. The Subsequent Disambiguation sentences were used for this test because we were interested in comprehension with this more difficult sentence construction, where there is a need to inhibit the dominant meaning. Participants were given example sentences and responses before starting the task, and were

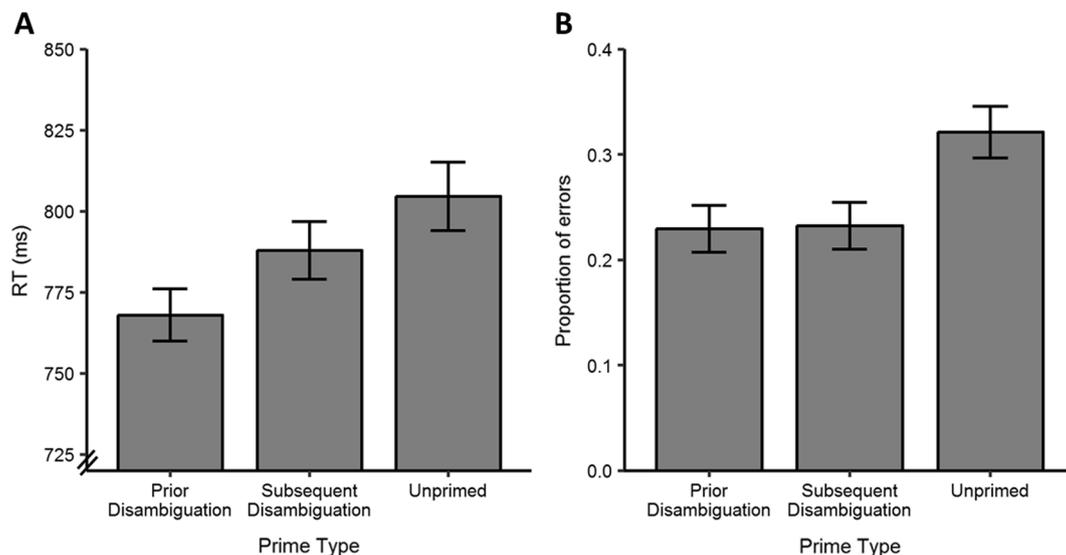
allowed to take breaks throughout the session.

## Results

Performance in the sentence exposure task was good (mean proportion correct = 0.93, range = 0.76–0.98). Accuracy for Prior Disambiguation sentences was similar to that for Subsequent sentences (mean proportion correct 0.92 and 0.93 respectively). One item (“star”) was removed from the analyses because the post-test showed that the sentence did not fully disambiguate the intended subordinate meaning (celebrity) from a third meaning (newspaper title).

Fig. 4 shows the participant grand mean response times for correct responses (A) and proportions of errors (B) in the meaning access task according to whether the target ambiguous word was primed with a Prior Disambiguation sentence, primed with a Subsequent Disambiguation sentence, or Unprimed. For the response time analysis, responses that were faster than 300 ms or slower than 2500 ms were removed, as these were assumed to reflect accidental key presses and lapses in attention, respectively. This trimming resulted in the exclusion of 16 (0.3%) out of 5785 (89 participants  $\times$  65 target items) total responses. Of the remaining 5769 responses, 1506 responses (26.1%) were incorrect and removed. The relatively high error rate in this task was expected given the use of subordinate meanings (mean dominance = 0.21, range = 0.00–0.50), and item mean error rates in the Unprimed condition correlated with the baseline dominance of the sentence-consistent meanings, with increased errors for less frequent meanings,  $r(63) = .48$ ,  $p < .001$ . The log-transformed RT data met the assumptions of homogeneity of variance and normally distributed residuals required for LME modelling, so no further trimming was needed (Baayen & Milin, 2015).

The data was analysed with linear mixed effects modelling using R software (R Core Team, 2013) and the “lme4” (Bates et al., 2015), “lmerTest” (Kuznetsova, Brockhoff, & Christensen, 2015) and “phia” (Rosario-Martinez, 2015) packages. The fixed factor was Prime Type, which was Helmert coded. The first contrast coded the primed vs unprimed conditions (Prior: 1/3, Subsequent: 1/3, Unprimed: -2/3) and the second contrast coded the Prior vs Subsequent Disambiguation sentence types (Prior: -1/2, Subsequent: 1/2, Unprimed: 0). The model was first constructed with the maximal random effects structure. However, the by-subject random slopes for Prime Type produced perfect correlations (both with the by-subject intercept and between the two



**Fig. 4.** Experiment 2 response times (A) and proportions of errors (B) from relatedness decisions to ambiguous and subordinate-related word pairs (e.g. bank – RIVER) in the meaning access task. Target ambiguous words were primed with Prior or Subsequent Disambiguation sentences, or were Unprimed. Bars reflect the participant grand means, and error bars show 95% confidence intervals of the participant means, adjusted to remove between-subject variance (Morey, 2008).

contrasts), indicating that the model was over-parameterised (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). We therefore removed this term, leaving a by-subject random intercept and by-item random intercept and slope for Prime Type.

The model revealed a significant effect of Prime Type,  $\chi^2(2) = 27.25$ ,  $p < .001$ ; see Fig. 4A. The contrast comparing primed vs unprimed levels was significant,  $\beta = -0.04$ ,  $SE = 0.01$ ,  $t(53.6) = -4.69$ ,  $p < .001$ . This effect was characterised by faster responses to the word pairs when the target ambiguous word was primed, compared to when it was unprimed. The contrast comparing the two prime sentence types was also significant,  $\beta = 0.02$ ,  $SE = 0.01$ ,  $t(416) = 2.78$ ,  $p = .006$ . Consistent with the results of Experiment 1, there was a priming advantage for the Prior Disambiguation condition; responses in the meaning access task were faster when target words were primed with a Prior sentence than with a Subsequent sentence. Pairwise comparisons with Holm adjustment for multiple comparisons showed that both prime sentence types produced significant priming: Prior vs Unprimed ( $\chi^2(1) = 31.32$ ,  $p < .001$ ), Subsequent vs Unprimed ( $\chi^2(1) = 8.90$ ,  $p = .006$ ).

In the accuracy analysis, we again removed the 16 responses that were faster than 300 ms or slower than 2500 ms. The remaining 5769 trials were entered into a logistic mixed effects model with accuracy as the dependent variable, and with the same structure as in the RT analysis. Prime Type was a Helmert coded fixed factor (contrast 1: Prior = 1/3, Subsequent = 1/3, Unprimed = -2/3; contrast 2: Prior = -1/2, Subsequent = 1/2, Unprimed = 0). The full model failed to converge, so the two terms that accounted for the least variance were removed: the by-subject and by-item random slopes for Prime Type. This left by-subject and by-item random intercepts in the final model.

Prime Type had a significant effect on accuracy in the meaning access task,  $\chi^2(2) = 69.65$ ,  $p < .001$ ; see Fig. 4B. Accuracy was significantly higher in the primed vs unprimed conditions,  $\beta = 0.60$ ,  $SE = 0.07$ ,  $z = 8.43$ ,  $p < .001$ . The prime sentence type contrast (Prior vs Subsequent Disambiguation) was not significantly related to response accuracy,  $\beta = -0.02$ ,  $SE = 0.09$ ,  $z = -0.17$ ,  $p = .866$ . Pairwise comparisons with Holm adjustment for multiple comparisons confirmed that, compared to the Unprimed condition, there were fewer errors in both the Prior Disambiguation ( $\chi^2(1) = 53.07$ ,  $p < .001$ ) and Subsequent Disambiguation ( $\chi^2(1) = 50.74$ ,  $p < .001$ ) conditions.

#### Analysis of sentence probe type

Unlike Experiment 1, here the target sentences in the sentence exposure task were followed by a mix of semantically-related and unrelated probe words. We took the opportunity to conduct an exploratory analysis of the RTs and accuracy from the meaning access test task, using both the Prime Type (Prior, Subsequent, or Unprimed) and Sentence Probe Type (Related or Unrelated) as predictors. The Sentence Probe Type factor was coded as: Related = 0.5, Unrelated = -0.5.

In the RT analysis, the main effect of Prime Type remained significant,  $\chi^2(2) = 25.59$ ,  $p < .001$ ; see Fig. 5A. The model coefficient for the contrast comparing primed vs unprimed levels was significant,  $\beta = -0.04$ ,  $SE = 0.01$ ,  $t(53.60) = -4.74$ ,  $p < .001$ , as was the coefficient for the Prior vs Subsequent Disambiguation contrast,  $\beta = 0.02$ ,  $SE = 0.01$ ,  $t(492.90) = 2.83$ ,  $p = .005$ .

There was also a significant main effect of Sentence Probe Type such that response times were faster for target words that were primed with a

sentence that was followed by a Related vs Unrelated probe word,  $\beta = -0.08$ ,  $SE = 0.02$ ,  $t(61.20) = -3.17$ ,  $p = .002$ ,  $\chi^2(1) = 9.60$ ,  $p = .002$ . The main effect of Sentence Probe Type was not solely driven by a significant difference in the Subsequent sentence condition ( $\chi^2(1) = 12.51$ ,  $p = .001$ ), rather it was also significant in the Prior Disambiguation condition ( $\chi^2(1) = 11.82$ ,  $p = .001$ ) but not in the Unprimed<sup>2</sup> condition ( $\chi^2(1) = 2.88$ ,  $p = .090$ ).

There was a significant interaction between Prime Type and Sentence Probe Type,  $\chi^2(2) = 8.17$ ,  $p = .017$ . The model coefficient for the interaction between Sentence Probe Type and priming (primed vs unprimed) was significant,  $\beta = -0.05$ ,  $SE = 0.02$ ,  $t(53.60) = -2.83$ ,  $p = .007$ , whereas the coefficient for the interaction with prime sentence type (Prior vs Subsequent) was non-significant,  $\beta = -0.01$ ,  $SE = 0.02$ ,  $t(493.90) = -0.81$ ,  $p = .420$ . Pairwise interactions with Holm adjustment for multiple comparisons showed that the interaction between Subsequent vs Unprimed conditions and Sentence Probe Type was significant,  $\chi^2(1) = 8.09$ ,  $p = .013$ . Meaning access response times were faster for Subsequent vs Unprimed target words, but only when the Subsequent Disambiguation prime sentence was followed by a semantically-related probe word,  $\chi^2(1) = 19.20$ ,  $p < .001$ . When the Subsequent sentence was followed by an unrelated probe word, then there was no significant priming effect,  $\chi^2(1) < .01$ ,  $p = .953$ . The interaction between Prior vs Unprimed and Sentence Probe Type was non-significant,  $\chi^2(1) = 4.88$ ,  $p = .055$ .

In the accuracy analysis, the main effect of Prime Type remained significant,  $\chi^2(2) = 30.18$ ,  $p < .001$ ; see Fig. 5B. The priming contrast was significant, with more errors in unprimed vs primed conditions,  $\beta = 0.59$ ,  $SE = 0.09$ ,  $z = 6.61$ ,  $p < .001$ . The prime sentence type contrast remained non-significant,  $\beta = -0.08$ ,  $SE = 0.11$ ,  $z = -0.72$ ,  $p = .474$ . Pairwise comparisons with Holm adjustment showed that the Unprimed condition resulted in more meaning access errors compared to both the Prior ( $\chi^2(1) = 9.64$ ,  $p = .006$ ) and Subsequent ( $\chi^2(1) = 6.40$ ,  $p = .023$ ) primed conditions.

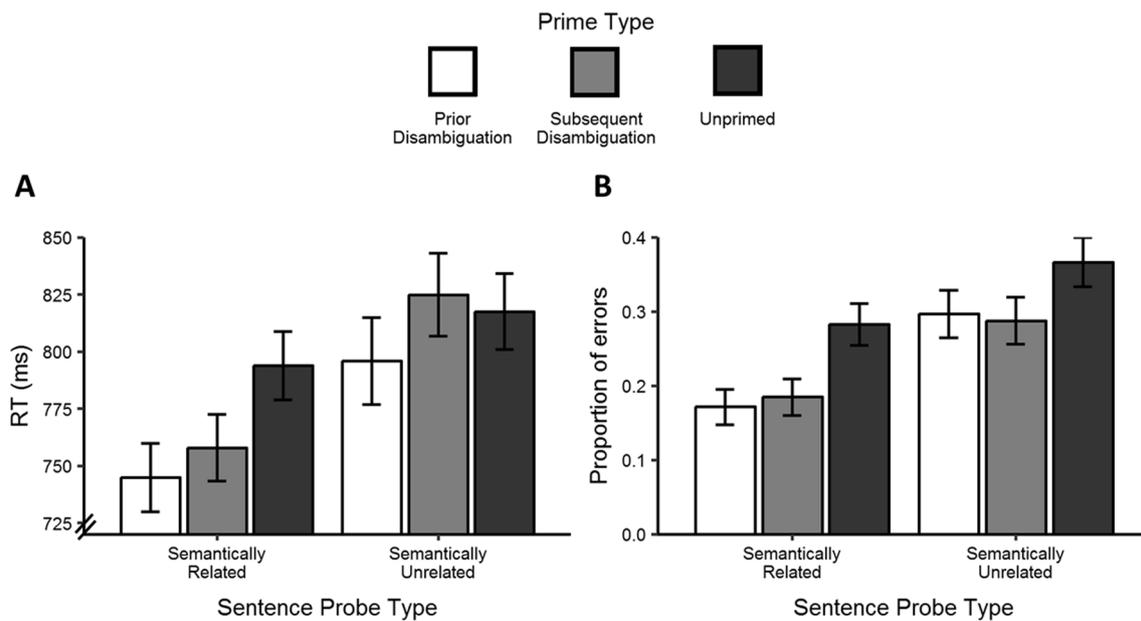
There was also a significant main effect of Sentence Probe Type on error rates,  $\beta = -0.70$ ,  $SE = 0.27$ ,  $z = -2.57$ ,  $p = .010$ ,  $\chi^2(1) = 6.30$ ,  $p = .012$ . There were more errors in the meaning access task for words that were primed with a sentence followed by an Unrelated probe word, compared to when the prime sentence was followed by a Related probe word.

Model comparisons revealed that there was no significant interaction between these variables,  $\chi^2(2) = 5.46$ ,  $p = .065$ . However the model coefficient for the interaction between the priming contrast and Sentence Probe Type was significant,  $\beta = -0.37$ ,  $SE = 0.17$ ,  $z = -2.13$ ,  $p = .03$ , due to a larger priming advantage for Related vs Unrelated probe types on errors for the primed vs unprimed conditions. The coefficient for the interaction between Sentence Probe Type and the prime sentence type contrast was non-significant,  $\beta = 0.23$ ,  $SE = 0.20$ ,  $z = 1.15$ ,  $p = .251$ , showing that the effect of Sentence Probe Type was not reliably different for Prior and Subsequent sentence conditions.

#### Disambiguation post-test

Seventy-three (82.0%) of the 89 participants completed the post-test. Overall performance was very good, with a near-ceiling distribution of proportions of correct responses (i.e. responses consistent with the meaning used in the sentence). Across participants, the median

<sup>2</sup> The probe relatedness during the sentence exposure task could not have had any effect on response times at test for unprimed words. However, Sentence Probe Type was not counterbalanced across words via different versions of the experiment because we did not predict that this factor would have any effect on priming. Instead, the prime sentences for each item were always paired with either a semantically-related or unrelated probe word. Thus it is possible that the effect of Sentence Probe Type on response times reflects (at least in part) a difference between the two word sets, and if so, this difference could also be present in the unprimed condition.



**Fig. 5.** Experiment 2 response times (A) and proportions of errors (B) from relatedness decisions to ambiguous and subordinate-related word pairs (e.g. bank – RIVER) in the meaning access task. Responses are shown according to the prime sentence type and the type of probe word that was presented after the prime sentence. The probe words presented after each prime sentence were either semantically-related (left bars) or semantically-unrelated (right bars). Target ambiguous words were primed with a Prior (white) or Subsequent (light gray) Disambiguation sentence, or were Unprimed (dark gray). Bars reflect the participant grand means, and error bars show 95% confidence intervals, adjusted to remove between-subject variance (Morey, 2008).

proportion of correct responses was 0.97 (IQR = 0.06, range = 0.55–1.00). A response was coded as incorrect if it was clearly related to the dominant meaning or another context-inappropriate meaning, and across participants, these incorrect responses were typically infrequent (mdn = 0.02, IQR = 0.05, range = 0.00–0.45). The remaining responses were categorised as ‘other’ (mdn = 0.02, IQR = 0.03, range = 0.00–0.25) because they met one of the following conditions: (1) the meaning of the response was ambiguous, (2) the participant misheard the word, (3) the participant responded with “I don’t know”, or (4) a technical problem prevented the audio from playing. We repeated the analyses of Prime Type on this subset of participants, and on a further subset of the meaning access trials in which the participant responded correctly to the item in the post-test. This subset of participants and items allowed us to be more confident that participants were familiar with the subordinate meanings used in the prime sentences and interpreted them as intended. The subset analyses showed the same pattern of results in both RTs and accuracy as in the full set of participants and trials.

### Discussion

Recent experience with ambiguous words in subordinate-meaning contexts increased the speed with which that same meaning was accessed when the ambiguous word was encountered again later on. The results were consistent with Experiment 1: the priming effect was significantly larger for words primed in Prior Disambiguation versus Subsequent Disambiguation sentences. These results suggest that the meaning preference results of Experiment 1 reflect a modulation of immediate word-meaning access (which are apparent in speeded tasks), rather than a slower, more deliberate strategy (apparent only in non-speeded tasks like word association).

The priming advantage for words presented in Prior Disambiguation sentences appears to provide support for the hypothesis that priming is driven by initial meaning access and selection during the prime encounter. However, this conclusion is modified by the finding that the type of probe word that appeared after the prime sentence also modulated the magnitude of word-meaning priming. There was an overall

reduction in priming for sentences that were followed by unrelated probe words, regardless of the sentence type. Furthermore, for the Subsequent Disambiguation sentences, which were associated with reduced priming compared to Prior sentences, priming was only reliable when the sentence was followed by a *related* probe word. When the Subsequent Disambiguation sentence was followed by an *unrelated* probe word, there was no significant priming effect. This is consistent with the results of Experiment 1, in which the experimental sentences were always followed by unrelated probe words and we saw no evidence of priming in the Subsequent Disambiguation condition. The present experiment goes further in showing that priming from Subsequent Disambiguation sentences is possible when the sentence is followed by a semantically-related probe word.

This effect of Sentence Probe Type on word-meaning updating was surprising. The two theoretical possibilities that we anticipated in the introduction both involved learning mechanisms that operate during initial sentence processing and comprehension. However, since word-meaning priming appears to depend on the semantic relationship of the probe word that followed the sentence, this conversely suggests that word-meaning priming might involve cognitive operations that arise throughout an entire trial and not just during initial sentence comprehension. The impact of the change in probe relatedness in Experiment 2 was that word-meaning priming for ambiguous words in Subsequent Disambiguation sentences became apparent, but only when the prime sentence was followed by a semantically-related probe. While the presence of word-meaning priming indicates successful disambiguation, the absence of significant word-meaning priming (for Subsequent Disambiguation sentences followed by unrelated probes) provides uncertain evidence regarding the degree of comprehension that occurred during the prime encounter. Nonetheless, in the absence of a statistical interaction between prime sentence type and sentence probe type we cannot conclude that there is any effect of related probes that is specific to Subsequent Disambiguation sentences.

The specific reason for the effect of Sentence Probe Type on word-meaning priming remains unclear. The experiments that follow were designed to test two possible explanations for this effect. First, it is possible that the semantically-related probe words that appeared after

the prime sentences provided an additional disambiguation cue that increased the probability of successful disambiguation and hence enhanced the magnitude of word-meaning priming. While, numerically, this effect seems most apparent for the relatively difficult to understand Subsequent Disambiguation sentences, (e.g. “The girl wanted to ask when the post would be advertised” - EMPLOYMENT), the results indicated that Prior Disambiguation sentences seemed to be similarly influenced by the type of sentence probe word. The second, converse explanation is that, in the context of the speeded decision-making task, the semantically-unrelated probe words interfered with normal sentence comprehension, for instance because these unrelated words were presented quickly after the sentence, and because a correct ‘no’ (unrelated) response could be made to the probe word without successful (re) interpretation of the ambiguous word.

Importantly, under both these accounts, the failure to find significant priming for the Subsequent sentences does not indicate that participants *cannot* learn from exposure to ambiguous words presented in this Subsequent Disambiguation sentence structure, but instead was a side-effect of participants’ failure to fully disambiguate the word meanings from the encounter with the prime sentence. Indeed, the present experiment showed that significant priming from Subsequent sentences is possible when the sentence is followed by a related probe word. The results of the post-test confirm that, in an unsped situation, our participants correctly disambiguated 97% of these sentences on average. Furthermore, analysis of a subset of trials in which participants could correctly disambiguate the Subsequent Disambiguation sentence at post-test revealed the exact same pattern of results. This suggests that any failure to disambiguate the Subsequent Disambiguation sentences during the sentence exposure task was not the result of an intrinsic inability to reinterpret these sentences correctly (even in the absence of a semantically-related probe word), but was instead caused by the specific prime task conditions.

However, because we did not anticipate an effect of the type of sentence probe word, the Sentence Probe Type factor was not counter-balanced across words; rather, each item was always paired with either a semantically-related or unrelated sentence probe word in the prime phase. Thus, it is important to investigate this possible task effect and the role of additional post-sentence disambiguating context in a fully counter-balanced design. Experiments 3 and 4 that follow were conducted to determine whether the type of probe word presented after the prime sentence had a true effect on priming, and whether exposure to the prime sentences in a different task context would still produce greater priming in the Prior versus Subsequent sentence condition.

### Experiment 3

There are at least two non-mutually-exclusive explanations for the unanticipated effect of Sentence Probe Type on word-meaning priming in Experiment 2: (i) the semantically-related probe words were boosting the priming effect, and/or (ii) the semantically-unrelated probe words were disrupting normal sentence comprehension and thereby suppressing word-meaning priming. The results from Experiment 2 do not distinguish between these possibilities and led us to ask additional questions in Experiments 3 and 4. One question is whether the magnitude of lexical-semantic retuning can be affected by additional biasing context that follows after the word can be disambiguated. A second question is whether and how lexical-semantic retuning might be sensitive to the situational context, for instance, when disambiguation is not strictly necessary in order for the listener to achieve their goal. The answers to these questions are important for understanding precisely when and how lexical-semantic representations are updated in response to experience.

For the first explanation of the Experiment 2 Sentence Probe Type effect mentioned above, there are multiple possible mechanisms through which this ‘boosting’ effect might occur. One mechanism is that the semantically-related probe word initiated a second association

between the ambiguous word and subordinate meaning, separate from the first word-meaning association that forms from the sentence context. However, previous studies show that the word-meaning priming effect does not reliably increase with multiple repetitions of an ambiguous word that is disambiguated toward the same subordinate meaning, when these repetitions are presented in immediate succession, compared to a single word-meaning encounter (Betts et al., 2018). We therefore consider this explanation unlikely.

Instead, we propose that any potential ‘boosting’ effect of the semantically-related probe word was caused by a more direct influence on prime sentence comprehension. For instance, it may be that, at least in some cases, the semantically-related probe word was either necessary for successful disambiguation, or it provided confirmation that the unexpected subordinate interpretation was correct. Thus the semantically-related probe words could have produced an overall increase in the priming effect due to, for instance, a greater probability of successful disambiguation of the ambiguous word in the prime sentences, or greater confidence that the appropriate meaning was selected. However, we note that our participants were able to correctly comprehend the Subsequent Disambiguation sentences in our post-test (when there were no probe words that followed the sentences), so it might be that this explanation would only apply when sentences are processed under time pressure and/or as part of a semantic decision-making task. In any case, this possibility would suggest that lexical-semantic retuning increases with additional disambiguating context that is encountered after the ambiguous word and disambiguation, and would therefore challenge the view that word-meaning updating predominantly occurs during initial word activation and meaning selection.

By the second account, it may be that the sentence-probe relatedness task with an unrelated probe encourages relatively shallow processing; listeners can successfully complete the task (i.e., correctly deciding that the probe word is unrelated to the sentence) without forming a detailed representation of the sentence meaning and disambiguating the ambiguous word. Indeed, it is possible that interpretation of the prime sentences was delayed until the probe word appeared, and only then was the sentence analysed to find a relationship with the probe word. This possibility would be consistent with research from the ‘good enough’ processing literature showing that sentence interpretations can remain in an underspecified state until further processing is required (e.g. Christianson, 2016; Ferreira, Bailey, & Ferraro, 2002). For instance, there is some evidence for (i) longer reading times for comprehension questions that follow after garden-path sentences (Wonnacott, Joseph, Adelman, & Nation, 2016), and (ii) an effect of comprehension question wording on the ultimate interpretation of garden-path sentences (Christianson & Luke, 2011). Critically, the proposed effect of the unrelated probe word on sentence interpretation would be specific to conditions in which precise disambiguation is not strictly required in order for the listener to achieve their goal, and so we would not necessarily expect to see the same effect of additional disambiguation in other listening conditions.

In order to address this latter explanation, in the present experiment we removed the sentence probe words from the sentence exposure task and instead used a written continuation sentence after each prime sentence. We also made the sentence exposure task self-paced. Participants were asked to attend to sentence pairs and respond to occasional comprehension questions presented at the end of a trial (questions were only presented after filler sentences). The reason for these changes to the sentence exposure task was to allow language comprehension processing to continue as normal after the experimental prime sentences, without disruption from the appearance of the probe word, and without the need to make a speeded decision. Thus, if the *unrelated* probe words were disrupting processing for the sentences, then we should see priming for the Subsequent sentences in the present experiment because here typical processing can continue for a longer period of time, without disruption from a decision-making task.

Moreover, in order to investigate the first explanation, that is, any

potential beneficial effect of the *related* probe word on priming, we manipulated the content of the continuation sentences. In one condition (Biased), the continuation sentence contained additional context that confirmed the meaning used in the first sentence, and in another condition (Neutral), the continuation sentence did not contain any additional disambiguating information. For example, the Prior Disambiguation prime sentence “Sally worried about how crowded the ball would be.” was either followed with the Biased continuation sentence “She had already bought a formal dress so decided to attend anyway.” or the Neutral continuation sentence “She was the type of person who was concerned about all sorts of things.” Therefore if the *related* probe words were increasing the magnitude of priming in Experiment 2, then here we should see greater priming from Biased compared to Neutral continuation sentences, because the Biased continuation sentences (like the semantically-related probe words from Experiment 2) provide additional cues to disambiguation.

We used this design in two experiments: Experiment 3 used word association as a test of meaning preference, and Experiment 4 used speeded semantic relatedness as a test of meaning access. The former allowed us to examine word-meaning preferences in the absence of any biasing context, while the latter provided a speeded measure of access to the primed meaning. The use of these two measures also allowed for a direct comparison with the results of Experiments 1 and 2.

## Method

### Participants

Data was collected over the internet from 121 volunteers (67 women; mean age = 30, range = 18–49). Of these, 12 participants were excluded (11 had prime-to-test durations that were over the 30 min limit, and one had an unusually high number of incorrect responses to comprehension questions and ambiguous/uninterpretable meaning preference responses). The excluded participants were replaced until we reached the *a priori* aim of 18 participants per version of the experiment. There was one additional participant in one version of the experiment due to accidental over-recruitment, resulting in 109 participants included in the analysis. Inclusion criteria, recruitment and payment rates were the same as in Experiment 2.

### Design

The experimental design was three (Prime Type: Prior Disambiguation, Subsequent Disambiguation, Unprimed) × two (Continuation Sentence Type: Biased, Neutral). Each factor was both within-subject/between-item and within-item/between-subject. The crossing of Prime Type and Continuation Sentence Type resulted in six conditions, so six versions of the experiment were created to counterbalance the assignment of experimental word list (65/6 = 10–11 items) to condition. Participants were assigned to one of the six versions, and only encountered each experimental word in a single condition.

Because the two levels of Continuation Sentence Type were only relevant to the two primed levels of Prime Type (Prior and Subsequent), there were two equivalent conditions in this design: Unprimed-Biased and Unprimed-Neutral. For this reason, if there were a main effect of Continuation Sentence Type then we expected this to manifest as an interaction with Prime Type, where a true effect of Continuation Sentence Type should only be observed in the primed conditions.

### Materials and procedure

The stimuli were the same as those used in Experiment 2, with a few exceptions. First, the experimental ambiguous word ‘star’ was removed, leaving 65 items (mean dominance of prime-consistent meanings for words in isolation = 0.21, SD = 0.14, range = 0.00–0.50; mean dominance of prime-consistent meanings for words in Subsequent Disambiguation cut-off sentences = 0.24, SD = 0.15, range = 0.00–0.50; see [Tables S1 and S2 in the Supplementary Material](#) for a list of experimental words and sentences). Second, the sentence probe words from the

sentence exposure task were replaced with full continuation sentences (see [Table S5 in the Supplementary Material](#) for continuation sentences for experimental items and [Table S6](#) for continuation sentences for filler items). These continuation sentences either provided additional disambiguating context that was consistent with the meaning used in the prime sentence (Biased), or did not provide any additional disambiguating context (Neutral). For each experimental item, there was a Biased and Neutral sentence. The Biased sentence elaborated on an aspect of the sentence related to the subordinate meaning of the ambiguous word, whereas the Neutral sentence often shifted focus to another aspect of the person/situation and was always compatible with multiple meanings of the ambiguous word. The Biased and Neutral sentences were matched for length. See [Table 3](#) for example stimuli.

We also created 22 comprehension questions that were presented after an unpredictable subset of trials (see [Table S7 in the Supplementary Material](#)). These questions only followed filler sentences so that we could be sure that priming for experimental words was not affected by a follow-up question. The questions were designed to be relatively easy single-word answers based on the preceding sentence pair. For instance, for the sentence pair: “Tim told his wife that the suit he had chosen was hearts. His wife was distracted by the television and wasn’t really listening.”, the comprehension question was: “What suit did Tim choose?” In some cases the question required the participant to make an inference about the situation described by the two sentences. Most questions required an understanding of the first (spoken) sentence, in order to encourage participants to attend to the spoken sentence, rather than rely solely on the written, self-paced continuation sentence to answer the question. The comprehension questions followed a mix of both high-ambiguity and low-ambiguity filler sentence pairs, so that the presence/absence of an ambiguity in the sentence did not become a valid cue to the upcoming presentation of a comprehension question.

The procedure was similar to Experiment 1, except that here we used a self-paced listening/reading task in the prime phase (see [Fig. 6](#)). In this task, each trial began with a spoken sentence. After the spoken sentence

**Table 3**

Experiment 3 stimuli for a single item (bank) and all six Prime Type × Continuation Sentence Type conditions. For each participant, each experimental item was assigned to only one of the six conditions. Ambiguous words are underlined and disambiguation words are in italics in the prime and continuation sentences. The Unprimed-Biased and Unprimed-Neutral conditions were equivalent, but were categorised separately to create a balanced factorial design.

		Sentence exposure task	Meaning preference task
Prime type	Continuation sentence type	Prime sentence (auditory) and continuation sentence (visual)	Target word (auditory)
Prior	Biased	The old man had a long way to <i>swim</i> as he headed for the <u>bank</u> . After reaching the <i>shore</i> , he rested briefly and then <i>swam</i> back.	bank
Prior	Neutral	The old man had a long way to <i>swim</i> as he headed for the <u>bank</u> . He thought about going back but then decided to continue.	bank
Subsequent	Biased	The old man headed for the <u>bank</u> but he had a long way to <i>swim</i> . After reaching the <i>shore</i> , he rested briefly and then <i>swam</i> back.	bank
Subsequent	Neutral	The old man headed for the <u>bank</u> but he had a long way to <i>swim</i> . He thought about going back but then decided to continue.	bank
Unprimed	Biased	–	bank
Unprimed	Neutral	–	bank

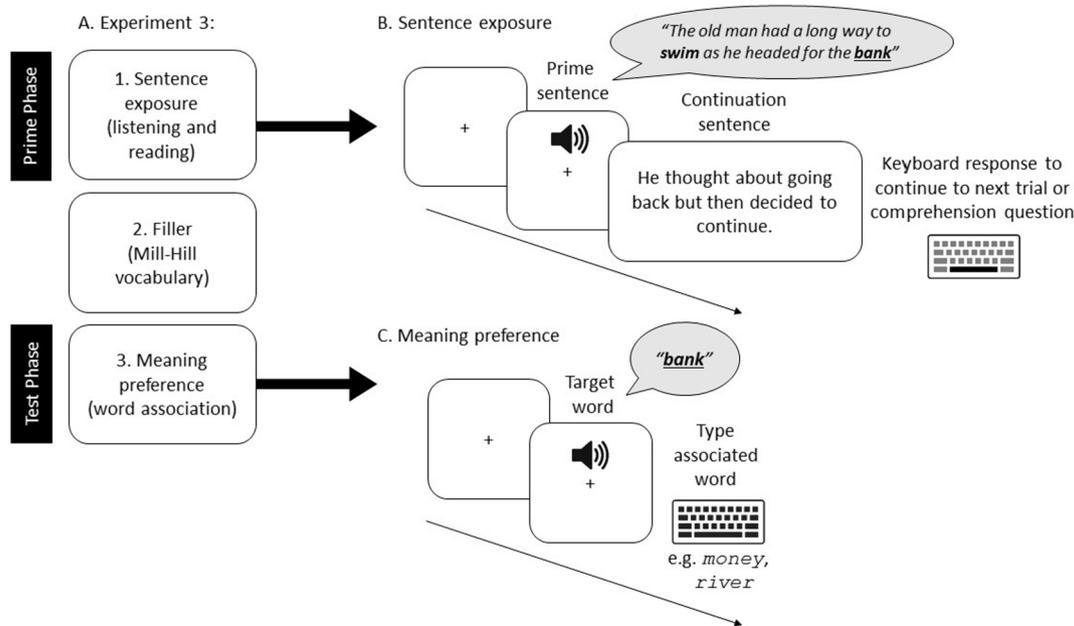


Fig. 6. Overview of Experiment 3 task order (A), and procedures for the sentence exposure task (B) and the meaning preference task (C).

ended, there was a 500 ms fixation, and then the written continuation sentence was displayed on the screen. Participants were asked to press the space bar to move on once they were certain that they understood the sentences. Participants were told that they would sometimes be given a comprehension question at the end of the trial to make sure that they were paying attention.

As in Experiment 2, the Mill-Hill vocabulary test was used as a filler task between the sentence exposure and meaning preference tasks, and served as an indicator of English proficiency. After the sentence exposure and filler tasks, participants were given instructions for the meaning preference task. They were told that they would hear words, one at a time, and for each word they should type in an associated word (unlike Experiment 1, here the participant did not need to first type the target word). There was no meaning clarification task in this experiment because the results of Experiment 1 showed that this was not a reliable method for categorizing responses. The mean duration between the presentations of each ambiguous word in the prime and test phases was 21.1 min (SD = 3.2, range = 15.9–29.4). There was no disambiguation post-test in this experiment because the results of Experiment 2 showed that (1) participants generally interpreted the subsequent disambiguation sentences with the intended meaning, and (2) the overall pattern of results did not change when we excluded trials in which the participant responded incorrectly in the disambiguation post-test for that item.

#### Response coding

Unlike Experiment 1, here participants did not self-code their meaning preference responses in a meaning clarification task. This change was made because, in Experiment 1, participants' response coding contained errors that needed correction, and because we developed a more efficient system for automatically coding responses based on a large set of previously-coded responses to the same items. The automated coding script coded 72% of the total responses, and 27% of responses could not be automatically coded because the participant's response to that item was unique to our data set. The remaining 1% of trials were not automatically coded because they were in response to an item ("record") that was not present in our set of previously-coded responses. The responses that were not automatically coded for meaning were manually coded by the authors (RAG and JMR).

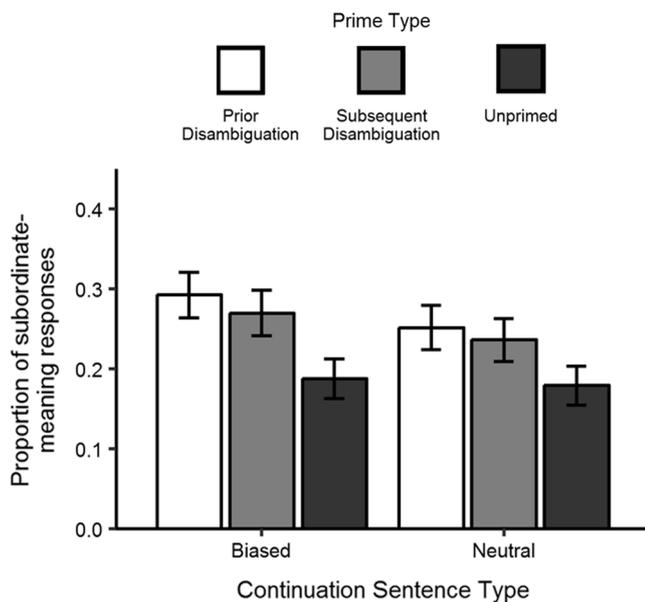
#### Results

Performance on the comprehension questions in the sentence exposure task was good (mean proportion correct = 0.95, SD = 0.06, range = 0.73–1.00). Of the incorrect responses to comprehension questions, most were indicative of the participant mishearing a word (e.g. responding with 'mat' rather than 'bat', 'shelves' rather than 'shells'). Overall these results show that participants were attending to the sentence pairs and able to answer basic questions about the sentences, including some that contained ambiguous words in subordinate meaning contexts. The average proportion correct on the Mill Hill vocabulary test was 0.61 (SD = 0.13, range = 0.27–0.88).

Fig. 7 shows the participant grand mean proportions of meaning preference responses that were consistent with the primed meaning within each of the Prime Type  $\times$  Continuation Sentence Type conditions. Of the 7085 meaning preference responses (109 participants  $\times$  65 items), 334 were removed due to being ambiguous between meanings, uninterpretable, or because they indicated that the participant misheard the word. The remaining 6751 responses were analysed with a logistic mixed effects model. The fixed factors were Prime Type and Continuation Sentence Type. Prime Type had three levels (Prior Disambiguation, Subsequent Disambiguation, Unprimed), which was assigned Helmert contrasts to provide separate estimates for effects of primed vs unprimed conditions (Prior:  $-1/3$ , Subsequent:  $-1/3$ , Unprimed:  $2/3$ ) and of Prior vs Subsequent Disambiguation sentence types (Prior:  $-1/2$ , Subsequent:  $1/2$ , Unprimed: 0). Continuation Sentence Type had two levels, which were sum-coded (Neutral:  $-1/2$ , Biased:  $1/2$ ).

The model revealed a significant effect of the Prime Type factor,  $\chi^2(2) = 34.96$ ,  $p < .001$ . The primed vs unprimed contrast was significant,  $\beta = -0.62$ , SE = 0.10,  $z = -6.06$ ,  $p < .001$ . The contrast for Prior vs Subsequent sentence types was non-significant,  $\beta = -0.17$ , SE = 0.10,  $z = -1.67$ ,  $p = .095$ . Pairwise comparisons with Holm adjustment for multiple comparisons revealed that both the Prior and Subsequent Disambiguation prime conditions significantly differed from the Unprimed condition ( $\chi^2(1) = 36.98$ ,  $p < .001$  and  $\chi^2(1) = 22.69$ ,  $p < .001$ , respectively).

There was no reliable effect of the Continuation Sentence Type,  $\chi^2(1) = 2.86$ ,  $p = .091$ . Because the two Continuation Sentence Type levels were equivalent in the Unprimed condition, we did not predict an effect of Continuation Sentence Type in this condition; rather, any effect of



**Fig. 7.** Experiment 3 proportions of meaning preference responses that were consistent with the primed subordinate meaning for ambiguous words in the Prior (white) and Subsequent (light gray) Disambiguation prime conditions, and in the Unprimed (dark gray) condition. The prime sentences were followed by either Biased (left bars) or Neutral (right bars) Continuation Sentences. The Biased and Neutral conditions were identical for the Unprimed words; the Continuation Sentence Type factor did not apply to the Unprimed condition, however the Unprimed words were split into two levels to create a factorial design. Bars show the participant grand means, and error bars show 95% confidence intervals, adjusted to remove between-subject variance (Morey, 2008).

Continuation Sentence Type should only affect the two prime levels of the Prime Type factor (Prior and Subsequent Disambiguation). We therefore expected that any main effect of the Continuation Sentence Type would manifest as an interaction between the two factors. The interaction between the Prime Type and Continuation Sentence Type was non-significant,  $\chi^2(2) = 0.35, p = .841$ . Thus we have no evidence for a reliable difference in priming between the Biased and Neutral Continuation Sentence Types, or for a greater effect of the Biased (vs Neutral) Continuation Sentences on Subsequent (vs Prior) Disambiguation conditions.

### Discussion

In Experiment 3 we replicated the word-meaning priming effect with Prior Disambiguation sentences, which has now been observed many times (Betts et al., 2018; Gaskell et al., 2019; Gilbert et al., 2018; Rodd et al., 2016, 2013). Importantly, we also found significant priming for words presented in Subsequent Disambiguation sentences. This effect was not observed with the same meaning preference task in Experiment 1, when participants were exposed to the prime sentences as part of a semantic relatedness task, and where the prime sentences were always followed by semantically-unrelated sentence probe words. The null effect of priming with Subsequent Disambiguation sentences in Experiment 1 is difficult to interpret because it could have been the result of several factors including insufficient statistical power, the changes to the materials, and/or the change in task during the test phase. Nonetheless, the results of Experiment 2 are consistent with those of Experiment 1 in showing an absence of reliable priming from Subsequent Disambiguation sentences paired with an unrelated sentence probe word. The Experiment 2 results go further in showing that priming from the Subsequent Disambiguation sentences can occur when the sentence is followed by a semantically-related probe word. The results of Experiment 3

are therefore consistent with those from Experiment 2 in showing that word-meaning priming with Subsequent Disambiguation sentences is possible, this time using a different sentence exposure task (sentence pairs with occasional comprehension questions, rather than semantic relatedness), and a different test (meaning preference, rather than meaning access). In the present experiment, we observed a numerical difference in priming between Prior and Subsequent sentences in the same direction as in previous experiments (approximately 10% for Prior Disambiguation versus approximately 7% for Subsequent Disambiguation) however this difference was not statistically significant.

We also did not find any evidence of a main effect of Continuation Sentence Type, or interaction between the sentence structure (Prior or Subsequent Disambiguation) and Continuation Sentence Type (Biased or Neutral). In particular, there was no indication of greater priming from Biased (vs Neutral) Continuation sentences, either in general or specifically when following the more difficult Subsequent Disambiguation sentences. Although it is difficult to interpret null results, the absence of this interaction suggests that additional disambiguating context given after the prime sentence does not significantly affect the magnitude of word-meaning priming, and (to the extent that word-meaning priming is an indicator of comprehension) does not appear to change the likelihood of sentence comprehension. Thus the effect of Sentence Probe Type seen in Experiment 2 was likely due to task-specific interference from the *unrelated* probe words that appeared after the prime sentences, rather than due to an enhancing effect of the semantically-related sentence probes. Similarly, the design used in Experiment 1, in which experimental prime sentences were always followed by semantically-unrelated probes, may have underestimated the magnitude of word-meaning priming in other sentence exposure conditions.

### Experiment 4

The pattern of results from Experiments 2 and 3 show that experience with ambiguous words in Subsequent Disambiguation sentences can bias the later interpretation of those words toward the same subordinate meaning, at least in some sentence exposure situations. While Experiments 1 and 2 showed a reduction in priming for words in Subsequent Disambiguation sentences compared to those in Prior Disambiguation sentences, Experiment 2 also showed that priming in the Subsequent Disambiguation condition can occur when the prime sentence is followed by a semantically-related probe word. Furthermore, Experiment 3 showed that word-meaning priming from Subsequent Disambiguation sentences can be reliable when the sentence exposure task changes to one with a second continuation sentence and without the need for speeded decision making. We interpret this pattern of results as suggesting that word-meaning priming occurs when the listener is given enough time to complete the reinterpretation process that is triggered by the disambiguating context, and when there is no decision-making task that interferes with this process. In other situations, such as those that allow for more shallow or gist-like processing, there appears to be a priming advantage for words presented in Prior Disambiguation sentences, probably because the words in these sentences are more likely to be correctly disambiguated during initial processing. In the present experiment we sought to replicate the results of Experiment 3 with the meaning access task used in the Experiment 2 test phase.

### Method

#### Participants

One hundred and eighty native British English speakers completed the study. The participant inclusion criteria, recruitment method, and payment rates were the same as in Experiments 2 and 3.

#### Design, materials and procedure

The materials and design were the same as in Experiment 3. The

target-probe word pairs used in the meaning access task were the same as those used in Experiment 2 (see Tables S8 and S9 in the Supplementary Material). See Table 4 for example stimuli.

The procedure was the same as that used in Experiment 3 (see Fig. 6), except that the meaning preference task was replaced with the meaning access task that was used in Experiment 2 (see Fig. 3C). After the self-paced listening/reading sentence exposure task, participants completed the Mill Hill vocabulary test, followed by the meaning access task. The meaning access task began with brief instructions and practice trials before the main task began (see Experiment 2 method). The average delay between ambiguous word presentations in the prime and test phases was approximately 25 min (range = 18–36 min). As in Experiment 3, there was no disambiguation post-test.

Results

Performance on the comprehension questions in the sentence exposure task was good (mean proportion correct = 0.95, SD = 0.06, range = 0.64 – 1.00), showing that participants were attending to the sentence pairs. The average proportion correct on the Mill Hill vocabulary test was 0.60 (SD = 0.13, range = 0.27–0.88).

As in Experiment 2, responses that were faster than 300 ms or slower than 2500 ms were assumed to reflect accidental button presses or lapses in attention, respectively, and removed from analysis. Of the 11,700 total responses from the meaning access task, 32 trials were excluded

Table 4

Experiment 4 stimuli for a single item (bank) and all six Prime Type × Continuation Sentence Type conditions. For each participant, each experimental item was assigned to only one of the six conditions. Ambiguous words are underlined and disambiguation words are in italics in the prime and continuation sentences. In the meaning access task, experimental target words were always paired with a probe word related to the primed subordinate meaning. The Unprimed-Biased and Unprimed-Neutral conditions were equivalent, but were categorised separately to create a balanced factorial design.

Prime type	Continuation sentence type	Sentence exposure task	Meaning access task	
		Prime sentence (auditory) and continuation sentence (visual)	Target word (auditory)	Probe word (visual)
Prior	Biased	The old man had a long way to <i>swim</i> as he headed for the <u>bank</u> . After reaching the <i>shore</i> , he rested briefly and then <i>swam</i> back.	bank	RIVER
Prior	Neutral	The old man had a long way to <i>swim</i> as he headed for the <u>bank</u> . He thought about going back but then decided to continue.	bank	RIVER
Subsequent	Biased	The old man headed for the <u>bank</u> but he had a long way to <i>swim</i> . After reaching the <i>shore</i> , he rested briefly and then <i>swam</i> back.	bank	RIVER
Subsequent	Neutral	The old man headed for the <u>bank</u> but he had a long way to <i>swim</i> . He thought about going back but then decided to continue.	bank	RIVER
Unprimed	Biased	–	bank	RIVER
Unprimed	Neutral	–	bank	RIVER

based on the RT thresholds. Of the remaining trials, 2,856 responses (25%) were incorrect and removed, leaving 8,812 trials in the RT analysis. As in Experiment 2, the relatively high error rate in this task was expected given the use of subordinate meanings in a speeded task, and item mean error rates correlated with the dominance of the subordinate meanings,  $r(63) = .47, p < .001$ .

Model diagnostic plots revealed that log transformation of RTs best met the assumption of normally-distributed residuals. The model with the full random effects structure did not converge, so random effect terms were removed one at a time, starting with the term that accounted for the least variance, until the model converged. This procedure resulted in the removal of by-item slopes for Prime Type and Continuation Sentence Type, and the by-subject slope for Continuation Sentence Type. Thus the final model contained a by-item intercept, a by-subject intercept, and by-subject slopes for the Prime Type and the interaction between the Prime Type and Continuation Sentence Type.

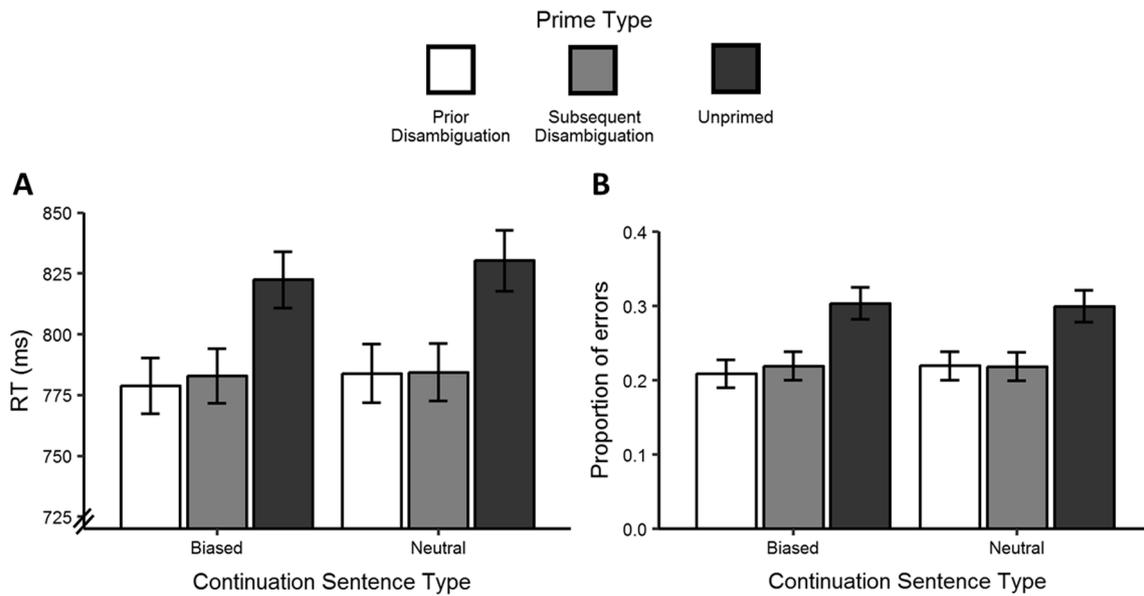
Model comparisons revealed a main effect of Prime Type on RTs,  $\chi^2(4) = 86.54, p < .001$ ; see Fig. 8A. The model summary showed that the coefficient for the contrast between the primed vs. unprimed conditions was significant,  $\beta = 0.06, SE = 0.01, t(178.4) = 10.63, p < .001$ . The coefficient for the contrast between the two prime sentence types was not significant,  $\beta = 0.01, SE = 0.01, t(210.5) = 0.79, p = .431$ . Pairwise comparisons with Holm adjustment for multiple comparisons confirmed that RTs for words in both the Prior and Subsequent sentence conditions were significantly faster than Unprimed ( $\chi^2(1) = 92.79, p < .001$  and  $\chi^2(1) = 82.24, p < .001$ , respectively). The model comparisons showed no significant effect of Continuation Sentence Type,  $\chi^2(3) = 2.23, p = .527$ , or interaction between Prime Type and Continuation Sentence Type,  $\chi^2(2) = 0.76, p = .685$ .

In the accuracy analysis, there were 11,668 trials included after removing the 32 responses that were faster than 300 ms or slower than 2500 ms. The subject mean error rates by condition are shown in Fig. 8B. Model comparisons showed a significant main effect of Prime Type,  $\chi^2(4) = 49.61, p < .001$ . There were significantly more errors in the Unprimed condition relative to the two primed conditions,  $\beta = -0.62, SE = 0.07, z = -8.91, p < .001$ , while errors in the Prior and Subsequent prime conditions did not significantly differ,  $\beta = -0.07, SE = 0.08, z = -0.86, p = .388$ . This pattern of results was reflected in the pairwise comparisons, with Holm adjustment for multiple comparisons, which showed significantly more errors in the Unprimed condition relative to both the Prior ( $\chi^2(1) = 63.18, p < .001$ ) and Subsequent ( $\chi^2(1) = 57.41, p < .001$ ) prime conditions. The likelihood ratio tests from model comparisons revealed no significant main effect of Continuation Sentence Type,  $\chi^2(3) = 0.40, p = .940$ , and no significant interaction,  $\chi^2(2) = 0.08, p = .960$ .

Discussion

The results of Experiment 4 are clear in replicating the priming effects for Prior Disambiguation sentences (Experiments 1–3) and Subsequent Disambiguation sentences (Experiments 2–3), as well as the null effect of the Continuation Sentence Type from Experiment 3. As in Experiment 3, we found significant priming from Subsequent Disambiguation sentences when participants were exposed to the sentences via a self-paced listening/reading task. This suggests that priming from Subsequent Disambiguation sentences can occur as long as the listener has sufficient time to reanalyse the sentence and correctly disambiguate the ambiguous word, without interruption from unrelated sentence probes in the context of a speeded decision task.

Here we again found no evidence of an effect of additional disambiguating context after the prime sentence; there was no reliable difference between Biased versus Neutral Continuation Sentences. Null effects are difficult to interpret, but we can have greater confidence given that this is essentially the same result as found in Experiment 3. The lack of a clear influence of the Continuation Sentences suggests that semantically-related sentence probe words were not critical in



**Fig. 8.** Experiment 4 response times (A) and proportions of errors (B) from relatedness decisions to ambiguous and subordinate-related word pairs (e.g. bank – RIVER) in the meaning access task. Target ambiguous words were primed in a Prior Disambiguation (white) or Subsequent Disambiguation (light gray) sentence, or they were Unprimed (dark gray). The prime sentences were followed by either Biased (left bars) or Neutral (right bars) continuation sentences, which did or did not provide additional cues to disambiguation, respectively. The two Continuation Sentence Type conditions were equivalent in the Unprimed condition. Bars show the participant grand means, and error bars show 95% confidence intervals, adjusted to remove between-subject variance (Morey, 2008).

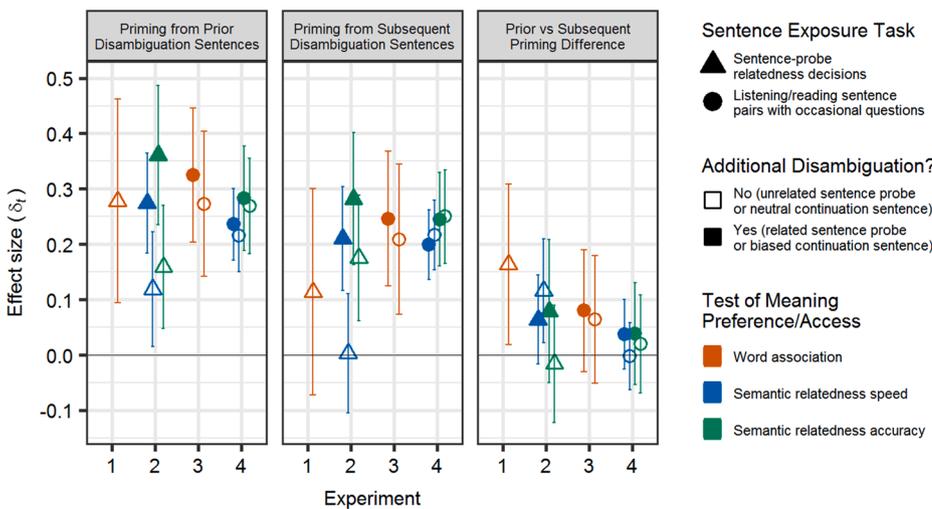
producing the effect of Sentence Probe Type in Experiment 2. That is, it does not seem to be the case that additional disambiguation cues that follow the prime sentence (either from the semantically-related probe words in Experiment 2, or from the Biased sentence condition in Experiments 3 and 4) are crucial for determining the presence or magnitude of priming effects.

Consistent with Experiment 3, the results of Experiment 4 showed that when Subsequent Disambiguation prime sentences are followed by a continuation sentence in a self-paced task, we see priming effects that are statistically indistinguishable from those produced by Prior Disambiguation sentences. Thus it appears that encountering an ambiguous word in a subordinate-meaning context can bias later interpretation of that word toward the same meaning regardless of the relative positioning of the ambiguous word and disambiguating context.

**General discussion**

The results of our four experiments are summarised in Fig. 9, which

shows effect sizes for priming from prior disambiguation sentences, priming from subsequent disambiguation sentences, and the difference in priming between these sentence types. In all experiments, we observed reliable word-meaning priming in all prior disambiguation conditions (Fig. 9, left panel), which is consistent with previous studies using this sentence structure (2013; Betts et al., 2018; Gaskell et al., 2019; Gilbert et al., 2018; Rodd et al., 2016). Priming from subsequent disambiguation sentences was less consistent (Fig. 9, middle panel). We found some limited evidence of greater priming from prior versus subsequent disambiguation sentences, with the effect of this contrast diminishing across the series of experiments (Fig. 9, right panel). Experiment 2 showed that the type of probe word presented after the prime sentence significantly modulated priming. For both prior and subsequent disambiguation sentences, there was greater priming when the sentence was followed by a semantically-related vs unrelated sentence probe word (Fig. 9, Experiment 2, left and middle panels, filled vs unfilled triangles). This raised the possibility that the absence of priming for subsequent disambiguation sentences (in Experiment 1, and for one



**Fig. 9.** Effect sizes ( $\delta_t$ ) across the four experiments. The  $\delta_t$  standardised effect size is a Cohen’s-*d*-like value for mixed-effects models that was calculated by dividing the unstandardized effect by the sum of all variances (from random effects and residual; Brysbaert & Stevens, 2018; Hedges, 2007; Westfall, Kenny, & Judd, 2014). Unstandardized mean differences were calculated using the “emmeans” package for R (version 1.3.2; Lenth, 2019). The left and middle panels show priming effects with prior and subsequent disambiguation sentences, respectively, where positive values indicate mean differences versus unprimed conditions in the expected direction. The right panel shows the effect of prior versus subsequent priming, where positive values indicate more priming for prior sentences, negative values indicate more priming for subsequent sentences, and 0 indicates no effect of disambiguation context position in the prime sentence. Error bars show 95% confidence intervals.

condition of Experiment 2) occurred when the sentence exposure task did not mandate the formation of a detailed representation of the sentence meaning (i.e. when performing a semantic relatedness judgement on an unrelated probe word). In Experiments 3 and 4, we changed to a self-paced listening/reading paradigm to encourage more natural processing of the prime sentences. The continuation sentences were either biased toward the primed subordinate meaning or were neutral, allowing us to assess whether additional disambiguating information influences priming. In both of these experiments, we found reliable priming from subsequent disambiguation sentences, regardless of the presence/absence of additional disambiguating context (Fig. 9, middle panel, Experiments 3 and 4), and priming was of a similar magnitude for prior and subsequent disambiguation sentences (Fig. 9, right panel, Experiments 3 and 4).

#### *Effects of reinterpretation and listening task on lexical-semantic retuning*

The aim of this study was to determine whether lexical-semantic retuning in response to recent experience with ambiguous words is affected by the need to revise an initial misinterpretation of the word. We investigated whether word-meaning priming can occur not only after exposure to prior disambiguation sentences, in which the context-appropriate subordinate meaning can be selected immediately, but also after exposure to subsequent disambiguation sentences, in which the inappropriate dominant meaning is likely to be selected first before later revision. We predicted that there may be greater priming for either of these two sentence types, on the basis of the two possible learning mechanisms described in the introduction.

While the results of Experiments 1 and 2 showed a priming advantage for words in prior disambiguation sentences, the results of Experiment 2 also suggested a possible task effect. This led us to ask an additional question in Experiments 3 and 4: is the magnitude of lexical-semantic retuning affected by additional biasing context that follows after the context-appropriate meaning has been selected during the prime encounter? As summarised above, we found that (1) there is a priming advantage for prior disambiguation sentences when the sentence exposure task encourages faster and/or shallower processing, but (2) word-meaning priming can occur from both sentence types, and (3) when the sentence exposure task encourages slower and/or deeper processing of sentence meaning, there is equivalent word-meaning priming from both prior and subsequent disambiguation sentences.

Based on these findings, we can rule out the hypothesis that word-meaning priming can *only* occur when there is immediate activation and selection of the contextually-appropriate subordinate meaning when listeners encounter an ambiguous word. Such an account would predict the absence of (or a substantial reduction in) learning from subsequent disambiguation sentences, which was not always observed. We can also rule out the hypothesis that word-meaning priming is driven primarily by error detection or effortful reinterpretation processes. This account predicts greater priming from subsequent than prior disambiguation sentences, which was never observed.

Instead, our findings suggest that word-meaning priming can operate regardless of the initial meaning activation and need for reinterpretation during the prime encounter, as long as the listening situation is likely to result in correct (re)interpretation of the ambiguous word. The combination of sentences that require more effort to understand and a situation that does not mandate a precise representation of the sentence meaning can be sufficient to reduce or prevent the updating of word-meaning representations, even when the listener is attending to the prime sentences.

It is perhaps remarkable that a task requiring a decision about the *semantic* content of each sentence (the semantic relatedness task used in sentence exposure phases for Experiments 1 and 2) can result in shallower sentence processing compared to a more passive listening/reading task in which participants merely attend to sentences and answer occasional comprehension questions (as in Experiments 3 and 4).

However, this result becomes less surprising in the light of the ‘good enough’ view of language processing (Christianson, 2016; Ferreira & Lowder, 2016; Ferreira & Patson, 2007; Ferreira et al., 2002; Karimi & Ferreira, 2016). According to this view, linguistic representations computed during sentence comprehension are only as precise as necessary to perform a given task, and can remain in an underspecified or ‘gist’ state until further elaboration of these representations becomes necessary. The impact of sentence processing depth is most apparent when sentences are more complex (e.g. greater distance between the ambiguous and disambiguation words in garden-path sentences; Christianson, Hollingworth, Halliwell, & Ferreira, 2001) and/or when the content deviates from expectations, such as an unexpected thematic role assignment (e.g. “The dog was bitten by the man.”; Ferreira, 2003) or, as in the present study, when a single word rules out the high frequency meaning of an ambiguous word (e.g. “Sally worried that the ball would be too crowded.”).

Although testing this view was not an initial aim of our study, the task effects that we saw are consistent with ‘good enough’ theories of language processing. Word-meaning priming seems to operate on the output of good-enough processing: for lower frequency meanings of ambiguous words, a subsequent context may not be enough to easily repair an initial misinterpretation, whereas for prior disambiguation sentences, the ambiguity is more easily resolved and priming occurs more reliably. This is supported by brain imaging studies showing differences in fMRI bold responses and event-related potentials for subsequent disambiguation sentences that resolve to a subordinate meaning, which presumably reflects the detection of misinterpretation and/or effortful repair process (Macgregor et al., 2020; Rodd et al., 2012). More generally, the effect of the probe task on sentence processing underscores the importance of conceptual replications for investigating the phenomenon of interest separate from any potential task-specific effects.

It is interesting to note that there was no suggestion of a ‘good enough’ processing effect in Experiments 3 and 4, despite the inclusion of a neutral continuation sentence condition that did not mandate correct disambiguation in order to be integrated with the first prime sentence. In this condition, participants could have initially formed shallow, ‘gist’-like representations of the sentence pairs until presented with a (relatively rare) comprehension question at the end of the trial, which would then signal the need for further elaboration of the sentence meanings. However, in Experiments 3 and 4, we found no evidence of greater priming from prior versus subsequent disambiguation sentences despite comprehension questions never being presented after the experimental prime sentences. So why did we not find evidence of ‘good enough’ processing in this task? This may be due to the additional cost of holding more information in memory (two sentences, rather than one sentence in Experiments 1 and 2) for the purpose of potentially needing to answer a comprehension question, versus the benefit of avoiding deep/detailed sentence processing when it is not required by the situation. The most efficient strategy for participants in Experiments 3 and 4 might have been to form a complete and specific interpretation of the first sentence. This would minimise memory demands and facilitate integration with the second sentence, irrespective of whether the continuation sentence contained additional subordinate-meaning context that would help with disambiguation.

#### *Implications for learning mechanisms during speech comprehension*

What do these findings tell us about the mechanisms that underpin word-meaning priming? Based on previous work showing equivalent cross-modal (speech-to-reading and vice-versa) and uni-modal (speech-to-speech, reading-to-reading) word-meaning priming, we have proposed that word-meaning priming occurs via changes to connection weights within the lexical-semantic system (Gilbert et al., 2018). By this account, word-meaning priming does not change the mapping from form to meaning (which would be modality-specific), but rather changes long-term lexical-semantic connections such that, after priming, the

semantic system becomes more biased toward settling into the primed meaning of an ambiguous word (Rodd, 2020). This effect is also compatible with the alternative view that meaning representations are not directly activated by word forms, but rather are mediated by an intermediate ‘lemma’ level, where each lemma is associated with one meaning of an ambiguous word (Taft & Nguyen-Hoan, 2010). Under this view, one particular word-meaning lemma would become more readily activated following the prime exposure. But regardless of this aspect of the lexical-semantic structure, the critical point for our purposes is that word-meaning priming is directly affecting long-term amodal word-meaning representations.

Our results have ruled out two plausible mechanisms by which this ‘direct alteration’ account might operate, thereby substantially narrowing the space of viable options. For the direct alteration account to explain word-meaning priming for subsequent disambiguation sentences requires that changes to stored lexical-semantic knowledge do not necessarily occur immediately after the ambiguous word has been encountered, but can instead be delayed until the comprehender is more certain of the intended meaning. It is only at or after the end of a sentence like “Sally worried that the ball would be too *crowded*” that listeners could know that the “party” meaning of “ball” was intended, and consequently, the meaning that is primed for the word “ball”. This learning process therefore requires sufficient short-term memory for listeners to retain the ambiguous word until the end of a sentence when correct interpretation is possible and updating occurs. Since the ambiguous word itself must be used in the prime sentence in order for long-term word-meaning priming to occur (Rodd et al., 2013, Experiment 3), this implies that the ambiguous word representation is a critical part of this learning process and it must remain accessible during ongoing language processing in order for the lexical-semantic connections for that particular word to be adjusted in favour of the final interpretation.

By this account, and given the role for short-term memory of some kind, it is entirely feasible that factors such as the relative position of the disambiguating context or the presence/absence of a semantically-related sentence probe word can influence word-meaning priming. These factors could change the nature and timing of the training signal that is responsible for altering lexical-semantic connection weights. Nonetheless, since learning from subsequent disambiguation sentences is possible, it follows that short-term memory in this account must generally be sufficient to retain the critical ambiguous word until the end of a sentence when correct disambiguation can take place.

In this respect, word-meaning priming shows a striking difference from perceptual learning of degraded or ambiguous speech sounds. As mentioned in the introduction, for speech-sound learning, the relative timing of the perceptually-ambiguous segment(s) and disambiguating information has been shown to be critical. For example, listeners’ comprehension of distorted (e.g. noise vocoded) speech is enhanced across trials when, after each sound exposure and word report, the listener is given feedback in which disambiguating information (e.g. matching written word or clear speech) is presented *before* a repetition of the same distorted speech sound. However, perceptual learning across trials is reduced or absent if, during this post-trial feedback, the disambiguating information is only presented *after* the repeated distorted speech sound (Davis et al., 2005; Hervais-Adelman et al., 2008). A similar advantage for prior versus subsequent disambiguation has been shown for learning ambiguous speech sounds; ambiguous /s/-/f/ sounds at the end of words like *carcass* or *giraffe* produce reliable perceptual learning many minutes later (Jesse & McQueen, 2011). In other words, prior lexical context that constrains the interpretation of an ambiguous sound leads to a long-term change in later interpretations of the same sound. However, long-term perceptual learning is absent if the disambiguating lexical context only follows after a /s/-/f/ ambiguity at word onset, as in words like *syrup* or *phantom*. These results indicate that, for ambiguous speech sounds, the disambiguating context must occur before the ambiguity in order to affect future interpretation. This is in

contrast with our results showing that learning about ambiguous word meanings is relatively insensitive to the position of the ambiguity and disambiguating context. The apparent inconsistency in these results suggests that there may be a critical difference in the underlying mechanisms that support these forms of learning.

We suggest that the dependence of perceptual learning on prior disambiguating information can be explained as due to a limitation of acoustic or echoic short-term memory (Crowder & Morton, 1969). Auditory echoic memory has a very limited capacity – perhaps as short as a 200 ms. Listeners therefore cannot retain the acoustic details of degraded or ambiguous speech sounds long enough for subsequent disambiguation information to become available (see Sohoglu, Peelle, Carlyon, & Davis, 2014, for discussion). By this view, then, our observation of reliable word-meaning priming for subsequent disambiguation sentences may be a reflection of the longevity of the phonological short-term memory system that supports word recognition and sentence comprehension. We might therefore predict that listeners with impaired phonological short-term memory would also be impaired at word-meaning priming when there is a long delay between the ambiguity and subsequent disambiguation.

An alternative account of word-meaning priming, proposed by Gaskell et al. (2019), does not involve the direct alteration of lexical-semantic connection weights, but rather depends on a separate system that facilitates long-term sentence memory. Gaskell et al. propose that the comprehension of an ambiguous word in a disambiguating sentence context generates a new hippocampally-based memory representation that binds, at some level, the words in each sentence to a specific meaning context. This more episodic representation could then be consolidated during sleep, with systems consolidation gradually updating the long-term lexical-semantic weights as described in our first account above. This alternative might relate the ambiguous word form to the outcome of the comprehension process at a situation model level (Kintsch, 1988). As such, any factors that affect the outcome of a good-enough comprehension process (e.g., disambiguation context position, sentence exposure task) would equally affect the newly-bound memory and its ability to influence meaning retrieval processes for the ambiguous word many minutes or hours later. Nothing in the current results can discriminate between this account and an account in which priming directly alters semantic attractor weights with support from phonological short-term memory. However, future research could investigate word-meaning priming in patients with impaired episodic memory (e.g. due to hippocampal damage or neurodegeneration) in order to tease apart the relative contributions of these two proposed mechanisms. Indeed it is interesting to note that patients with hippocampal amnesia tend to have impoverished lexical-semantic representations of words, and are specifically disadvantaged at retrieval of multiple senses of lexically ambiguous words, suggesting that the hippocampus may be involved in the updating of meanings for these words (Klooster & Duff, 2015).

This discussion of different mechanisms underpinning word-meaning priming emphasizes the important but often overlooked role of dynamic learning processes in skilled comprehension. While similar issues have been addressed in syntactic and perceptual learning, the word-meaning priming effect further shows that adults’ lexical-semantic knowledge is not static, but rather is continuously updated in response to experience. Thus, lexical-semantic learning must be incorporated into models of language comprehension. Our results rule out two candidate mechanisms for the way in which this type of learning relates to sentence comprehension, and yet at least two neurocognitive mechanisms remain viable for how lexical-semantic knowledge is updated. Only by understanding precisely when and how language exposure and comprehension leads to long-term learning about words and their meanings can we achieve a satisfactory account of language processing.

## Author contributions

RAG contributed to the experimental design, carried out the experiments, analysed the data, and wrote the manuscript. JMR acquired the funding, led the conception and design of experiments, supervised the work, and wrote the manuscript. MHD and MGG contributed to the experimental design and analysis, and wrote the manuscript.

## Declaration of Competing Interest

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

## Acknowledgments

This work was supported by an Economic and Social Research Council grant [ES/K013351/1] awarded to Jennifer M. Rodd, and UK Medical Research Council funding [SUAG/044 G101400] awarded to Matthew H. Davis. The authors would like to thank Jane Warren for the stimuli pre-testing, and Elinor Lighter for her help with Experiment 1 data collection.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jml.2020.104188>.

## References

- Baayen, R. H., & Milin, P. (2015). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. <https://doi.org/10.21500/20112084.807>.
- 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). <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, D. M., 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>. Computation.
- 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). <https://doi.org/10.1037/xlm0000507>.
- Blott, L. M., Rodd, J. M., Ferreira, F., & Warren, J. (2019). Semantic ambiguity resolution during sentence comprehension is more efficient in individuals with greater lexical expertise. *PsyArXiv*. <https://doi.org/10.31234/OSF.IO/3EJQY>.
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129(2), 177–192. <https://doi.org/10.1037/0096-3445.129.2.177>.
- Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, 55, 22–31. <https://doi.org/10.1016/j.learninstruc.2018.01.013>.
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1). <https://doi.org/10.5334/joc.10>.
- Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29(2), 217–230. <https://doi.org/10.1023/A:1005101313330>.
- Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 817–828. <https://doi.org/10.1080/17470218.2015.1134603>.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. Retrieved from *Cognitive Psychology*, 42(4), 368–407 <http://linkinghub.elsevier.com/retrieve/pii/S0010028501907522>.
- Christianson, K., & Luke, S. G. (2011). Context Strengthens Initial Misinterpretations of Text. *Scientific Studies of Reading*, 15(2), 136–166. <https://doi.org/10.1080/1088431003636787>.
- Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 5(6), 365–373. <https://doi.org/10.3758/BF03210660>.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGgettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2), 222–241. <https://doi.org/10.1037/0096-3445.134.2.222>.
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, 48(1), 1–12. <https://doi.org/10.3758/s13428-015-0567-2>.
- Duffy, S. A., Kambe, G., & Rayner, K. (2001). The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 27–43). Washington, DC, US: American Psychological Association.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27(4), 429–446. [https://doi.org/10.1016/0749-596X\(88\)90066-6](https://doi.org/10.1016/0749-596X(88)90066-6).
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47(2), 164–203. [https://doi.org/10.1016/S0010-0285\(03\)00005-7](https://doi.org/10.1016/S0010-0285(03)00005-7).
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11–15. <https://doi.org/10.1111/1467-8721.00158>.
- Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation*, 65, 217–247. <https://doi.org/10.1016/BS.PLM.2016.04.002>.
- 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>.
- Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 37(3), 578–591. <https://doi.org/10.1111/cogs.12022>.
- Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition*, 182, 109–126. <https://doi.org/10.1016/j.cognition.2018.09.007>.
- Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and readers generalize their experience with word meanings across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(10). <https://doi.org/10.1037/xlm0000532>.
- Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, 94, 149–165. <https://doi.org/10.1016/j.jml.2016.11.001>.
- 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>.
- Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341–370. <https://www.jstor.org/stable/20172092>.
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20(11), 1315–1326. <https://doi.org/10.1002/hipo.20857>.
- Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 460–474. <https://doi.org/10.1037/0096-1523.34.2.460>.
- Hilbig, B. E. (2015). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-015-0678-9>.
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition*, 127(1), 57–83. <https://doi.org/10.1016/j.cognition.2012.10.013>.
- Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, 18(5), 943–950. <https://doi.org/10.3758/s13423-011-0129-2>.
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013–1040. <https://doi.org/10.1080/17470218.2015.1053951>.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>.
- Klooster, N. B., & Duff, M. C. (2015). Remote semantic memory is impoverished in hippocampal amnesia. *Neuropsychologia*, 79, 42–52. <https://doi.org/10.1016/j.neuropsychologia.2015.10.017>.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. Retrieved from <https://cran.r-project.org/package=emmeans>.
- Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, 106(50), 21086–21090. <https://doi.org/10.1073/pnas.0907664106>.
- Macgregor, L. J., Rodd, J. M., Gilbert, R. A., Hauk, O., Sohoglu, E., & Davis, M. H. (2020). The neural time course of semantic ambiguity resolution in speech comprehension. *Journal of Cognitive Neuroscience*, 32(3), 403–425. [https://doi.org/10.1162/jocn\\_a.01493](https://doi.org/10.1162/jocn_a.01493).
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, 4(2), 61–64.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9).

- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225–236. <https://doi.org/10.3758/BF03196957>.
- Palan, S., & Schitter, C. (2017). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*. <https://doi.org/10.1016/J.JBEF.2017.12.004>.
- Peer, E., Samat, S., Brandimarte, L., & Acquisti, A. (2015). Beyond the Turk: An empirical comparison of alternative platforms for online behavioral research. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2594183>.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raven, J., Raven, J. C., & Court, J. H. (1998). The Mill Hill Vocabulary Scale. In Manual for Raven's progressive matrices and vocabulary scales. San Antonio, TX: Harcourt Assessment.
- Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods*, 47(2), 309–327. <https://doi.org/10.3758/s13428-014-0471-1>.
- Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science*, 15(2), 411–427. <https://doi.org/10.1177/1745691619885860>.
- Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, 87, 16–37. <https://doi.org/10.1016/j.jml.2015.10.006>.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. (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., Johnsruide, 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., Johnsruide, I. S., & Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex*, 22(8), 1761–1773. <https://doi.org/10.1093/cercor/bhr252>.
- 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>.
- Rosario-Martinez, H. De. (2015). phia: Post-Hoc Interaction Analysis. Retrieved from <https://cran.r-project.org/package=phia>.
- Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5), 781–794. <https://doi.org/10.1037/xlm0000341>.
- Schubert, T. W., Murteira, C., Collins, E. C., & Lopes, D. (2013). ScriptingRT: A Software Library for Collecting Response Latencies in Online Studies of Cognition. *PLoS One*, 8(6), Article e67769. <https://doi.org/10.1371/journal.pone.0067769>.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14(4), 489–537. [https://doi.org/10.1016/0010-0285\(82\)90017-2](https://doi.org/10.1016/0010-0285(82)90017-2).
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104–120. <http://linkinghub.elsevier.com/retrieve/pii/S0749596X13000247>.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 186–199. <https://doi.org/10.1037/a0033206>.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645–659. [https://doi.org/10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4).
- Taft, M., & Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes*, 25(2), 277–296. <https://doi.org/10.1080/01690960903403261>.
- van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, 55(3), 335–362. <https://doi.org/10.1016/J.JML.2006.06.004>.
- Vitello, S., & Rodd, J. M. (2015). Resolving semantic ambiguities in sentences: Cognitive processes and brain mechanisms. *Language and Linguistics Compass*, 9(10), 391–405. <https://doi.org/10.1111/Inc3.12160>.
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>.
- Wiley, J., George, T., & Rayner, K. (2016). Baseball fans don't like lumpy batters: Influence of domain knowledge on the access of subordinate meanings. *The Quarterly Journal of Experimental Psychology*, 1–11. <https://doi.org/10.1080/17470218.2016.1251470>.
- Wonnacott, E., Joseph, H. S. S. L., Adelman, J. S., & Nation, K. (2016). Is children's reading "good enough"? Links between online processing and comprehension as children read syntactically ambiguous sentences. *Quarterly Journal of Experimental Psychology*, 69(5), 855–879. <https://doi.org/10.1080/17470218.2015.1011176>.
- Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: A tutorial review. *PeerJ*, 3, Article e1058. <https://doi.org/10.7717/peerj.1058>.