Retuning lexical-semantic representations on the basis of recent experience

Hannah N. Betts

2018

Experimental Psychology
University College London

A thesis submitted for the degree of Doctor of Philosophy

Abstract

Classic studies on ambiguous words (e.g. '*bark*' – dog/tree) imply that comprehenders' lexical-semantic representations remain relatively stable across time. However, recent research has shown that a single encounter with a particular word-meaning biases interpretation up to 20 minutes later ("word-meaning priming"), suggesting that representations update to reflect recent experience. Nine experiments in this thesis investigate in detail the effects of recent experience on the comprehension of ambiguous words.

Using word association, Chapter 2 replicates the single-encounter subordinate priming effect and shows that this effect is reduced by a subsequent dominant meaning encounter. Three recent subordinate encounters boost priming compared to a single encounter but only when encounters are temporally spaced; massed encounters seem to provide no such boost. Chapter 3 assesses a newly-developed semantic relatedness test of word-meaning availability effects on comprehension, using picture probes. It shows that, compared to word association, semantic relatedness can detect dominance with the additional benefit of testing dominant and subordinate meaning availabilities independently. Chapter 4 shows that this semantic relatedness test can detect single-encounter word-meaning priming and that this effect is driven by increased availability of the primed meaning, not decreased availability of the unprimed meaning. Furthermore, an additional priming boost from three repetitions reflects an increase in primed meaning availability for both massed and spaced repetitions, with an additional *decrease* in *unprimed* meaning availability after spaced repetitions only; there was no evidence that massed repetitions reduced unprimed availability.

Possible mechanisms are discussed that account for these different repetition priming patterns observed with semantic relatedness and word association tests. The findings suggest that the word-meaning priming effect might be driven by episodic memory and consolidated lexical-semantic representations. Taken together, these experiments confirm that recent experience plays a key role in retuning lexical-semantic representations and can help to refine our theoretical accounts of this important phenomenon.

Impact Statement

The majority of English words have multiple possible interpretations. This means that difficulties to understand ambiguous words can be detrimental to comprehension. However, current accounts of semantic ambiguity resolution are, at best, incomplete and, at worst, incorrect. Until we fully understand efficient comprehension in healthy adults, we cannot begin to provide interventions for those challenged by ambiguity. The present research provides key insights into the learning mechanism(s) that improve the ability of healthy adult listeners to understand ambiguous words efficiently. In doing so, this research provides more of the necessary evidence-base for future research that will assess the precise nature of the comprehension difficulties for particular groups and individuals. This will facilitate the development and evaluation of interventions aimed at improving comprehension skills.

Declaration

I, Hannah N. Betts, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Statement of contributions

Whilst data for Experiment 1 were collected for my previous MSc degree, the re-analysis of its data with mixed effects modelling was conducted as part of this PhD and is therefore included in this thesis. Five UCL students kindly contributed to Chapters 2 and 3: the word association data for Experiment 2 were coded by Lauren Farrar, a research assistant in the lab; the data for Experiment 3 were collected by Zainab Okedara, an undergraduate student in the lab; the data in Chapter 3 were partially collected by three other undergraduate students. A modified version of Chapter 2 was published in JEP:LMC prior to the completion of this thesis.

Acknowledgments

I am incredibly grateful to the Economic and Social Research Council for their funding of the studentship covering my MSc, MPhil and PhD, and to UCL for affording this research. I would especially like to thank my parents, Susie and Phil, my brother, Dan (and of course Marsha!), and my boyfriend, James, for their continued support during the rollercoaster that is a PhD; I couldn't have done it without their patient listening and encouragement. Also, my lovely Nana Hilda, who was so proud of me that she told everyone she met that her granddaughter was going to be a Doctor. I wish that she and Grandad Dennis could have seen me complete it. I would like to thank my other family members, my friends and my lab colleagues, but particularly Eva, Rachael and Becky, for making such an enjoyable working environment and for keeping me sane during the data analysis phases. I also thank Jane and Matt for an interesting discussion during my viva, and of course Jenni for the numerous sketched diagrams and her invaluable supervision.

Contents

List of Figures

List of Tables

Chapter 1: General introduction

The importance of studying semantic ambiguity resolution

Communication is a vital aspect of human life; the ability to understand language is therefore invaluable. However, language is universally ambiguous and, irrespective of whether this is an adaptive element, or an inconvenient by-product, of language development (for a range of arguments in psycholinguistics and philosophy, see: Chomsky, 2002; Piantadosi, Tily, & Gibson, 2012; Wittgenstein, 1953; Zipf, 1949), ambiguity does complicate the comprehension process. In particular, some words have more than one possible interpretation. In English, over 80% of words have multiple meanings, and are therefore ambiguous (e.g. 'bark': the noise made by a dog, or the covering of a tree; Rodd, Gaskell, & Marslen-Wilson, 2002). This not only demonstrates that ambiguous words occur in the English language, but also that they are in fact an integral part of it.

Ambiguous words require additional processing compared to unambiguous words because they are open to multiple possible interpretations, and misinterpretation of these words can be costly (e.g. Christianson, 2016). For instance, an instruction to 'bring a *mac*' could either indicate that a trench coat would be useful due to rain, or that a particular brand of computer is required. Here, misinterpretation of 'mac' would be inconvenient at best. It follows, then, that difficulties to understand ambiguous words would be detrimental to comprehension, and these difficulties have been found in a range of clinical and non-clinical populations (e.g. Gernsbacher, Varner, & Faust, 1990; Norbury, 2005).

A considerable amount of research has focused on the comprehension of ambiguous words, providing evidence and models primarily for the representations of, and processes involved in, disambiguating ambiguous words (e.g. McClelland & Rumelhart, 1981; Swinney, 1979). More recently, research has turned to investigate the learning mechanism(s) underlying semantic ambiguity to reveal more about comprehension (Gilbert, Davis, Gaskell, & Rodd, 2018; Rodd, Cai, Betts, Hanby, Hutchinson, & Adler, 2016; Rodd, Lopez Cutrin, Kirsch, Millar, & Davis, 2013). The existing literature does not, however, provide the full picture. For instance, it is not yet clear exactly how, and in which circumstances, recent experience can inform subsequent comprehension. Without a complete picture of the representations, processes and learning mechanisms associated with ambiguous word comprehension, we cannot begin to build interventions to facilitate understanding in those challenged by ambiguity.

The present research provides key insights into the learning mechanism(s) that improve the ability of healthy adult listeners to interpret ambiguous words in a fluent and efficient manner. In doing so, this research adds to the necessary evidence-base for future research into the precise nature of the comprehension difficulties for particular groups and individuals. The present experiments will focus on learning mechanisms to investigate how the availabilities of ambiguous word-meaning representations adapt on the basis of recent experience to maintain processing efficiency.

Research into different aspects of semantic ambiguity resolution

As mentioned earlier, there are three elements of semantic ambiguity that can be examined: the representations of word-meanings, the processes involved in their understanding, and the learning mechanisms that allow representations and processes to be adapted over time. Here, the relevant existing literature on these three elements is reviewed.

Representations

A considerable amount of research has investigated how ambiguous words are represented in the mental lexicon. Using a lexical decision task, a range of early experiments on ambiguous word representations suggested that lexical decisions were faster for ambiguous words than for unambiguous control words (e.g. Kellas, Ferraro,

& Simpson, 1988; Rubenstein, Garfield, & Millikan, 1970). A localist approach was proposed to explain this ambiguity advantage, which was based on the assumption that one word-meaning was associated with one unit in a connectionist network (e.g. Interactive Activation and Competition Model; McClelland & Rumelhart, 1981). Under this view, it was argued that a comprehender would be able to make a lexical decision as soon as one meaning representation unit is sufficiently active; the more meaning units that exist, the more likely that *one* of them will reach the activation threshold level quickly (Jastrzembski, 1981).

However, it has been shown more recently that participants respond more slowly to words with multiple distinct meanings (e.g. 'bark') but more quickly to words with multiple related senses (e.g. 'run'), compared to unambiguous control words (Klepousniotou & Baum, 2007; Rodd et al., 2002), a distinction that had not been clarified previously. A localist model (e.g. McClelland & Rumelhart, 1981) cannot account for this finding, yet an alternative, distributed connectionist model (Rodd, Gaskell, & Marslen-Wilson, 2004) *can* accommodate the co-existing distinct word-meaning disadvantage and related word-sense advantage. This distributed connectionist model assumes that a word is not represented by a single unit, but is instead characterised by a specific pattern of activation across multiple units in the network that represent different lexical and semantic features of the word. The model suggests that participants are slower to disambiguate words with distinct meanings because the patterns of activation for each meaning are very different (i.e. they do not overlap), causing interference between meaning representation units within the semantic layer (through inhibitory connections between competing units). This slows ambiguity resolution because the model avoids settling in a blend state by forcing a shift towards an attractor basin corresponding to a particular meaning (taking more time than if only one possible pattern of activation existed for a word). Participants are faster to disambiguate words with related senses because the patterns of activation for each meaning are very similar. These overlapping patterns mean that activation of one sense is likely to facilitate activation of a second related sense through the partial activation of some of its units, making the second sense more available. This model shows that it is important to understand the representations of ambiguous word-meanings, as only by knowing how meanings are represented can we fully understand the processes that concern them.

Processes

Findings from the literature have also led to models that explain the processes behind the disambiguation of ambiguous words, either in the presence or absence of biasing context. Disambiguation involves the parallel retrieval and consideration of each of a word's possible meanings (Swinney, 1979). For instance, after encountering the ambiguous word 'pipe' in context, lexical decisions to targets related to the contextually appropriate and the contextually inappropriate meaning were both facilitated when tested immediately. However, when the target was delayed by just a second or less, only targets related to the contextually appropriate meaning were facilitated (Swinney, 1979). Additionally, the presence of an ambiguous word in a sentence increases processing time compared to an unambiguous control word (Foss, 1970; Rayner, Pacht, & Duffy, 1994, Experiment 1). Together, this evidence indicates that the comprehension of ambiguous words requires more processing than unambiguous words. It seems that a multi-stage process occurs whereby all possible word-meanings are activated and then the appropriate meaning is rapidly selected, whilst the inappropriate meaning(s) is (/are) rejected (Onifer & Swinney, 1981; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). This is a seemingly autonomous process of which the comprehender tends to be unaware (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982).

Since listeners must carry out this complex process with the majority of English words (Rodd et al., 2002), disambiguation is clearly critical to language comprehension. It has been shown that comprehenders make use of a range of cues to determine the most appropriate meaning of these semantically ambiguous words. These cues include the relative frequency with which a word-meaning occurs in the language (known as *dominance*) and the immediate sentence context in which the word is encountered.

Dominance

Although all meanings of an ambiguous word tend to be retrieved in parallel (Swinney, 1979), the dominance of a word's alternative meanings is useful in determining the most likely meaning of a word. Most ambiguous words have a higher frequency, dominant meaning (e.g. *bank* – financial institution) and one (or more) lower frequency, subordinate meaning (e.g. *bank* – riverside land). Research has shown that an interlocutor's dominant (more frequently used) meaning tends to be the default interpretation of the word unless immediate sentence context exists to steer interpretation towards a different meaning (e.g. Chen & Boland, 2008; Colbert-Getz & Cook, 2013; Foss, 1970; Rayner & Duffy, 1986; for an overview, see Vitello & Rodd, 2015). That is, when ambiguous words are encountered within a neutral context, or in the absence of context altogether, people are more likely to interpret it with its dominant meaning (Rayner & Duffy, 1986; Twilley, Dixon, Taylor, & Clark, 1994, respectively). For example, 'she sat by the *bank'* is more likely to be interpreted as the more common financial institution meaning than the less common riverside land meaning. Listeners tend to settle on the dominant meaning since it is more readily available than the subordinate meaning (Duffy, Morris, & Rayner, 1988); it tends to be encountered more often in everyday experience and is therefore more likely to be the correct interpretation (Twilley et al., 1994). The use of meaning dominance reflects an optimal strategy in word interpretation on the part of the comprehender: when there is no cue to indicate otherwise, it makes sense that the listener is likely to interpret a word with its most frequent, 'default' meaning.

Context

The presence of context can also help to rapidly select the appropriate meaning. The highly influential reordered access model (Duffy et al., 1988) indicates that strong context can serve to increase the availability of the consistent meaning, such that access to meanings can be *reordered* to make interpreting the correct word-meaning more efficient. For example, 'she sat next to the river on the grassy *bank'* strongly constrains interpretation of 'bank' towards the riverside land meaning. Here, activation of the subordinate riverside meaning is increased compared to when a neutral context is provided, confirmed by the finding of an increase in looks towards the subordinate meaning referent in these cases (using an eye tracking paradigm, Chen & Boland, 2008, Experiment 2). Access to the subordinate meaning can also be faster following subordinate context compared to a neutral context (Colbert-Getz &

Cook, 2013). Clearly, context can increase processing efficiency of ambiguous words during natural language comprehension.

The effects of context and dominance also interact to further improve the efficiency of the disambiguation process (Tabossi, 1988; Tabossi & Zardon, 1993) such that activation of the dominant meaning would be faster following dominant context than if such context had not been present. Moreover, weak dominant context biases disambiguation more than weak subordinate context (Martin, Vu, Kellas, & Metcalf, 1999). Clearly, comprehenders can take into account information from multiple cues to maximise the likelihood of correct interpretation of ambiguous words, minimising the risk of misunderstanding. Whilst these processes are largely understood and accepted, it is not clear how or whether context affects the availability of the inappropriate meaning, as well as the availability of the appropriate meaning. Without being able to account for the effects of meaning availability, models of semantic ambiguity resolution are at best incomplete, and at worst incorrect.

Learning Mechanisms

Research from different areas of psycholinguistics indicates that adults update their knowledge of language (comprehension and production) based on experience, making for a continually evolving language system. Phonetic representations alter following recent exposure to particular phonemes (Norris, McQueen, & Cutler, 2003), expectations of syntactic structures are biased by recently encountered structures (Fine & Jaeger, 2013; Kaschak & Glenberg, 2004), and speakers align the production of sentence structure to the recent experiences with their fellow speakers' sentence structure (e.g. Levelt & Kelter, 1982), where these effects decay over time or with intervening sentences (Branigan, Pickering, & Cleland, 1999). Together, these experiments show that adults are continuously learning from experience with language. This kind of learning is evidently beneficial; it can ease the processing involved in subsequent encounters with that particular language feature to maximise processing efficiency. Learning mechanisms are therefore a crucial part of communication and must be understood before any model of semantic ambiguity resolution can be complete.

Learning from Recent Experience

There is increasing evidence to suggest that learning from recent experience also plays a role in guiding semantic ambiguity resolution, whereby comprehenders learn from experience with a word-meaning to improve the likelihood of correctly interpreting that ambiguous word in the future. The influence of a single word-meaning encounter on comprehension several seconds and minutes later has been observed across different tasks (e.g. sentence reading, speeded lexical decision) and measures (e.g. eye tracking, EEG). Where context constrains the meaning of the ambiguous word at test, it is consistently shown that word-meaning comprehension is facilitated on a second encounter when the meaning is consistent with the first encounter (Binder & Morris, 1995, 2011; Copland, 2006). Encountering the ambiguous word itself is crucial to this comprehension facilitation, since reading subordinate context alone in a prime sentence (i.e. without the ambiguous word itself being presented) does not facilitate comprehension of the subordinate word-meaning itself when it is read up to a few minutes later (Leinenger & Rayner, 2013).

Furthermore, comprehension can be (but is not always; Binder & Morris, 1995) impeded when the meaning of the second encounter is *inconsistent* with the first, showing that recent experience with a particular word-meaning can also hinder subsequent comprehension in cases where the subsequent encounter has the alternative meaning (Bainbridge, Lewandowsky, & Kirsner, 1993; Copland, 2006; Dholakia, Meade, & Coch, 2016; Simpson & Kang, 1994; Simpson & Kellas, 1989). Together, these very short-term (up to only a few minutes) priming studies clearly demonstrate that word-meaning representations are sensitive to very recent experience with those words, and can update rapidly to accommodate that experience. However, due to the prime-test delays being less than a few minutes, the time-course of the effect of recent experience and learning is not clear.

Slightly longer-lasting effects of recent experience on word-meaning interpretation have also been shown (Bainbridge et al., 1993). Participants completed a lexical decision task on an ambiguous probe word that was preceded by biasing sentential context (e.g. "the man kicked the machine after it returned his – *token*"). Participants encountered each ambiguous word once in block 1 and once in block 2.

Responses in block 2 were faster when the meaning was consistent with block 1, compared to inconsistent meanings where priming was eliminated. This indicates that priming effects with word-meanings surpass a few minutes, though Bainbridge et al. (1993) do not specify the exact time delay. Their findings also suggest that one encounter with each alternative word-meaning might balance each other out, such that one subordinate meaning and one dominant meaning encounter is the same as not encountering the word at all. However, Rayner et al. (1994, Experiment 2) failed to replicate this effect of recent experience using an eye tracking measure and it could be argued that the probe nature of the ambiguous words might have made the priming manipulation salient, which could have interfered with any observed effects. Clearly, more research is needed to investigate learning from recent experiences with ambiguous words.

Other research has concluded that interlocutors' interpretations of ambiguous words remain relatively stable over time (Geis & Winograd, 1974). This is also an implicit assumption of the highly influential reordered access model, which takes both immediate context and long-term knowledge into account, but does not mention possible changes in word-meaning representations through learning from experience over intermediate time periods (Duffy et al., 1988). This assumption of stable representations is undermined by the research investigating recent experience that has been outlined above, which has shown that interpretations of ambiguous words can change as a result of experience up to several minutes earlier. These experiments provide evidence to suggest that lexical-semantic representations might (usefully) update to incorporate information about a recently encountered word. Thus, it might be that preferred interpretations are stable over time but only provided that linguistic experience, or input, is stable. Perhaps if the less common meaning were regularly encountered, people would update their lexical-semantic representations to accommodate this information and change availabilities of word-meanings accordingly. This learning would benefit comprehension, since a representation would reflect the overall frequency with which a meaning is encountered and therefore an up-to-date likelihood of alternative meanings being encountered. This argument is consistent with the notion of dominance: the fact that people have dominant (more available because most often encountered) word-meanings shows that, at some level, people must learn from experience with those meanings and incorporate the evidence into their lexical-semantic representations.

Whilst a considerable amount of research has investigated how information about *new* words and meanings is learned/consolidated, particularly over a 24-hour period involving sleep (e.g. Dumay & Gaskell, 2007), or even over a week (Tamminen & Gaskell, 2013a), until recently relatively little work has focused on changes to the representations of *familiar* meanings of words (e.g. Fang & Perfetti, 2017). However, the few recent studies that do focus on changes to representations of familiar meanings confirm that recent and long-term linguistic experience can modulate, and sometimes even overturn, the meaning dominance of an ambiguous word (Leinenger & Rayner, 2013; Poort, Warren, & Rodd, 2016; Rodd et al., 2016; Rodd et al., 2013). These studies, along with others, use a "word-meaning priming" paradigm (Rodd et al., 2013) to investigate how listeners are able to learn from and develop lexical-semantic representations on the basis of experience. This recent research is building a picture, which suggests that we should move away from the view of adults having stable, unchanging lexical-semantic representations and towards a more flexible and dynamic view where representations continuously update to reflect experience with language.

What we know so far about word-meaning priming

Rodd et al. (2013) showed that, when listeners encountered ambiguous words such as 'fans' without any biasing context, they were 30-40% more likely to interpret the words as referring to the subordinate (less common) 'supporter' meaning if they heard that subordinate meaning in a sentence (e.g. 'the footballers were greeted warmly by the adoring *fans*') 20 minutes earlier. Hence, just a single subordinate encounter significantly increased the likelihood with which it is later used. This priming effect remained regardless of whether the same or a different voice was used for the prime sentence phase and the subsequent test phase, suggesting that word-meaning priming reflects an implicit updating of meaning frequencies in response to recent linguistic input, rather than relying purely on the conscious recall of episodic memories of the recently-used meanings (Experiment 2). Importantly, there was also evidence to suggest that this priming effect relied on repetition of the ambiguous word itself, and was not driven by a more general form of semantic priming (Experiment 3). Semantic priming from synonyms (e.g. *fan* – supporter) was evident at short prime-target delays (3 minutes) but was eliminated at the longer delays at which word-meaning priming has been studied (20 minutes or more). This finding is consistent with previous work showing that context alone (repetition of context *without* repetition of the ambiguous word per se) can affect later word interpretation over shorter prime-test intervals of a few minutes (Colbert-Getz & Cook, 2013). Finally, Rodd et al. (2013) showed that the more subordinate meanings at prime benefitted more from subordinate meaning priming than the more dominant meanings. In other words, participants showed a greater priming effect for less frequent word-meanings (Experiment 1). This suggests that people are able to learn more from recent experience with (on average) unexpected meanings.

In addition to these effects of prior experience with ambiguous words that occur within a single, controlled experimental setting, this priming effect has also been replicated in naturalistic settings (Rodd et al., 2016, Experiment 1). When priming was conducted over a radio show and participants took part in a web-based test in a location of their choice up to several hours later, the same word-meaning priming pattern emerged, showing that word-meaning priming extends beyond a controlled environment. Rodd et al. (2016) also demonstrated that if a person repeatedly uses/hears a word with its subordinate meaning over longer timescales of months or years, the meaning dominance for that word can be altered. Recreational rowers, who know additional rowing-related meanings for common English words (e.g. ‘feather’ and ‘square’ refer to positions of the oar), tend to interpret these words as the rowing-related meanings, in light of their experience with these meanings, even in non-rowing contexts. This tendency was significantly positively associated with additional years of rowing experience and significantly negatively associated with time since the most recent rowing practice. Moreover, rowers who had rowed the day of the test were significantly more likely to generate rowing responses than those who had just rowed the previous day, or had not rowed recently at all (Experiment 4). Together, these findings show that long-term and short-term experience both affect lexical-semantic representations.

Converging evidence comes from experiments using ambiguous words that have additional baseball-related meanings (Wiley, George, & Rayner, 2016). Baseball experts, compared to non-experts, have more difficulty disambiguating sentences when they are strongly biased towards the non-baseball meaning. Again, this shows a difficulty to disambiguate a word when the encountered meaning is inconsistent with one's prior long-term experience. Taken together, these studies show that adults accumulate evidence across their lifespan to build and update lexical-semantic representations, learning from linguistic experience across a range of timescales to guide interpretation.

Rodd et al. (2013) proposed that the mechanism for the updating of word-meaning representations involves changes to connection strengths among units in the distributed connectionist network (Rodd et al., 2004), as this would allow transient changes in meaning availability to slowly accumulate across a lifespan. This learning mechanism, which has been proposed as an explanation for other types of long-term priming (e.g. Becker, Behrmann, Moscovitch, & Joordens, 1997), involves small but persistent changes to connection strengths between the relevant units within and/or across representational layers. For the updating of word-meaning representations, the changes to connection strengths reflect a build-up of evidence about the likelihood of a given meaning. In this way, comprehenders can gradually and continually learn about language.

Aside from updating representations based on recent experience, there is recent evidence about other types of information that can be learned from encountering an ambiguous word (Cai, Gilbert, Davis, Gaskell, Farrar, Adler, & Rodd, 2017). British English participants were more likely to retrieve the American-dominant meaning of a word (e.g. the 'hat' meaning of *bonnet*) if they had previously heard that word in an American accent, than if they had previously heard it in a British accent (where the alternative, 'engine cover' meaning of bonnet is dominant). Whilst this was not a word-meaning priming experiment, it does demonstrate that listeners can perceive subtle details in language and can make use of them to influence the later interpretation of words. This is clearly an adaptive comprehension strategy; listeners use what they know about the identity of a speaker to assimilate their interpretation towards the most likely intended meaning, maximising their chances of correctly interpreting the word.

Although experience with a particular speaker's accent can affect word interpretation (Cai et al., 2017), comprehenders' interpretation does not seem to be influenced by all types of word-form information. The word-meaning priming effect has been shown to be resistant to changes in modality between prime and test (Gilbert et al., 2018). In the prime phase of the experiment, ambiguous words were presented in subordinate-biasing sentence context, either in spoken or written form (or not presented, as an unprimed baseline). Twenty minutes later, the words were presented in spoken or written form using word association (Experiment 1), or speeded semantic relatedness (Experiment 2), as a means of testing the interpretation of the words in light of recent experience with them in the prime phase. Both experiments showed that all primed meanings were retrieved more often and more quickly than unprimed meanings, regardless of the prime modality, test modality and congruence between prime and test modality.

These findings provide useful evidence in uncovering the mechanism(s) involved in word-meaning priming. They are inconsistent with the explanation from Rodd et al. (2013) that word-meaning priming is the result of changes to form-to-meaning connections, since this would assume a benefit for unimodal priming, which was not found by Gilbert et al. (2018). It seems more likely that, as Rodd et al. (2016) suggested, the changes to connections could happen *within* the lexical-semantic layer such that connections are strengthened with priming, which increases the width or depth of the attractor basin, making it more likely to be selected on a subsequent encounter with the ambiguous word. More work is needed to investigate the exact nature of mechanism(s) underlying word-meaning priming. Regardless, these findings demonstrate that adult comprehenders benefit from learning from recent experience in a flexible way and that the modality of recent experience is immaterial to this benefit. Together, these experiments on recent experience demonstrate the flexibility with which adults can disambiguate ambiguous words to maintain an up-to-date likelihood of occurrence.

My research topic

The aim of this thesis is to investigate how lexical-semantic representations can be retuned on the basis of recent experience. Word-meaning priming will be used as a tool to examine how interlocutors learn from experience to inform their subsequent comprehension. A total of nine experiments and one pretest were run, with the data from 986 participants analysed in total. Eight experiments investigate in detail the effects of recent experience on the comprehension of ambiguous words, whilst a further experiment and pretest provide a set of picture stimuli that can be used to measure these effects of recent experience. In particular, Experiments 1, 2 and 3 (Chapter 2) investigate how multiple recent encounters with a particular word-meaning affect the subsequent interpretation of that word. Using a newly-developed picture stimuli set and picture semantic relatedness test (Experiment 4, Chapter 3), Experiments 5, 6, 7, 8 & 9 (Chapter 4) investigate whether the word-meaning priming effect is driven by increased availability of the primed meaning alone, or by the combination of increased availability of the primed meaning and decreased availability of the unprimed meaning.

Chapter 2: Effect of multiple repetitions on lexical-semantic representations

Introduction

The continual updating of word-meanings, driven by recent experience, plays a critical role in maintaining a common ground among interlocutors in language communication (Rodd et al., 2016). It is also crucial for helping the listener to avoid misinterpreting a word and, as a result, having to engage in effortful reinterpretation processes (Rodd, Johnsrude, & Davis, 2010). It seems that interlocutors update their lexical-semantic representations based on their experience with the meanings of words. This allows comprehension to benefit from the most up-to-date likelihood of a particular meaning being the correct interpretation whenever an ambiguous word is encountered. People are able to capitalise on experience with words so that they can flexibly alter representations based on both longer-term (Rodd et al., 2016) and shorter-term (Rodd et al., 2013) experience. Unlike a view of lexical-semantics where representations remain stable throughout adulthood, this dynamic "updating" approach suggests that adults' comprehension is made more efficient by continuously learning from experiences with word-meanings to make a "best guess" about the most likely intended meaning at any point in time.

The recent experiments on shorter-term word-meaning priming (Rodd et al., 2016, Experiments 1 & 2; Rodd et al., 2013) have tended to investigate the impact of encountering only one prior instance of an ambiguous word, thus it is unclear how word-meanings are updated by *multiple* recent encounters. For instance, recent encounters could have the same or different meanings and could be clustered or more spaced over time. The present experiments investigate how these different types of recent encounters may differentially affect the updating of word-meaning representations.

However, the mechanism that allows for word-meaning updating in response to recent experience is not clear. The finding that priming effects persist over 20-40 minutes in lab-based experiments (Rodd et al., 2016; Rodd et al., 2013) and several hours in more naturalistic settings (Rodd et al., 2016, Experiment 1) means that these changes in word-meaning availability are not easily accounted for by short-term priming mechanisms such as residual activation (e.g. Dell, 1986; McClelland & Rumelhart, 1981; Meyer & Schvaneveldt, 1971). Similarly to the incremental learning account of repetition priming and semantic interference in speech production from Oppenheim, Dell, and Schwartz (2010), Rodd et al. (2013) suggest that every encounter with an ambiguous word strengthens the connection between the word and the encountered meaning, such that experience with word-meanings accumulates to enhance comprehension over time. More specifically, they proposed that the mechanism for the updating of word-meaning representations involves changes to connection strengths among units in a connectionist network (Rodd et al., 2004), as this would allow transient changes in meaning availability to slowly accumulate across a lifespan, which reflects a build-up of evidence about the likelihood of a given meaning.

As for the relative likelihood of *different* meanings, if listeners continue to encounter both the dominant and subordinate meanings of a word, it is likely that they strengthen the relevant connections in proportion to the overall frequency with which each meaning is encountered, such that the availability of the different meanings reflects the relative frequencies of these encounters. For example, disambiguation of 'bark' could be influenced by recent encounters of *both* the 'dog noise' and 'tree covering' meanings. If an individual's experience with a particular word changes systematically with time then, given sufficient experience, a previously subordinate meaning could eventually become the dominant meaning (which seems to be the case for the rowers reported in Rodd et al., 2016). As described by Rodd et al. (2013), connectionist models can accommodate this mechanism so long as they allow for updating/learning to continue throughout the model's "lifespan". In summary, it seems likely that repeated encounters with a word-meaning gradually strengthen the relevant connections in the lexical-semantic network and, over a relatively long period of time (e.g. months, years), can change an individual's preferred meaning.

What is less clear is whether repeated encounters within a relatively *short* period of time (e.g. 20-30 minutes, compared to a lifetime of experience) can lead to similar cumulative effects in updating the representations of word-meanings. Changes in representation availability following a single encounter with a particular meaning do occur (Rodd et al., 2013), (also see Bainbridge et al., 1993; Binder & Morris, 1995; Copland, 2006; Masson & Freedman, 1990, for comprehension facilitation from recent encounters in the space of a minute) but it is not known whether these relatively short-term changes in availability are sensitive to multiple, repeated encounters of a particular meaning within the same time-frame. It is also unclear whether repeated encounters of *different* meanings of an ambiguous word accumulate to have a combined effect on comprehension.

The repetition priming literature shows that multiple repetitions of words in a short space of time do increase the magnitude of priming compared to one repetition. This has been shown in lexical decision (Forbach, Stanners, & Hochhaus, 1974; Forster & Davis, 1984), word naming (Durso & Johnson, 1979), passage reading (Kolers, 1976), free recall, cued recall and recognition (Nelson, 1977). A similar effect of repetition has been found in a test of explicit recall of words from a sentence, in which two presentations of an ambiguous word in a sentence improved recall compared to one presentation (Thios, 1972). However, this improvement was lessened when the second presentation used the alternative meaning of the ambiguous word, suggesting that encountering the dominant meaning interfered with the updated representation from an earlier encounter with the subordinate meaning. Together, these results indicate that multiple repetitions of an ambiguous word might lead to greater word-meaning priming than only one repetition, and that the effect of an initial exposure to a word-meaning might be disrupted or abolished by a subsequent exposure to an alternative meaning of the word. However, the findings reported by Thios are in the explicit memory domain and therefore may be driven by different mechanisms to word-meaning priming (see Rodd et al., 2013), so it is not clear whether the repetition benefit and the interference from an alternative meaning would replicate in a less explicit learning paradigm.

Given the repetition literature, it seems possible that multiple repetitions of an ambiguous word-meaning increase the likelihood of interpretation of the word towards that meaning compared to a single repetition. As argued above, this could occur through a process of cumulatively updating the relevant connection strengths within the lexical-semantic system upon each encounter with the word and meaning. However it is not clear whether the temporal spacing of these updates would further influence any such repetition benefit. That is, it remains unclear whether a particular temporal distribution of repetitions is most effective in changing the availability of word meanings: repetitions that are massed (i.e. temporally compressed), or repetitions that are spaced (i.e. temporally distributed). The existing literature shows inconsistent findings, such as no spacing benefit for cued recall (Greene, 1989), spacing benefit over massed for free recall (Madigan, 1969; Melton, 1970; Underwood, 1970) and no spacing benefit for free recall (Paivio, 1974). Multiple repetitions must at some level influence meaning availability over one repetition, otherwise the overall meaning dominance effect, (i.e. more frequent meanings being easier to access than less frequent meanings), and the increased availability of rowing meanings for rowers (Rodd et al., 2016), would not exist. Furthermore, if repetitions of different meanings are encountered then they might strengthen the relevant connections in proportion to the overall frequency with which each meaning is encountered, suggesting that a single subordinate followed by a single dominant repetition would both have an effect on how that word is later interpreted. Another possibility is that the relatively short-lived word-meaning priming effects, lasting e.g. 20-40 minutes, are solely driven by the most recent word-meaning priming encounter and that earlier encounters during this same timescale leave no (or minimal) trace. Under this view, the fact that the most recent encounter takes precedence over prior recent encounters would mean that changes to word-meaning preferences that occur over longer timescales (e.g. from days onward) would involve a different or additional learning mechanism, such as overnight consolidation.

The experiments reported in this chapter investigate, for the first time, whether and how recent repetitive encounters of ambiguous words in particular meaning contexts affect the availability of the primed meanings. Each of the three experiments follow the word-meaning priming paradigm first used by Rodd et al. (2013). Participants were exposed to repetitions of ambiguous words in subordinate meaning contexts and, after a filler task, these words appeared in a word association test to assess how the availability of the subordinate meaning had changed as a result of the prior exposure. This word association task, in which participants must comprehend a given word in order to respond with the first word that comes to mind, allows us to assess how ambiguous words are interpreted in the absence of the constraining semantic contexts that are used in tasks such as semantic relatedness judgments and thus provides a straightforward measure of participants' default/preferred meanings. Broadly speaking, we assume that when participants provide an associate for a word, they first bring to mind one of the word's meanings, and then report the first-generated associate of that meaning. Importantly, it does not seem to be the case that priming, as measured by word association, is driven purely by words remembered specifically from the prime sentence for an ambiguous word (items referred to as "primed associates"). That is, the priming effect does not rely on participants producing a response word at test that was encountered within the specific prime sentence (e.g. producing at test 'footballers' after being primed with 'the footballers were greeted warmly by the adoring *fans*'), since removing these primed associates from the test data does not alter the pattern of priming (Rodd et al., 2013; Experiment 1). For these reasons, the word association test has become a commonly-used method for assessing word-meaning priming and will therefore be used in the present experiments (Cai et al., 2017; Rodd et al., 2016; Rodd et al., 2013).

In what follows we examine how multiple recent encounters with an ambiguous word, either in the same or a different meaning context, affect the later interpretation of these words (Experiment 1 [1]), and how this interpretation is influenced by the relative timing of multiple subordinate meaning repetitions (Experiments 2 and 3).

[1] Whilst the data for Experiment 1 were collected for a previous MSc degree, the re-analysis of its data using mixed effects modelling was conducted as part of this PhD and is therefore included in this thesis.

Experiment 1 – one & three massed subordinate repetitions, one dominant repetition

Experiment 1 had two aims. The first was to investigate whether multiple recent encounters with the *same* subordinate meaning boost the word-meaning priming effect compared to one encounter. Based on the mechanism for updating of word meaning representations proposed by Rodd et al. (2013) and Rodd et al. (2016), which assumes that the effects of multiple encounters with ambiguous will accumulate over time, we predict that multiple subordinate repetitions presented within the same spoken paragraph (i.e. massed presentation) will boost meaning priming compared to one subordinate repetition. If this is the case, then it suggests that lexical-semantic representations are sensitive to the frequency of encounters during this time period and update cumulatively during this process.

The second aim was to examine the effects of encounters with *different* meanings of an ambiguous word. Specifically, we examine the case where the listener first encounters the subordinate meaning and then encounters the dominant meaning of the same word. The view that the effects of multiple encounters will accumulate over time predicts that both of these encounters have an impact on subsequent disambiguation such that the dominant repetition will reduce the impact of the earlier exposure to the subordinate meaning. However, we also predict that there will still be a residual effect of the prior subordinate repetitions, compared to the case where only the dominant meaning is presented. If this were the case, then again it would support the view that lexical-semantic representations are updated in an incremental manner to reflect the relative frequency with which meanings occur.

This experiment used a modified version of the word-meaning priming paradigm developed by Rodd et al. (2013) with the addition of a dominant prime phase. That is, participants completed the subordinate prime phase, filler task, dominant prime phase and then a word association test phase (See Figure 1 for an overview of the procedure). In the subordinate prime phase, participants encountered a subset of the ambiguous words in the context of their subordinate meanings, either once or three times in massed presentation. The remaining (unprimed) ambiguous words were only presented during the test phase, which provided a baseline measure of meaning dominance for these items against which to compare the primed conditions. Hence, the prime phase involved three conditions: unprimed baseline, one repetition and three massed repetitions. After a filler task, which created a prime-test delay, participants encountered half of all words one more time, but in the context of their dominant meanings. Finally, in the word association test, participants heard all ambiguous words in isolation and responded with an associate, which provided a measure of each participant's interpretation of the words. The mean length of the tasks resulted in an average delay between each item in the subordinate prime task and the word association task of approximately 30 minutes.

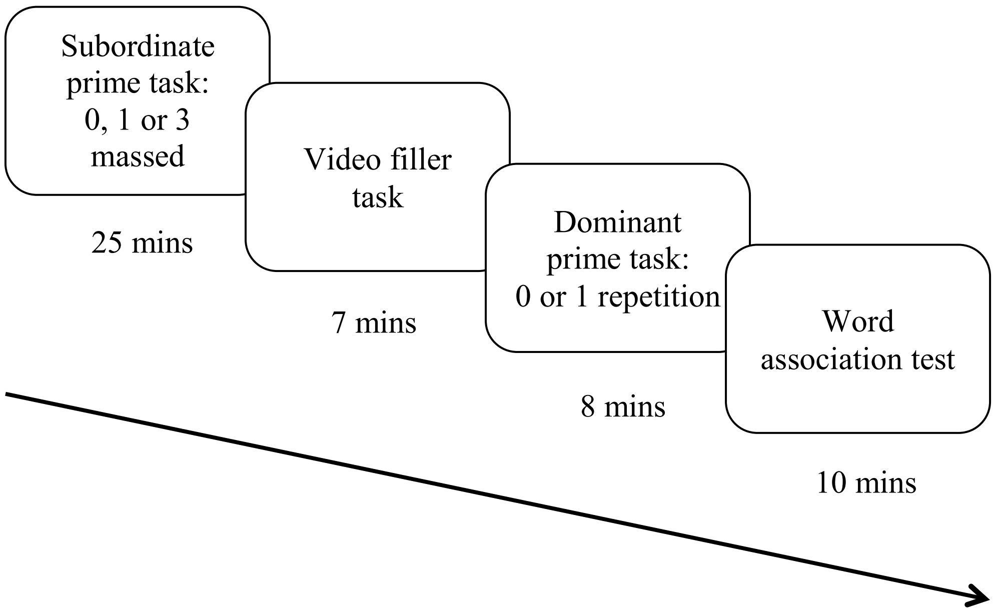

Figure 1. Experiment 1 task order, including prime phase elements, filler task and test. The mean duration of each task is displayed within the figure.

Method

Participants

Thirty-three native British English speakers participated in the current experiment. However, only the data from 30 participants (23 females; mean age = 24.8, range = 18 – 40) were analysed: one participant was excluded for exceeding age requirements and two participants were excluded due to a software error, which prevented task completion. All participants reported that they had no language, hearing or vision impairments (other than corrected-to-normal vision) and had lived in the UK for the majority of their lives, speaking English as their first language from birth. Participants were recruited via the University College London online recruitment system or advertisements on the university campus and paid the standard rate at the time of £6/hour[2].

Materials

Sixty ambiguous words (e.g. bark, cabinet) were selected from a pretested set that had assessed dominance using a standard word association test (Warren, Vitello, Devlin, & Rodd, in preparation); see Appendix A for ambiguous word list. These words had a dominance rating of 12-42% for the subordinate meaning (mean of 25%). In all cases the primed subordinate meaning had the same pronunciation and spelling as the dominant meaning, although in some cases there was an additional meaning with a different spelling (e.g. ‘break/brake’). Polysemous words were also included as long as the related meanings were judged by the author as sufficiently distinct that they could be distinguished on the basis of word association responses (e.g. typical associates related to the two related meanings of ‘wave’, *disturbance in water* or *hand gesture*, were deemed sufficiently distinct, whereas those to the two meanings of ‘passage’, *corridor/tunnel* or *journey over time/distance*, were not. Thirty-eight words were classed as polysemous; Parks, Ray, and Bland (1998)).

---

[2] All experiments in this thesis were approved UCL Division of Psychology and Language Sciences Ethics Committee, fMRI/2013/001.

For the subordinate prime task, a total of 60 short paragraphs (mean length of 70 words) were composed in the style of a media or literature excerpt. Each paragraph contained one of the 60 ambiguous words, disambiguated towards the subordinate meaning[3]. For the three repetition condition, the ambiguous word was used in the paragraph three times and was therefore massed in presentation (i.e. the three repetitions appeared in quick succession, within the same paragraph). The first presentation of the word always occurred in the first sentence, with the second and third repetitions distributed throughout the remainder of the paragraph, e.g.:

*'The **cabinet** concluded that a referendum would be unnecessary, since the time it would use might only worsen the financial situation. The **cabinet** had been in talks for several weeks about a plethora of problems, but had only discussed the idea of a referendum over the last few days. Their decision was not a popular one, since previous **cabinets** held many referenda, which had proven popular with the public.'*

For the one repetition condition, the paragraphs were identical to the three repetition condition except that the second and third repetitions were replaced with a substitute word of a similar meaning. This was done to remove the instance of the ambiguous word itself without altering the global meaning or length of the paragraph. For example, the one repetition version of the passage above was created by replacing '*cabinet*'/'*cabinets*' in the 2nd and 3rd sentence with '*politicians*'. To fully control the number of repetitions, the ambiguous words did not appear anywhere in the experiment except for their respective priming paragraphs and in the test task. The paragraphs were spoken by a female British English speaker and were digitally recorded in a sound-proof booth. For each paragraph, we created a written summary sentence (mean length 8.8 words), and participants rated how well this sentence summarised the paragraph (in order to encourage close attention to the paragraph; see Procedure). The summary for a given item was the same for both the one and three subordinate prime conditions. All summaries were designed to be a similarly reasonable level of quality (as quality-judgment/relatedness was the task for the participants, as explained in the Procedure).

[3] There were no unambiguous prime items.

These 60 ambiguous words formed the basis of the auditory word association test, with the addition of five unambiguous filler words that preceded these target items in the test. All words were recorded by the same female speaker as the prime paragraphs (see Rodd et al., 2013 for evidence that word meaning-priming is not dependent on, or enhanced by, consistency in speaker identity between prime and test).

Sixty sentences (mean length 9.2 words) were created for the dominant prime task. In each sentence, an ambiguous word was disambiguated towards the dominant meaning (e.g. 'the cherry wood *cabinet* looked magnificent'), that is, a different meaning from in the subordinate prime test. These sentences were digitally recorded by a male speaker with a similar accent to the female speaker of the paragraphs. Each sentence was coupled with a written probe word that was either related (50%) or unrelated to its content (e.g. 'furniture').

A video animation ('Shaun the Sheep', Aardman, 2010) was chosen as the filler task for several reasons. First, since controlling exposure to language is a key element to the word-meaning priming paradigm, this animation is ideal, as it does not involve any spoken or written words. Second, the content is not strongly related to any of the primed word meanings, and does not carry any strong emotional valence (strong valence stimuli were avoided for this task, as emotion can affect recall, e.g. Bock & Klinger, 1986; Cahill, Haier, Fallon, Alkire, Tang, Keator, Wu, & McGaugh, 1996). Third, the animation is engaging for participants.

Design

This experiment had a within-subject/between-item and within-item/between-subject experimental design with two independent variables: subordinate meaning repetitions (3 levels: unprimed (no repetition), one repetition, three massed repetitions) and dominant meaning repetition (2 levels: unprimed (no repetition), one repetition). The dependent variable was the proportion of responses from the word association test that were consistent with the *subordinate* meaning used in the priming paragraphs.

Each participant encountered each of the 6 conditions, with 10 items in each. The assignment of items to condition was rotated across six versions of the experiment, allowing each item to appear in only one priming condition for a given participant, yet across different participants, each item appeared in every priming condition. The number of items per condition and participant is shown in Table 1.

Table 1. Ambiguous word repetition design for the six experimental conditions in Experiment 1.

Task	Number of items encountered
Subordinate prime task	20 homophones – one repetition 20 homophones – three repetitions [20 homophones – unprimed baseline]
Filler task	(Video)
Dominant prime task	10 subordinate one repetition homophones 10 subordinate three repetitions homophones 10 subordinate unprimed homophones
Word association test	All 60 homophones

Note. Twenty ambiguous words (shown in grey) were not encountered in the subordinate prime phase but were later included in the word association test to act as an unprimed baseline against which to compare any word-meaning priming effects.

Procedure

The experiment was run in a cubicle, using Qualtrics survey software (Qualtrics Inc., www.qualtrics.com). The experiment was displayed on a desktop computer but the video for the filler task was presented to participants on an Apple iPad. Participants wore headphones for the whole experiment to ensure that the stimuli could be heard easily and to minimise any background noise. Each participant was randomly assigned to one of the six versions of the experiment. After giving their informed consent, participants' demographic data were collected and instructions for the experiment were displayed on screen. Trials within each task (subordinate prime task, dominant prime task, and word association) were randomised, each presented on a new page, with a mouse click (on-screen button) required to proceed to the next trial. Participants were given a practice trial and the chance to confirm instructions with the experimenter before each task. See Figure 1 for the sequence and timings of experimental tasks. To distract from the purpose of the experiment, participants were informed that they were taking part in two separate experiments. They were told that the "first experiment" (the subordinate prime task) was to pretest stimuli for another experiment and quality-check the summaries of the paragraphs, having been told that we were interested in their real opinion; the "second experiment", they were told, consisted of watching a video and carrying out a filler task and then a final main task (in fact the dominant prime task and then the word association task, respectively).

Subordinate Prime Task

In each of 40 trials participants heard an excerpt, which included the ambiguous word in the context of the subordinate meaning, either once or three times, and saw the accompanying summary on screen simultaneously. Participants were asked to rate on a five-point scale how well the summary sentence summarised the key information in the excerpt (1 – poorly to 5- excellently).

Filler Task

For the video animation, one of two selected episodes was played to participants (episode 1 length: 5 minutes, 55 seconds; episode 2 length: 5 minutes, 54 seconds). Participants were informed that they should pay attention to the content of the video, as they would be required to answer questions about it at the end of the experiment (although they were not asked questions, as this was only to disguise the aim of the experiment).

Dominant Prime Task

Participants subsequently completed the dominant prime task in which they were asked to listen to 30 sentences, each of which included an ambiguous word disambiguated towards the dominant meaning. For each sentence, they were asked to decide whether the sentence was semantically related to a probe word. The probe word was presented visually on-screen during the sentence presentation, with 'related' and 'unrelated' buttons displayed. Although participants could respond before the end of the sentence, they were encouraged not to do so and to be as accurate as possible (participants were less likely to be accurate if they responded before sentence offset). This relatedness task was included to ensure that participants attended to the sentences and processed their meanings.

Word Association Test

Although the presentation order of experimental items in the word association test was randomised, the five filler items were always presented at the start of the test to get participants used to the nature of the task. Items were presented auditorily and participants were asked to type the first word they thought of when they heard each word into a textbox on the screen[4]. They were asked to type '0' if they were unable to make out the word, unable to generate a response or felt uncomfortable giving one.

Post-Experimental Tasks

There were two tasks after the main experiment: awareness test and response-coding. For the awareness test, participants were asked two questions: 'What do you think the aim of the experiments was?' and 'How many words from the word association do you recognise from the tasks earlier in the experiment?' to measure awareness of the priming manipulation and investigate its impact on priming.

Participants were then asked to code their word association responses (blind to experimental condition) to clarify the meaning of each word that they had intended in their response. In this response-coding task, participants were presented with each

[4] The offset of the spoken word within the auditory file and the presentation of the type-in prompt were not synchronised, which meant that analyses of reaction times were not possible for the experiments in this chapter.

word and their response. Provided with short definitions of the dominant and subordinate meanings of each item, they were asked to select to which meaning their response was related (or 'other' meaning), following the method of Rodd et al. (2016). Finally, participants were debriefed and were given the opportunity to ask questions.

Task and Coding Checks

Subordinate Prime Task

All participants used the range of the five-point scale for the summary ratings adequately indicating that they were engaged in comprehending the paragraphs - 87% used the full range; those who did not use the full range did not rate any summaries as the lowest rating, which most likely reflects that the summaries were designed to be accurate. Summary rating means were consistent across subordinate prime conditions (one subordinate repetition *mean*: 3.56; *SD*: 1.25, three subordinate repetitions *mean*: 3.59; *SD*: 1.32).

Dominant Prime Task

All participants demonstrated accurate semantic relatedness judgments for the target words in this task (at least 80% correct responses), suggesting adequate engagement in the task.

Word Association Test

Responses were coded by participants as either (1) related to the dominant meaning of the homophone, (2) related to the subordinate meaning of the homophone, or (3) related to another meaning, 'other'. To check that participants had coded responses correctly, the experimenter verified a 5% subset of coded responses. Since there were several incorrect codes, *all* coded responses (1s, 2s and 3s) were then verified by the experimenter by checking each code alongside the respective word association response. Any word association responses that were clearly associates of either the dominant or the subordinate meaning were recoded as such. For example, where participants coded their response ‘hot’ as ‘other meaning’ to the item ‘cold’ (presumably because it has the opposite meaning), their response was recoded as being related to the dominant (temperature) meaning by the experimenter. Because we were primarily interested in changes in the proportion of responses consistent with the primed subordinate meaning, for the analyses, ‘other’ responses (6%) were removed to provide a coded data set that indicated whether a participant gave a subordinate prime-consistent response or the dominant meaning of the ambiguous word.

Results

Main Analyses

As is clear from the pattern of subject means in Figure 2, and as predicted, the subordinate priming increased the proportion of subordinate meaning responses, and the subsequent dominant priming reduced the proportion of subordinate responses. Interestingly, there seems to be little difference in priming between one and three subordinate repetitions.

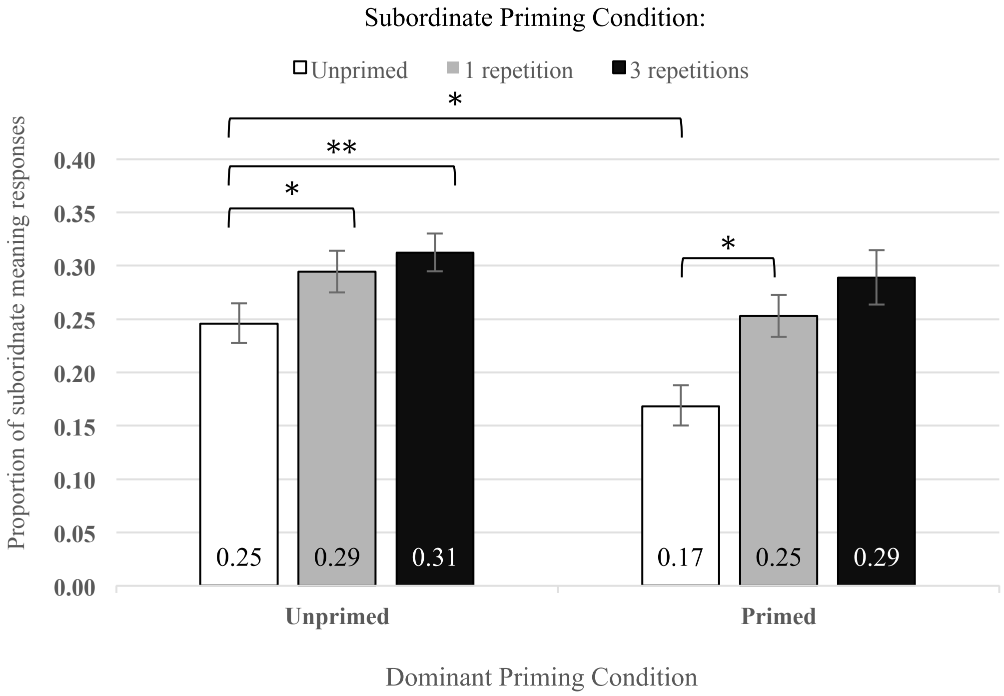

Figure 2. Experiment 1. Subject mean proportion of word association responses consistent with the primed subordinate meaning, with standard error bars adjusted for the within-subjects design[5]. Significance level indicated with asterisks (* < .05, **<.01) and simple effects shown for the theoretically important contrasts.

The word association data were modelled using logistic mixed effects modelling, with the glmer function from the lme4 package (version 1.1-7; Bates, Mächler, Bolker, & Walker, 2014) in R (version 3.3.1; R Core Team, 2016). Mixed effects modelling is the most appropriate form of analysis for the present data since these data are binary, responses being subordinate *or not,* and this form of analysis takes the within-subject and within-item dependencies into account within a single model (Jaeger, 2008). As the subordinate meaning repetitions factor had three levels, we used two Helmert contrasts for this factor. These contrasts allowed for separate estimates of i) the overall effect of subordinate priming (subordinate unprimed versus

[5] Whilst logistic mixed effects modelling was used to analyse all data in this thesis, it does not provide "interpretable" means, hence all relevant figures show the subject means. For this reason, there may be some slight discrepancies between the results of mixed effects analyses, which account for both item- and subject-specific effects, and the results implied by the subject means in the figures. However, this does not affect the pattern of results in any case.

the two subordinate repetition conditions combined) and ii) the effect of number of repetitions (one versus three subordinate repetitions, omitting the unprimed control). Both factors were deviation coded for ease of interpretation of the model coefficients (subordinate repetitions contrast 1: unprimed = -2/3, one repetition = 1/3, three repetitions = 1/3; subordinate repetitions contrast 2: unprimed = 0, one repetition = -1/2, three repetitions = 1/2; dominant repetition: unprimed = -1/2, one repetition = 1/2).

A model was then built with five fixed effect coefficients (two to represent the subordinate meaning repetitions factor, as defined by the Helmert contrasts, one fixed effect for dominant meaning repetition, and two to represent the interaction between each of the subordinate meaning contrasts and the dominant factor) with a maximal random effects structure, as recommended to protect against inflated Type I error (Barr, Levy, Scheepers, & Tily, 2013). This full model failed to converge across all tests of main effects and interactions (most likely due to the complex random effects structure), so here and in subsequent experiments we followed the recommended protocol for dealing with non-convergence from Barr et al. (2013). The random effects structure was simplified by removing one random effect term at a time (correlations removed first, then intercepts, then slopes[6]; the subject or item term that explained the least variance was removed first) until all of these nested models also converged. This resulted in the final model having a random effects structure comprising the subject and item intercepts-only[7]. A model comparison approach (e.g. Baayen, Davidson, & Bates, 2008) was then used to determine the significance of the main effects of the subordinate and dominant meaning repetitions and their interaction. This approach involved individually removing the fixed factor of interest (e.g. the interaction term) and comparing it to the main model using a likelihood ratio

[6] Where the slopes were removed, the intercepts were put back into the model.

[7] Whilst a maximal random effects structure does seem to protect against inflated Type I error (Barr et al., 2013), the size of this inflation is still under debate. More recent research has shown that an intercepts-only random effects structure does not necessarily inflate Type I error (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), and this model structure is still preferable to the equivalent separate within-subject/item ANOVAs, since mixed effects modelling allows these analyses within a *single* model.

test to examine whether the inclusion of the fixed factor of interest resulted in a significantly better model fit. Although the subordinate repetitions and interactions factors were each split into two by the Helmert contrast codes (see above for details), the two factors for each were either left in the model as a whole or removed as a whole for tests of the subordinate main effect and the interaction, respectively. In each case, a model without the fixed factor of interest was compared to the full model using a likelihood ratio test.

The main effect of subordinate repetitions was significant (X^2 (2) = 16.64, p < .001), showing that there were more subordinate-meaning word association responses following subordinate priming. The main effect of dominant repetition was also significant (X^2 (1) = 6.68, p = .009), indicating that dominant priming reduced the number of subordinate meaning word association responses. However, the interaction between subordinate and dominant repetitions was not significant (X^2 (2) = 1.71, p = .430), meaning that the interaction term did not significantly improve model fit compared to the model that only included the linear combination of the two predictors. This finding indicates that the reduction in subordinate meaning interpretations due to the dominant meaning encounter did not significantly vary as a function of the number of subordinate prime repetitions.

The overall significance of the subordinate repetitions factor appeared to be attributable to a significant difference between the subordinate primed and unprimed conditions; the model coefficient for the primed (both one and three subordinate repetitions) versus unprimed contrast was significant, (β = 0.49, *SE* = 0.13, z = 3.87, p < .001), while the model coefficient for the one versus three repetitions contrast was not significant, (β = 0.15, *SE* = 0.14, z = 1.12, p = .260). Pairwise comparisons with Tukey adjustment for multiple comparisons were conducted using the glht (general linear hypothesis testing) function in the multcomp package (version 1.4-1; Hothorn, Bretz, & Westfall, 2008). Comparisons confirmed that the one and three repetition conditions were both significantly different from the unprimed condition (β = -0.55, *SE* = 0.21, z = -2.54, p = .020 and β = -0.75, *SE* = 0.21, z = -3.58, p = .001, respectively).

In order to address questions about the significance of differences between specific conditions, we conducted a set of four simple effects analyses using subsets of the data, with Tukey-adjusted *p* values for *post-hoc* comparisons. First, for subordinate unprimed *and* dominant primed words (i.e. words *not* presented during the subordinate prime phase but later presented during the dominant prime phase), there was a significant dominant priming effect where one dominant repetition increased the number of dominant word association responses compared to the unprimed baseline condition (which was subordinate *and* dominant unprimed; $\beta = -0.52$, $SE = 0.21$, $z = -2.44$, $p = .010$). This confirmed that the main effect of dominant repetitions was applicable to this particular simple effect comparison, demonstrating that, like the subordinate meaning, a recent encounter with the *dominant* meaning of an ambiguous word biases the later interpretation of that word toward that same meaning, compared to when there is no recent encounter at all (i.e. the unprimed condition). Second, when words were primed with one subordinate repetition followed by one dominant repetition, this did not significantly alter word association responses compared to the unprimed baseline ($\beta = 0.03$, $SE = 0.20$, $z = 0.14$, $p = .890$). This result suggests that one subordinate-meaning exposure shifts meaning preferences towards the subordinate meaning, and a subsequent exposure to the dominant meaning shifts meaning preferences back again, so that the effects of exposures to the two different meanings balance each other out. In other words, the combination of one subordinate and one dominant meaning exposure results in the returning of meaning preferences to a net level that is not significantly different to the unprimed baseline.

Most importantly, the combination of one subordinate and then one dominant repetition resulted in significantly more subordinate-meaning responses than exposure to one dominant repetition alone ($\beta = -0.54$, $SE = 0.21$, $z = -2.50$, $p = .030$). This shows that it is not only the most recent encounter that affects the priming-related shift in meaning preferences, but that an earlier encounter with an alternative meaning leaves a residual effect on preferences. However, the trend that three subordinate repetitions prior to the dominant repetition resulted in more subordinate-meaning responses than one subordinate repetition prior to the dominant repetition was not significant ($\beta = -0.21$, $SE = 0.20$, $z = -1.06$, $p = .540$). This indicates that whilst an encounter with the subordinate meaning before exposure to the dominant meaning leaves a residual priming effect, three encounters with this subordinate meaning before the dominant meaning exposure do not significantly increase this residual subordinate priming effect further.

Awareness Checks

There were two awareness measures: awareness of experimental aim and awareness estimate, both of which were analysed with logistic mixed effects modelling to investigate their effect on priming. Two participants were removed due to missing data on the awareness test. One experimenter (HNB) coded the responses to the awareness of experimental aim question. If participants demonstrated some, or full, correct awareness of the experimental aim (e.g. 'to see if the original sentences influenced my later associations'), their responses were coded as aware, whereas if they demonstrated little/incorrect or no awareness of the aim (e.g. 'how large or small people's semantic fields are'), their responses were coded as unaware, hence these data were dichotomous. Fifteen participants were unaware of the aim (priming effect across subordinate repetition conditions *mean* = .33, *SD* = .09) and 13 participants were fully/partially aware of the aim (priming effect *mean* = .27, *SD* = .07). The awareness estimate data were continuous, indicating participants' estimates of the percentage of ambiguous words in the word association test that had been presented earlier in the experiment as a less explicit measure of awareness, (word estimate *median* = 33.5, *range* = 3-65, skewed distribution). These estimate data were rescaled (divided by 100) and centred.

Model comparisons[8] revealed that neither the interaction between awareness of the experimental aim and subordinate priming, nor the interaction between the

[8] We included only the dominant unprimed trials in this analysis, excluding the dominant primed condition, as we were interested in awareness of subordinate meaning encounters only. Each awareness factor was included as a fixed factor in a logistic mixed effects model along with the fixed factor of subordinate priming, which indicated whether an item was unprimed or subordinate primed (i.e. this factor combined one and three repetition items as 'primed'). The random effects structure was constructed with subjects and items intercepts and slopes for subordinate priming. The interaction

awareness estimate and subordinate priming, was significant (X^2 (1) = 1.34, p = 0.248; X^2 (1) = 0.16, p = 0.686, respectively), indicating that participants' awareness of the priming manipulation and how many test words were repeated from the prime phase did not influence subordinate meaning priming effects.

Discussion

The aim of the present experiment was to investigate how multiple recent experiences with either the same or different meanings of an ambiguous word affect subsequent disambiguation. Just one encounter with the subordinate meaning of an ambiguous word was sufficient to retune lexical-semantic representations 30 minutes later, thus replicating previous findings (Rodd et al., 2016; Rodd et al., 2013). A single encounter with an ambiguous word in the context of its subordinate meaning resulted in a significant increase in the proportion of responses consistent with this meaning, compared to the unprimed baseline. The average dominance of the primed subordinate meanings increased from a baseline of 25% to 29%, showing that although these subordinate meanings are, on average, still less preferred than the alternative dominant meaning, they are more readily available following recent exposure. Although there was a numerical effect suggesting that aware participants showed a smaller subordinate priming effect, analyses showed that this was not significant. Whilst it is reassuring that awareness of priming did not significantly alter subordinate priming, Experiments 2 and 3 will follow up on these awareness analyses with larger sample sizes and therefore more power.

Whilst both the one and three massed subordinate repetition conditions significantly shifted disambiguation towards the subordinate meaning compared to baseline (*relative* increases of 16% and 24%, respectively), three massed subordinate

---

between the relevant awareness factor (aim or estimate) and subordinate priming factor was the crucial test, as a significant interaction would indicate that priming varied as a function of the awareness factor. As before, a model comparison approach was used to determine the significance of this interaction, where a model with both fixed effects and their interaction was compared to a model with both fixed effects without the interaction term.

repetitions did not provide a significant additional biasing effect over and above one repetition of the subordinate meaning. In contrast to the mechanism proposed by Rodd et al. (2013) whereby every encounter with an ambiguous word produces a similar change to connections strengths, the present experiment finds no evidence to support the notion that each encounter with an ambiguous word increases the availability of the primed meaning to the same extent, at least when these encounters occur within a single paragraph (i.e. massed presentation).

One encounter with the dominant meaning was also sufficient to retune representations. This finding contradicts the predictions of the literature (Rodd et al., 2013, Experiment 1, Fig. 1b), which suggests that there would be little effect of dominant priming since the dominant meaning is already the most available meaning and therefore cannot be made much more available. However, the delay between the dominant prime phase and test is markedly shorter than the delay between the subordinate prime phase and test, which could account for the dominant priming effect and makes it difficult to compare the magnitudes of dominant and subordinate meaning priming.

Importantly, as predicted, there was still an observable effect of prior subordinate meaning repetitions following the dominant repetition: there were significantly more subordinate meaning responses when a word was primed with the subordinate and then dominant meaning, compared to priming the dominant meaning alone. In other words, prior subordinate priming has a residual effect that persists after exposure to the dominant meaning. Interestingly, one subordinate exposure followed by one dominant exposure was comparable to the unprimed baseline condition, with the effects of the two “opposite direction” manipulations effectively cancelling each other out. Clearly, it is not the case that only the most recently activated meaning drives subsequent disambiguation. Instead, at least in the case where *different* meanings of a word are encountered with a substantial (23.5 minutes) gap between the encounters, disambiguation seems to reflect a cumulative effect of recent experiences.

In contrast to this cumulative effect for encounters with *different* meanings of a word, this experiment found no evidence that multiple recent encounters with the *same* (subordinate) meaning can produce a significantly greater biasing effect compared to just one encounter. This finding is surprising: multiple repetitions must at some level influence disambiguation over and above the effect of one repetition, otherwise there would be no effect of relative meaning frequencies on word interpretation, nor would there be an effect of an individual's long-term experience with word meanings, ranging from hours to years (Rodd et al., 2016). Why, then, in the present experiment did multiple repetitions not significantly boost availability of the subordinate meaning any more than one repetition?

One possibility is that, in the one repetition condition, the synonymous words that were used in place of the second and third repetitions caused participants to re-activate the initial ambiguous word such that the priming effect in the one repetition condition was artificially inflated. Any semantic priming resulting from synonymous words is not likely to persist at a 30-minute delay (Rodd et al., 2013), so this account would have to assume that the ambiguous word itself was covertly re-activated. Another possibility is that it is the massed presentation of the multiple repetitions within single paragraphs that could explain the absence of any additional priming boost, and perhaps spacing these repetitions would increase priming compared to the single exposure condition. Indeed, for the condition in which participants encountered the subordinate and then the dominant meaning (where there is evidence of cumulative effects of multiple encounters), these encounters were spaced. The repetition priming literature provides some evidence to suggest that spacing might indeed boost priming (Glenberg, 1976; Greene, 1989; Madigan, 1969; Thios, 1972; Underwood, 1970), although not necessarily (Paivio, 1974). More specifically, the natural language processing literature suggests a "One Sense per Discourse" principle (e.g. Gale, Church, & Yarowsky, 1992) where an ambiguous word appearing multiple times within a discourse has a high (up to 98%) chance of each repetition having the same meaning. As a result, within-discourse repetition is most likely to (overall) provide *one* piece of information about only *one* meaning regardless of how many repetitions are encountered and is therefore unlikely to be representative of a wider language context. This within-discourse repetition would be less informative for improving future interpretation than between-discourse repetitions, which have multiple different contexts and would therefore provide *multiple* pieces of evidence about *one* meaning. Hence one or three subordinate repetition(s) within the same discourse (i.e. paragraph) would not lead to different levels of priming. In light of these possibilities, we further investigated the nature of multiple repetitions in Experiment 2.

Experiment 2 – one & three spaced subordinate repetitions

This experiment used single sentence primes rather than paragraphs to allow for the temporal spacing of repetitions (as in Rodd et al., 2016, Experiment 2; Rodd et al., 2013). The prime phase was divided into three blocks in order to allow for the three repetitions of an ambiguous word (each in a different sentence) to be spaced across the prime phase (i.e. one repetition in each block). We compared the word-meaning priming effect between these three spaced repetitions with that of one repetition, where the ambiguous word was only encountered once in the prime phrase. To ensure that any benefit seen in the spaced repetition condition over the one repetition condition did not arise as a result of a primacy or recency effects (i.e. greater priming for words encountered either early or late in the experiment), two 'one repetition' conditions were included: an early repetition condition, where the ambiguous word appeared in the first block, and a late repetition condition, where the ambiguous word appeared in the third block. Unlike Experiment 1, we did not include a dominant meaning priming manipulation. Hence, the experiment had four conditions: unprimed baseline, one early repetition (block 1), one late repetition (block 3) and three spaced repetitions (one repetition in each of blocks 1, 2 and 3). This subordinate meaning prime phase was followed by a filler task, which created a prime-test delay, and then by a word association task, where participants heard all ambiguous words in isolation and responded with an associate. See Figure 3 for an overview of the procedure.

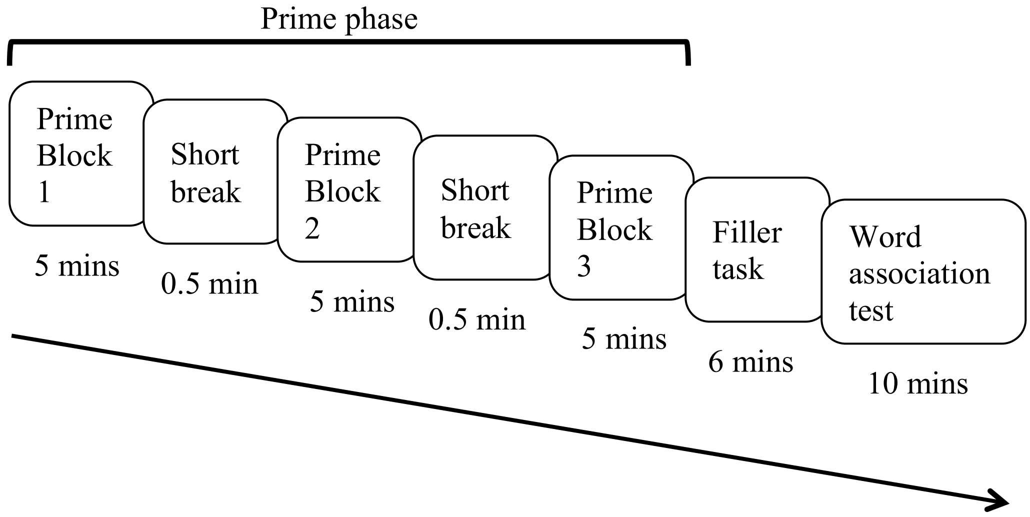

Figure 3. Experiment 2 task order, including prime phase elements, filler task and test, with the mean duration of each task.

Method

Participants

Sixty-four native British English speakers participated in the current experiment, although only the data from 55 participants (38 females; mean age = 21.5, range = 18 - 33) were analysed. The data from three participants did not save due to a technical issue and six participants were excluded for not meeting the eligibility requirements. All remaining participants met the requirements specified in Experiment 1 and were recruited in the same way but were paid the standard rate at the time of £8/hour.

Materials

The 88 ambiguous words were taken from Rodd et al. (2016, Experiment 2). These words were chosen to have a subordinate meaning that was semantically distinct from the dominant meaning (dominance range of the subordinate meanings = 0 - 0.48, mean = 0.24). Forty-nine (56%) of these ambiguous words had also been used in Experiment 1 (see Appendix B for full word list). As with Experiment 1, polysemous words were also included as long as the related meanings were judged by the author as sufficiently distinct that they could be distinguished on the basis of word association responses (this accounted for 50 words; Parks et al., 1998).

For the subordinate prime task, there were three sentences constructed for each of the 88 ambiguous words (mean length = 9 words; one sentence for each word was used in Rodd et al., 2016, Experiment 2). All three sentences disambiguated the word towards the same subordinate meaning but with different contextual details (see Table 2, below, for an example). This ensured that the multiple repetitions only primed the meaning of the word and not the entire sentence. Disambiguating context always preceded the ambiguous word so that upon encountering the homophone, only the intended subordinate meaning was appropriate. Each sentence was coupled with a probe word, which was either related or unrelated in meaning to the sentence (unrelated probes were not related to any meaning of the ambiguous word). The relatedness of probes was assigned at random to each sentence, although within each set of three sentences per ambiguous word, at least one probe was related and at least one was unrelated. Across the set of items, 50% of probe words were related. The target ambiguous words did not appear in any other sentences, instructions or other tasks, or as any of the probe words throughout the experiment. Sentences and probe words were presented in auditory form and spoken by a female native British English speaker with a Southern English accent (HNB).

Table 2. An example of the three sentences and probe words for the ambiguous word 'glasses' in Experiment 2.

Number	**Sentence (ambiguous word in italics)**	**Probe**
1.	The cupboard stored the mugs and *glasses*	Prefer (unrelated)
2.	She poured the champagne into the *glasses*	Fizz (related)
3.	The waiter set out the plates, cutlery and *glasses*	Table (related)

The 88 experimental ambiguous words were all included in the word association test, together with a further 20 unambiguous filler words, which were included to reduce the proportion of primed ambiguous words in the task with the aim of making the prime manipulation less salient. The first four ambiguous words in this task were filler ambiguous words, to allow participants to become accustomed to the task. All words were presented auditorily, in the same voice as the prime sentences. As with Experiment 1, a video animation ('Shaun the Sheep', Aardman Animations Ltd., 2010) was chosen as the filler task (see Experiment 1 for details).

Design

This experiment had a within-subjects design where all participants encountered all conditions but with a different set of items in each condition, so that each item appeared in every condition across participants. There was a single factor, subordinate prime repetitions, which had four levels: unprimed, one early repetition, one late repetition and three spaced repetitions. The dependent variable was the number of word association responses consistent with the primed subordinate meaning.

In the subordinate prime task there were three experimental blocks (see Figure 3). Participants encountered 22 ambiguous words in the first experimental block that were assigned to the one early repetition condition, 22 ambiguous words in the third experimental block that were in the one late repetition condition, and 22 ambiguous words in the three spaced repetition condition, which had one repetition in each of the three blocks. Participants therefore encountered 66 experimental sentences in total in the prime phase. To achieve an equal number of sentences in each block, 22 unambiguous fillers were added to block 2 for a total of 44 sentences per block. There were five additional unambiguous filler sentences presented at the start of each experimental block. Finally, 22 ambiguous words were assigned to the unprimed condition and thus were not encountered in the prime phase, but were presented in the word association test to provide an unprimed baseline.

Four versions of the experiment were created so that each ambiguous word appeared in each condition but for different participants, ensuring that participants saw each ambiguous word in only one condition. Thus, all ambiguous words and all participants contributed to all conditions. Within each version, three subversions were created, since there were three sentences for each ambiguous word but only one of which would be displayed in the one repetition conditions. In the multiple repetition condition, participants saw all three sentences for each ambiguous word, but the order of these three sentences varied across participants in different subversions. In the single repetition condition, across participants, a different sentence of the three was presented, rotated across subversions, to control for any potential differences between the three sentences.

Procedure

The experiment was presented using MATLAB (R2013b, 2013; version 8.2.0.701). All details regarding experiment set-up and preparation (e.g. demographics and instructions) were identical to Experiment 1 with the exceptions of a key press being required to proceed to the next screen or trial (as opposed to the mouse click in Experiment 1), and here the filler video was presented on the same screen as the other tasks (rather than via an iPad). See Figure 3 for a summary of the sequence and timings of the tasks.

Across all conditions there was an average delay of approximately 19 minutes between an ambiguous word in the subordinate meaning prime task and the same ambiguous word in the word association task. The average delays between an ambiguous word in block one and block three of the prime task and the same word in the word association task were 24.5 minutes and 13.5 minutes, respectively. Hence, there was an 11-minute average difference between the one early repetition and one late repetition conditions.

Subordinate Prime Task

Participants heard each sentence and, upon sentence offset, saw the probe word on-screen and were asked to respond as quickly and accurately as possible to the probe by either pressing the 'r' key for related or the 'u' key for unrelated. Response times longer than 3 seconds prompted a message encouraging faster responses on subsequent trials. The key press response triggered the next trial. There was a 30 second break for participants between each of the three experimental blocks. Five filler trials started each block, with the remaining items presented in a random order after the initial filler trials. The fillers at the start of each block were included to prevent the possibility that two of the spaced sentences for the same ambiguous word were encountered in close proximity (i.e. at the very end of one block and then at the very start of the subsequent block).

Filler Task

Video animation. See Experiment 1 for details.

Word Association Test

The procedure was the same as that used in Experiment 1, with the addition of a message encouraging faster responses on subsequent trials when the time to first key press exceeded 3 seconds.

Post-Experimental Tasks

The awareness questions were the same as those used in Experiment 1. Participant self-coding was not used in this experiment, or in Experiment 3, as the quality of participant coding in Experiment 1 was low and therefore required recoding by an experimenter (HNB).

Task Checks and Coding

All participants had at least 75% semantic relatedness accuracy, suggesting adequate engagement in the subordinate meaning prime task.

There were two coders (HNB and a research assistant) for the word association response data and coders were blind to the condition. Each word association response was coded either as being related to (1) the dominant meaning, (2) the primed subordinate meaning, (3) 'other', which included alternative meanings of the word, responses which were ambiguous/unclear and '0' responses (which participants were instructed to give if they could not think of a response or felt uncomfortable giving a response). For example, for the subordinate meaning of 'glasses' as in the sentence 'she poured the champagne into the glasses', the word association response 'eyes' would indicate the dominant meaning, whereas the response 'drink' would indicate the primed, subordinate meaning. Each experimenter coded half of the data. Any uncertainties were discussed with another researcher and if any doubt remained as to which meaning a participant intended, the response was coded as 'other'. For the analyses, 'other' responses (10%) were removed, as in Experiment 1.

Results

Main Analyses

As the subject means in Figure 4 indicate, relative to the unprimed condition, the proportion of subordinate responses increased following one repetition of the subordinate meaning, and increased again following three spaced repetitions.

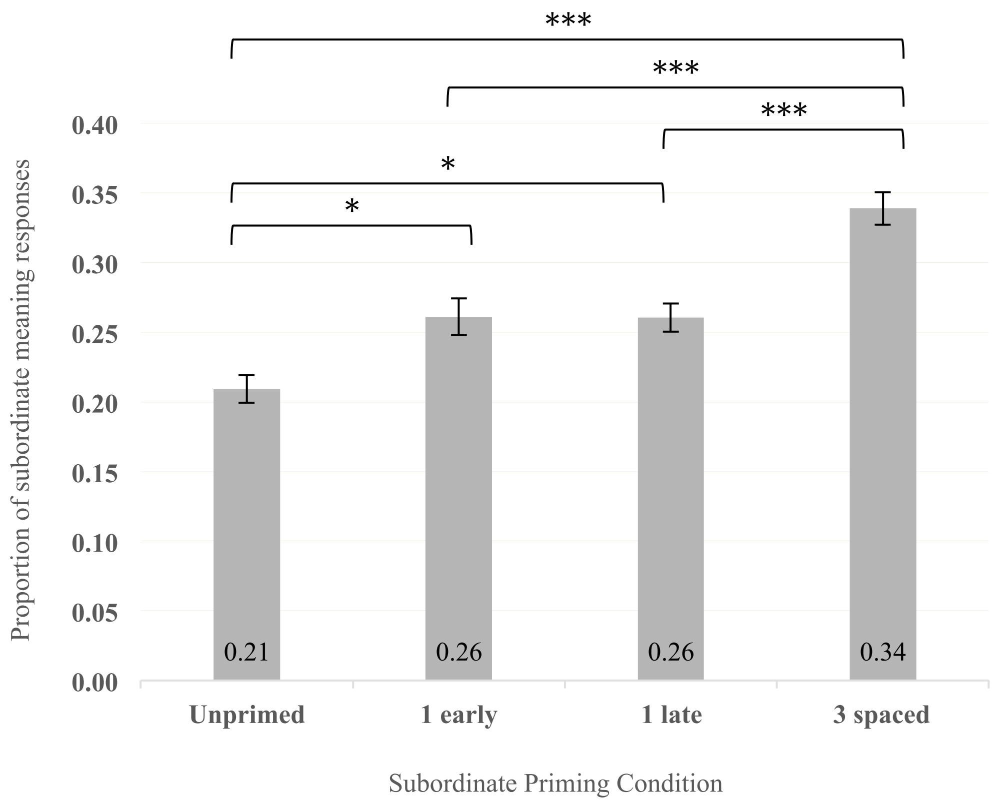

Figure 4. Experiment 2. Subject mean proportion of word association responses consistent with the primed subordinate meaning, with standard error bars adjusted for the within-subjects design and significance level indicated with asterisks (* < .05, *** <.001).

As with Experiment 1, a model with a maximal random effects structure was built with a fixed effect for subordinate meaning repetitions. The full model failed to converge. Following the recommended protocol for this issue (see Experiment 1 analyses for details; Barr et al., 2013), the correlations between the intercepts and slopes for subjects and items were removed, allowing the model to converge.

The model comparisons revealed a significant main effect of subordinate meaning repetitions (X^2 (3) = 69.60, $p < .001$), indicating that responses to ambiguous words varied as a function of the number of subordinate meaning repetitions encountered in the prime task. Pairwise comparisons with Tukey adjustment compared each level of the repetitions factor (unprimed baseline, one early repetition, one late repetition, three spaced repetitions) with one another (adjusted p values reported). Comparisons revealed significantly more subordinate prime-consistent responses following one early repetition (β = -0.38, SE = 0.15, z = -2.50, p = .050), and following one late repetition (β = -0.38, SE = 0.14, z = -2.70, p = .030), compared to the unprimed baseline. However, there was no significant difference between the single early and late repetitions (β = 0.002, SE = 0.13, z = 0.01, p = .990). Importantly, there were significantly more subordinate prime-consistent responses following three spaced repetitions than the one early repetition condition (β = 0.49, SE = 0.12, z = 4.06, p < .001), one late repetition condition (β = 0.49, SE = 0.12, z = 4.13, p < .001) and the unprimed baseline (β = -0.88, SE = 0.13, z = -6.71, p < .001).

Awareness Checks

The two awareness measures, awareness of experimental aim and awareness estimate, were analysed with logistic mixed effects modelling to investigate their effect on priming as outlined in Experiment 1. Two participants were removed due to missing data on the awareness test. Twenty-eight participants were unaware of the aim (priming effect across subordinate repetition conditions *mean* = .28, *SD* = .05) and 25 participants were fully/partially aware of the aim (priming effect *mean* = .30, *SD* = .05), where the word estimate gave an overall implicit measure of awareness (*median* = 60, *range* = 0-150, skewed distribution).

Model comparisons[9] revealed that neither the interaction between awareness of the experimental aim and subordinate priming, nor the interaction between the awareness estimate and subordinate priming, was significant (X^2 (1) = 1.34, p = 0.247; X^2 (1) = 0.002, p = 0.967, respectively), indicating that participants' awareness of the

[9] The logistic mixed effects models were constructed as in Experiment 1, again with the crucial test being the interaction between the relevant awareness factor (aim or estimate) and subordinate priming factor, as a significant interaction would indicate that priming varied as a function of the awareness factor. Whilst the models including the subjects and items slopes for subordinate priming failed to converge, the removal of these random effects allowed for convergence, leaving intercepts-only models for both of the following analyses.

priming manipulation and how many test words were repeated from the prime phase did not influence subordinate meaning priming effects.

Discussion

The aim of Experiment 2 was to investigate the impact of spacing repetitions of a word-meaning to see how multiple recent experiences with the same meaning affect how that word is later interpreted. First, the results indicate that just one encounter with the subordinate meaning of an ambiguous word can influence how that word is disambiguated approximately 19 minutes later. This word-meaning priming effect replicates the corresponding comparison from Experiment 1 (subordinate one repetition vs. subordinate unprimed, without dominant meaning priming) as well as previous findings (Rodd et al., 2016; Rodd et al., 2013). Moreover, awareness analyses supported findings from Experiment 1 that awareness does not significantly alter priming, although Experiment 2 showed a non-significant numerical increase in subordinate priming for aware participants rather than the non-significant numerical decrease seen in Experiment 1.

Second, the meaning priming effects for the early and late single repetition conditions did not significantly differ. The average time difference between these conditions was 10 minutes, hence a 24-minute prime-test delay for the early repetition condition and a 14-minute prime-test delay in the late repetition condition. This is consistent with previous findings: after a rapid decline during the first few minutes, word meaning-priming effects seem relatively stable across this time window (Rodd et al., 2016, Experiment 2). Whilst the prime-test delay for the late condition was less than the 19-minute delay used by Rodd et al. (2013), which showed that semantic priming did not persist, the similarity in priming effects from the early and late conditions is in contrast to what would be expected if the late condition were advantaged by semantic priming additional to word-meaning priming. Additionally, we would suggest semantic priming is unlikely given that semantic priming is generally short-lived, where an effect is considered ‘long-term’ if it survives a few minutes and intervening items (Becker et al., 1997).

Third, repeating the same subordinate word-meaning three times, spaced over the prime phase, increased the priming effect beyond that of one repetition. Compared to the unprimed baseline, one repetition provided a relative increase in the number of subordinate meaning preferences of 24%, whereas three spaced repetitions provided a more substantial relative increase of 62%. As there was no significant difference between the early and late one repetition conditions, it seems that there was no presence of a primacy or recency effect (from an encounter in the first or third prime block, respectively) and hence the benefit of spacing is not simply due to this condition consistently containing a prime in the first or last block, but is instead due to the multiple spaced repetitions themselves. This benefit of spaced repetitions shows that, at least in some cases, multiple individual encounters with an ambiguous word in a particular meaning context might further strengthen the relevant connections in the lexical-semantic network, producing a greater biasing effect over a single encounter (Rodd et al., 2013). This is consistent with the findings by Thios (1972) that spacing of repetitions improves task performance (recall of words in a sentence) compared to massed and single presentations.

Whilst the present findings suggest that the absence of a priming boost following three repetitions in Experiment 1 was due to their massed nature, these two experiments differ in several ways other than the spacing of the ambiguous words. Most notably this experiment used separate unrelated sentences and not connected paragraphs as in Experiment 1. Therefore, to be sure that it is the spacing of the ambiguous words that is key to determining the presence/absence of a boost in priming for multiple repetitions relative to one repetition, the three massed and three spaced repetition conditions need to be directly compared in the same experiment using the single sentence stimuli. Experiment 3 will therefore directly compare one repetition, three massed repetitions and three spaced repetitions in their word-meaning priming effects.

Experiment 3 – one, three massed & three spaced subordinate repetitions

This experiment includes four conditions: unprimed baseline, one repetition, three massed repetitions and three spaced repetitions. As in Experiment 2, the three spaced repetitions were spread across the three blocks of the prime phase, with one sentence per block. The three massed repetition sentences were presented as consecutive sentences within the same (randomly selected) block. The one repetition sentences were also distributed randomly across the three blocks. Since block position did not affect the magnitude of priming in Experiment 2, we did not counterbalance the block position in the one repetition condition. After the filler task, participants heard all ambiguous words in isolation and responded with an associate as a measure of their interpretation of the ambiguous word. See Figure 5 for an overview of the procedure.

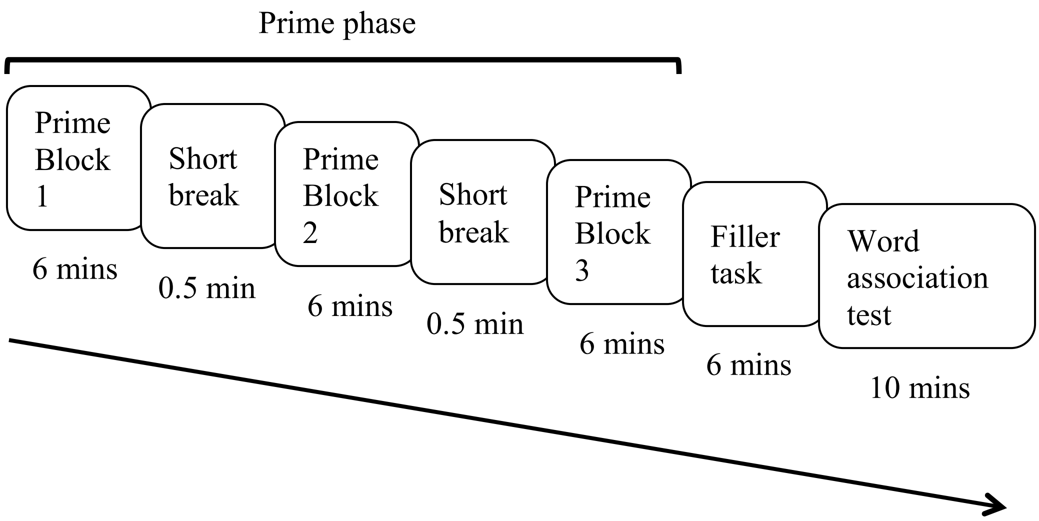

Figure 5. Experiment 3 task order, including prime phase elements, filler task and test, with the mean duration of each task.

Method

Participants

Sixty-one native British English speakers participated in the current experiment. Three participants were excluded for not meeting the eligibility requirements (see Experiment 1) and the remaining 58 participants (46 females; mean age = 20, range = 18 - 32) were entered into the analyses. All remaining participants met the requirements specified in Experiment 1 and were recruited in the same way but were paid the standard rate at the time of £8/hour.

Materials

See Experiment 2 Materials for details. The materials used in the current experiment are identical; only the design differed.

Design

In a within-subjects/between-item and within-item/between-subjects experimental design, the independent variable was the number of subordinate prime repetitions, which had four levels: unprimed, one repetition, three massed repetitions and three spaced repetitions. The dependent variable was the number of word association responses consistent with the primed subordinate meaning.

In each version, 22 of the total 88 ambiguous words were included in each of the four conditions. The 22 items in the one repetition condition and the 22 3-sentence sets in the massed repetition condition were distributed across the three experimental blocks (for each of these two conditions: 8 items in block 1, 7 items in block 2, 7 items in block 3), whereas for the 22 spaced repetition items, one sentence was allocated to each block. For each participant there were 22 ambiguous words that were not encountered in the prime phase but were included in the word association test to act as an unprimed baseline.

Four versions of the experiment were created so that each ambiguous word appeared in each condition but for different participants, ensuring that participants saw each ambiguous word in only one condition.

Procedure

The general procedure used in the current experiment is the same as in Experiment 2; only the design of the repetition differed. As the inclusion of the massed condition involved two additional sentences per item (compared to the single repetition conditions in Experiment 2), the prime phase was longer (timings shown in Figure 5): the average delay between prime and test encounters increased from 19 minutes in Experiment 2 to 21 minutes here.

The sets of three sentences that were presented in the massed and spaced conditions were always presented in the same order (the order of the three sentences was randomised following creation of the sentences). For the one repetition condition, one of the three sentences was randomly selected for each participant.

Task Checks and Coding

All participants had at least 75% accuracy on the semantic relatedness task, indicating adequate engagement in the prime task.

For the word association test responses the coding scheme was the same as for Experiment 2. One coder (a research assistant) completed all response coding, a subset of which was then verified by the second coder (HNB). Any uncertainties were discussed with another researcher and if any doubt remained as to which meaning a participant intended, the response was coded as 'other'. The item 'cold' was excluded from all analyses as there were too many responses coded as 'other' (28 out of 61), reflecting the fact that many common responses were indistinguishable between the 'temperature' and 'viral illness' meanings. For the analyses, 'other' responses (11%) were removed, as in Experiment 1.

Results

Main Analyses

As the subject means in Figure 6 indicate, the proportion of subordinate responses increased following both one repetition and three massed repetitions of the subordinate meaning, relative to the unprimed condition. There was a further increase following three spaced repetitions.

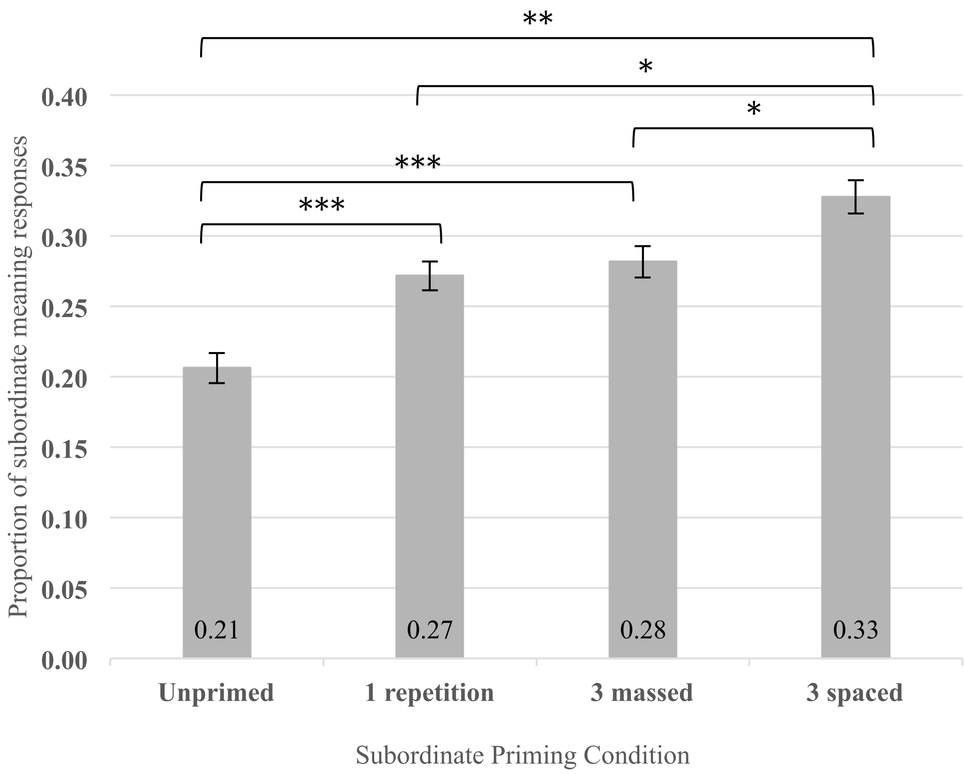

Figure 6. Experiment 3. Subject mean proportions of word association responses consistent with the primed subordinate meaning, with standard error bars adjusted for the within-subjects design and significance level indicated with asterisks (* < .05, ** <.01, *** <.001).

As with Experiments 1 and 2, a model with a maximal random effects structure was built (Barr et al., 2013) with a fixed effect for subordinate meaning repetitions and random effects for subjects and items. The full model failed to converge so the random effects structure was progressively simplified until the model converged, resulting in an intercepts-only random effects structure.

As with Experiment 2, a model comparison approach revealed a significant main effect of subordinate meaning repetitions, (X^2 (3) = 58.7, p < .001), indicating that responses to ambiguous words varied as a function of the number of subordinate meaning repetitions in the prime task. Pairwise comparisons with Tukey adjustment compared each level of the repetitions factor (unprimed baseline, one repetition, three massed repetitions, three spaced repetitions) with one another (adjusted p values reported). Comparisons revealed significantly more subordinate prime-consistent responses following one repetition compared to the unprimed baseline (β = -0.45, *SE* = 0.11, z = -4.23, p < .001). There were also significantly more subordinate responses following three massed repetitions compared to the unprimed baseline (β = -0.53, *SE* = 0.11, z = -4.96, p < .001), and no significant difference between the one repetition and three massed repetition conditions (β = -0.08, *SE* = 0.10, z = -0.80, p = .880). Critically, there were significantly more subordinate responses following three spaced repetitions compared to all other conditions: three massed repetitions (β = 0.26, *SE* = 0.09, z = 2.62, p = .040), one repetition (β = 0.34, *SE* = 0.10, z = 3.37, p = .004) and the unprimed baseline (β = -0.80, *SE* = 0.10, z = -7.53, p < .001).

Awareness Checks

The two awareness measures, awareness of experimental aim and awareness estimate, were prepared for logistic mixed effects modelling to investigate their effect on priming as outlined in Experiment 1. One participant was removed due to missing data on the awareness test. Thirty-one participants were unaware of the aim (priming effect *mean* = 0.27, *SD* = 0.05) and 29 participants were fully/partially aware of the aim (priming effect *mean* = 0.28, *SD* = 0.05), where the word estimate gave an overall implicit measure of awareness (*median* = 50, *range* = 1-100, skewed distribution).

Model comparisons[10] revealed that neither the interaction between awareness of the experimental aim and subordinate priming, nor the interaction between the awareness estimate and subordinate priming, was significant (X^2 (1) = 0.01, p = 0.923; X^2 (1) = 1.15, p = 0.282, respectively), indicating that participants' awareness of the priming manipulation and how many test words were from the prime phase did not influence subordinate meaning priming effects.

Discussion

The aim of the present experiment was to investigate the impact of spacing the priming encounters to see how recent experiences with a particular meaning of an ambiguous word affect subsequent disambiguation. As with Experiments 1 and 2, just one encounter with the subordinate meaning of an ambiguous word influenced how that word is disambiguated approximately 21 minutes later: there was a 29% relative increase in the proportion of subordinate responses from the unprimed to the one repetition condition, thus replicating the word-meaning priming effect (Rodd et al., 2016; Rodd et al., 2013). Moreover, awareness analyses supported findings from Experiments 1 and 2 that awareness does not significantly alter priming, with a small numerical effect consistent with Experiment 2 suggesting a non-significant *increase* in subordinate priming for aware participants.

As in Experiment 1, the magnitude of the word-meaning priming effect did not significantly increase following three massed presentations of sentences with the subordinate meaning compared to the condition with only one priming sentence. In contrast, priming did significantly increase when the three sentence presentations were spaced, resulting in a sizeable 22% *relative* increase compared with the one repetition condition. Critically, spaced repetitions also significantly increased the priming effect compared to massed repetitions with the same number of sentences (an

[10] The logistic mixed effects models were identical to those in Experiment 2 (intercepts-only random effects structures due to convergence failure when slopes for priming were included). As with Experiments 1 and 2, the crucial test was the interaction between the relevant awareness factor (aim or estimate) and subordinate priming factor, as a significant interaction would indicate that priming varied as a function of the awareness factor.

18% relative increase). It seems that when multiple repetitions occur in quick succession they act similarly to a single instance, and it is not until those repetitions are separated that there is an additional effect of multiple encounters with the word and its subordinate meaning. Hence, it seems that the spacing of experiences with ambiguous words is key to producing greater alterations to the lexical-semantic network than that of one experience.

General Discussion

The aim of the experiments in this chapter was to explore how listeners update their lexical-semantic knowledge on the basis of recent experience. Specifically, using a contextual prime and word association test paradigm, three experiments investigated how single and multiple experiences with ambiguous word-meanings influence the later interpretation of these words in isolation. The results can be grouped into three main findings.

Effects of single subordinate and dominant encounters

All three experiments show that a single encounter with a subordinate word-meaning was sufficient to bias how that word was interpreted when presented in isolation after a 20-30 minute delay. These findings replicate four experiments from the literature (Rodd et al., 2016, Experiments 1 & 2; Rodd et al., 2013, Experiments 1 & 3), providing a total of 7 experiments that have consistently shown this robust word-meaning priming effect within the subordinate prime/word association test paradigm. These experiments also replicate the finding that participants' awareness of the experimental aims is not a critical factor for priming to occur. In all three experiments, there was no significant interaction between the magnitude of priming and participants' awareness of the experimental manipulation. Further, the numerical effects of awareness on priming were inconsistent across experiments: while in Experiments 1 and 2 we observed (non-significantly) more priming for the 'unaware' participants compared with the 'aware' participants, for Experiment 3 we observed the reverse (non-significant) effect. This suggests that the word-meaning priming observed in this paradigm is *not* driven by conscious attempts to recall previous sentences.

Experiment 1 goes beyond this replication; while previous studies of word-meaning priming have focused on the situation where participants are primed with the subordinate (less frequent) meaning, we observed, for the first time, a significant effect of prior experience with the word's *dominant* meaning. Although the dominant prime-test delay was shorter than the subordinate prime-test delay (by approximately 15 minutes), this finding suggests that even when the meaning of an ambiguous word is encountered that is already (on average) preferred by participants, it is still possible to boost its availability. As a result of the different prime-test delays, the size of the dominant and subordinate meaning priming effects cannot be directly compared, although Rodd et al. (2013) provide evidence that larger priming effects can be seen for the more highly subordinate meanings, indicating that the initial dominance of the primed meaning may indeed moderate the magnitude of priming.

These subordinate and dominant priming findings are consistent with our current view of lexical-semantic representations (Rodd et al., 2016; Rodd et al., 2013), which suggests that the mechanism for updating word-meaning representations involves changes to connection strengths among units in a connectionist network (Rodd et al., 2004). According to this view, each individual encounter with a word-meaning strengthens the relevant connections in proportion to the overall frequency with which each meaning is encountered. This theoretical view would therefore predict that an encounter with either the subordinate or the dominant meaning would alter the connection strengths related to the representation of the word's subordinate or dominant meaning, respectively, increasing the availability of the relevant meaning representation so that when the word is later encountered in isolation, there is a relatively greater bias toward interpreting the word with this same meaning. In other words, Experiment 1 shows that lexical-semantic representations are sensitive to a single meaning encounter regardless of the initial availability of the meaning itself (i.e. whether it is the dominant or subordinate meaning). This is consistent with our view that lexical-semantic representations are dynamic even in adults, such that they flexibly adapt to reflect the up-to-date likelihood of occurrence in order to maintain efficient processing of ambiguous words.

Cumulative effects of multiple encounters

Experiments 2 and 3 go beyond previous findings in showing that repeated word-meaning encounters within a relatively short period of time (e.g. 20-30 minutes) can lead to cumulative effects in updating the representations of word-meanings similar to those shown in the literature (Rodd et al., 2016) with longer-term (e.g. days/months/years) cumulative effects from experience with ambiguous words. Both Experiments 2 and 3 showed that three spaced encounters of the same subordinate word-meaning biased the later interpretation of that word (in isolation) towards that subordinate meaning over a single encounter. The impact of three spaced repetitions was not threefold the magnitude of one repetition: this is consistent with an asymptotic nature of repetition effects found in the repetition priming field, such as with a lexical decision task (Logan, 1990). This finding is consistent with previous accounts of word-meaning priming and the view that the effect of experience is cumulative. In contrast, it rules out an account of word-meaning priming in which only the most recent encounter is critical in determining the accessibility of word-meanings. This latter view predicts that there would be no difference between the one and three spaced conditions, as they both involved the same single sentence encounter with the subordinate meaning as the most recent encounter of the word. However, this was not the case; three spaced subordinate repetitions made participants more likely to retrieve the subordinate meaning at test. Thus it is not only the most recent encounter that affected word interpretation, it is the effect of multiple recent encounters of the same meaning that accumulate to produce an additional influence on later interpretation.

Furthermore, Experiment 1 showed a residual effect of the initial subordinate meaning even after a subsequent encounter with the dominant meaning; there were more subordinate responses when the subordinate prime had preceded the dominant prime than when the dominant prime had been presented alone. Again, if only the most recent encounter were critical, the subordinate plus dominant condition and the dominant only condition would show equal priming, as they both involve the same dominant prime sentence being encountered most recently. As the former condition resulted in more subordinate responses than the latter, we can conclude that the dominant meaning does not completely ‘cancel out’ the earlier subordinate encounter, rather the effect of the recent dominant encounter in fact *adds* to the effect of the earlier subordinate encounter. Once more, it is the cumulative effect of multiple recent encounters of different meanings that combine to influence interpretation.

In summary, these data provide clear evidence that multiple encounters with ambiguous words can, when spaced throughout the prime phase, have a cumulative effect on how these words are interpreted in the future. We have now shown that for repeated encounters with the *same* meaning (Experiments 2, 3) and for repeated encounters with *different* meanings (Experiment 1), subsequent interpretation is not driven solely by the individual's most recent encounter with that word. These data can only be explained by assuming that recent experience with word meanings can accumulate across multiple exposures, such that earlier experience with the word meanings is not fully overwritten by the most recent encounter. This aspect of the data is fully consistent with the mechanism put forward by Rodd et al. (2013) to explain how lexical-semantic representations update. The proposed mechanism involves changes to connection strengths among units in a connectionist network, which would allow transient changes in meaning availability to accumulate slowly across the lifespan based on each individual experience with a word. These changes appear to reflect a build-up of evidence about the relative likelihoods of different word-meanings across a wide range of timescales. In this view, lexical-semantic representations subtly but continually update based on experience with word meanings, so that these representations adapt dynamically to the listener's environment. This view is consistent with the finding that rowers show a long-term preference for rowing-related meanings that increased for those rowers with more years of rowing experience (Rodd et al., 2016).

Whilst the present findings are lab-based, Rodd et al. (2016) revealed two findings indicating the real-world generalisability of updating meaning representations. First, rowers' long-term experience with specific meanings generalised to non-rowing settings (they were not informed that it was a rowing-related experiment and the experiment was not performed in a rowing environment). Second the radio study shows that the word-meaning priming paradigm was also successful outside of the lab, as participants heard the prime sentences over a radio show, later finished the experiment in their own time and place (i.e. not in a lab setting) and were not aware that the test was in fact linked to the radio prime phase.

Taken together with these earlier findings, the present results suggest that repeated encounters with a word-meaning gradually strengthen the relevant connections in the lexical-semantic network, which can change an individual's meaning dominance both in the shorter-term (present experiments) and longer-term (Rodd et al., 2016).

Benefit for spaced over massed repetitions

Experiments 2 and 3 demonstrated that when three subordinate meaning repetitions were presented in a spaced manner (i.e. with a 5-minute delay between each), this produced significantly more priming than when only one repetition had been presented. Moreover, Experiment 3 demonstrated that these three spaced repetitions also produced significantly more priming than three massed repetitions (i.e. each repetition presented in succession). It seems that when repetitions were massed, they did not bias responses towards the subordinate meaning any more than one repetition (Experiments 1, 3). Unlike the more general effect of repeated exposures discussed above, this specific spacing (over massed) benefit was not predicted by our current mechanism for updating meaning representations (Rodd et al., 2013). For decades, practice and spacing benefits for memory have been studied using a variety of different paradigms (Karpicke & Bauernschmidt, 2011; Madigan, 1969; Melton, 1970), yet there has been little agreement on the mechanism underlying these spacing effects (Delaney, Spirgel, & Toppino, 2012; Gotts, Chow, & Martin, 2012; Pavlik & Anderson, 2005; Raaijmakers, 2003; Shea, Lai, Black, & Park, 2000). Thus the specific mechanism for the spacing advantage here, as in other memory and learning paradigms, is an ongoing area of debate that warrants future investigation. Furthermore, the word association test used here reflects the ultimate outcome of multiple processes involved in word interpretation, including word recognition, meaning access, and word associate retrieval. Consequently we cannot draw a strong conclusion about which process(es) are affected by the spacing of prior exposures to word meanings, and other measures of word-meaning priming might yield different results.

Previous accounts of word-meaning priming do not provide an explanation for why the extra learning from additional repetitions should be impeded when the temporal spacing *between* repetitions is removed. There are two logical possibilities for why the additional massed repetitions do not contribute to learning. One possibility is that learning is primarily driven by the first of the massed repetitions, but is absent (or significantly reduced) for subsequent massed presentations. Alternatively, learning may be driven (primarily) by the most recent of massed repetitions and, for some reason, this final encounter reduces the extent to which the listener learns from the previous massed encounters. Knowing which of these possibilities drives the lack of a massed repetition benefit would help to elucidate the mechanism underlying the updating of meaning representations.

One example of a class of model in which listeners benefit primarily from the first of multiple massed encounters is the activation account (Pavlik & Anderson, 2005, 2008). This model suggests that with each encounter of an item, activation strength increases, but this increase decays as a power function of time. The *rate* of decay is greater when activation is higher, such that the benefit from highly active items will decay faster than for less active items. Hence, providing space between repetitions means that activation has time to decrease between each repetition, thus the rate of decay is slow and the benefit of repetitions lasts longer. Without this spacing between repetitions, as in the massed repetition case, there is not enough time for activation to decrease. This higher initial activation therefore means that the rate of decay is relatively fast and the benefit of massed repetitions does not last as long as for spaced repetitions. This notion is similar to that of a refractory period, where, post repetition there is a period during which activation cannot be further increased by (i.e. is unresponsive to) further repetitions (e.g. Hintzman, Block, & Summers, 1973; Welford, 1952).

In contrast, the consolidation account is an example of a class of model in which individuals learn primarily from the most recent of multiple massed encounters (e.g. Landauer, 1969; and specifically relevant to the present consolidation explanation, proposed for motor skill learning, Shea et al., 2000). This view suggests that memory formation is an ongoing consolidation process following the presentation of a stimulus that can result in transfer from short- to long-term memory, which is more resistant to forgetting and interference (e.g. Brashers-Krug, Shadmehr, & Bizzi, 1996). However, if this consolidation process is interrupted, then the long-term memory does not form properly, or indeed at all. Thus interruption of consolidation (even by a new encounter with the same stimulus, Shadmehr & Brashers-Krug, 1997) could reduce or prevent learning. Applying this to word-meanings, with three massed repetitions, the memory trace for the first repetition would start consolidation after presentation but this process would be interrupted by the presentation of the *same* word-meaning just seconds later. As the third repetition is the final encounter, this word-meaning would have more uninterrupted time for consolidation, although it is the only repetition out of the three to consolidate fully, making the massed condition similar to the one repetition condition in terms of consolidation. In contrast, spaced repetitions would show a priming benefit in this account because it allows sufficient time between repetitions for the word-meaning to be (partially) consolidated after each encounter.

Finally, in contrast to these two views, which both assume that it is the timing of the events that drives the observed spacing effect, we must consider an alternative view that this effect is instead driven by differences in contextual variation between massed and spaced exposures. This account proposes that spacing benefits can be explained by an encoding variability mechanism (Maddox, 2016). According to Mensink and Raaijmakers (1989) and Raaijmakers (2003), the general context surrounding a stimulus naturally fluctuates over time and this context is encoded with each presentation of a stimulus. As the temporal spacing of repetitions gets longer, the natural context is more likely to vary and that variation between stimulus encodings increases the likelihood/magnitude of learning from that stimulus. Hence, this account would suggest that the spacing benefit arises due to the increase in different encoded contexts for the spaced word-meaning exposures, which would subsequently make the meaning more available. This model is akin to the concept of contextual diversity (Adelman, Brown, & Quesada, 2006; van Heuven, Mandera, Keuleers, & Brysbaert, 2014), which has been shown to affect word processing (lexical decision performance is better explained by contextual diversity across word occurrences than by just the frequency of occurrence). Similarly, the "One Sense per

Discourse” principle (e.g. Gale et al., 1992) is based on the finding that an ambiguous word encountered multiple times within a discourse is highly likely to be used in the same meaning across those encounters, and suggests that an interlocutor would treat one subordinate repetition and three subordinate repetitions within the same discourse/paragraph as equivalents because they both provide *one* overall piece of evidence about *one* meaning (as opposed to *multiple* separate/spaced pieces of evidence of that one meaning).

However, this encoding variability/contextual diversity/ “One Sense per Discourse” type of account is less likely to provide an explanation for the current data. Although this account can explain the observed boost for spaced presentations compared with massed presentations, it cannot explain why three massed repetitions did not boost priming compared to one repetition, given that in Experiments 2 and 3 its two additional repetitions were presented in three separate sentences that did not link together into a coherent discourse. Even in the massed condition, these three sentences provided different contextual information and were distinctly presented in separate pieces of discourse (each sentence was followed by the judgment of relatedness of a probe word, and the sentences were unrelated) so this should provide enough contextual variation to see an increase in priming (compared to one repetition) even for the massed condition and even though the overall situational context did not vary a great deal. Yet, the massed condition provided no additional priming compared to one repetition, despite its two *additional and distinct* sentences/discourses of varying contextual information. Whilst contextual variation accounts consider the general surrounding context rather than context within the sentence, it seems unlikely that additional sentential context would not boost priming if context were such an integral factor in priming. This makes the contextual variation account an unlikely explanation for the present findings. Clearly, it seems that there are several possible mechanisms underlying the spacing benefit but, as aforementioned, this requires further research to disentangle.

Importantly, the observed lack of benefit for multiple massed repetitions is likely to be advantageous from a communication point of view, as these instances are not always representative of the broader word usage. For instance, a conversation with a tree surgeon might involve the tree meaning of 'bark' multiple times in a short passage/time-frame of perhaps minutes. If meaning preferences updated cumulatively with each of these repetitions, then this conversation alone would have a disproportionately large effect on meaning preferences for 'bark' compared to hearing the same number of 'tree bark' repetitions over a longer time-frame of perhaps days or weeks. In this case, the overly sensitive change in meaning preferences would be inefficient. In contrast, if additional word-meaning repetitions only alter representations when sufficiently spaced, lexical-semantic representations might still be somewhat sensitive to the listener's immediate environment but would primarily reflect the listener's long-term, temporally-distributed (spaced) experience with word usage, which are more likely to accurately predict how these words are used in the future. Under this account, exposure to multiple instances of a word used with its low-frequency meaning would produce a smaller biasing effect on its lexical-semantic representation, and thus this representation would more likely generalise to future encounters.

Conclusions

Adults' lexical-semantic representations are updated dynamically in response to on-going experience in order to reflect the most likely meaning of words. The present studies investigated the changes that occur as a consequence of exposure to the meanings of an ambiguous word. The results replicate the word-meaning priming effect and go further in showing that multiple subordinate repetitions provided an additional boost to priming compared to one repetition when these encounters were spaced, although this boost was eliminated when multiple repetitions were massed, at least in a word association test. Moreover, one repetition of the dominant meaning reduced, but did not eliminate, the effect of prior subordinate meaning priming. These results indicate that the experience-based changes to lexical-semantic representations are not solely based on the most recent encounter with a word meaning, nor does the effect occur with the same magnitude across repeated encounters. Rather, word-meaning interpretation appears to reflect the accumulation of recent experiences with word-meanings, where the temporal spacing of multiple encounters is key to producing additional learning effects. This seems to provide a balance among the influences of word usage patterns across a range of timescales, such that listeners can dynamically retune and update their lexical-semantic representations in response to recent experience while maintaining their longer-term knowledge of word-meaning dominance.

Chapter 3: Validation of picture-based test methodology

Introduction

Semantic ambiguity is ubiquitous in language, with over 80% of English words having multiple dictionary entries (e.g. '*bark* of the dog/tree'; Rodd et al., 2004). It is also arguably a useful and interesting component of language, where new words need not be invented for new concepts; existing word-meanings can be creatively extended to accommodate new concepts (Srinivasan & Rabagliati, 2015). Understanding semantic ambiguity resolution is therefore a critical component of any language comprehension model, and a large proportion of the literature has focused on how meaning dominance (the prevalence of each of a word's individual meanings) affects comprehension both in the presence and absence of context (e.g. Duffy et al., 1988; Foss, 1970; Rayner & Duffy, 1986; Seidenberg et al., 1982; Swinney, 1979). Due to limitations in the existing measures of dominance effects on comprehension, the present chapter provides a semantic relatedness picture test as an alternative. The development of the newly-developed picture stimulus set will be outlined, as well as the validation of their use in the semantic relatedness task. Since these pictures are also ideal for use in a wide range of language experiments, dominance norms and information on the ambiguous words and picture stimuli will be provided for use by other researchers.

A considerable number of experiments on semantic ambiguity have used word association to measure how meaning availability influences comprehension (e.g. Geis & Winograd, 1974; Rodd et al., 2016; Rodd et al., 2013; Twilley et al., 1994), where the ambiguous word is presented to participants who then provide an associate (i.e. interpretation) of that word. Word association is suitable for answering a range of questions regarding comprehension. For instance, it can measure whether experience with a particular word-meaning biases the later interpretation of that word (Rodd et al., 2016; Rodd et al., 2013). Here, the benefit of word association is that context is recently experienced but is not present at test, which measures whether or not comprehenders *learn* from this recent experience with language to guide their subsequent understanding.

There are, however, several limitations of word association as a measure of how meaning availability affects comprehension. Firstly, as noted by Cai et al. (2017), word association is a relatively slow, offline task measuring the end-point of comprehension. This means that any influence of, for instance, recent experience on word interpretation could occur either during or after meaning access. If the effect of experience occurs during meaning access, then experience must alter the pattern of activation of alternative word-meanings autonomously, making the recently encountered meaning more active and therefore more available for selection. Alternatively, if the effect of experience occurs after meaning access, then experience does not alter the pattern of activation of alternative word-meanings, but the listener could subsequently use the experience to select the recently encountered meaning in a strategic manner. Since word association measures comprehension *after* any/all of these processes have occurred, it cannot distinguish between these possibilities. Only by measuring the speed of a response can we determine whether or not such recent experience effects occur during or after meaning access. This is a significant limitation of word association; being able to determine the cause of any such effect is crucial, as the difference between these alternative processes is a fundamental element of developing any comprehension model.

Furthermore, since a word association response is the result of a completed disambiguation process guided by multiple meaning availabilities, the response is necessarily the combined effect of separate underlying dominant and subordinate meaning availabilities. Word association measures the relative availability of the different meanings, rather than the absolute availability of each meaning separately. For example, following a subordinate priming manipulation (e.g. 'the woodpecker clung onto the *bark*'), a 10% boost in subordinate meaning availability would appear the same as a 5% boost in subordinate meaning availability plus a 5% reduction in dominant meaning availability, as, in both cases, there would be a 10% change between meaning availabilities. Two different effects from the same manipulation would have very different effects on underlying meaning representations, and word association cannot show these theoretically interesting differences. Whilst this test may be suitable for answering some research questions (e.g. whether or not priming can bias comprehension in general), it is not suitable if the aim is to investigate changes to different underlying meaning availabilities (e.g. whether word-meaning priming can boost availability of one particular meaning and reduce the availability of a another meaning).

Finally, the test requires a participant to generate an associate for each ambiguous word, therefore providing only one data point per ambiguous word at test. For example, in response to hearing the ambiguous word 'bark', a participant might respond with 'tree'. For data analysis, each response must be coded as either related to the word's dominant or subordinate meaning. Since participants are not always successful or consistent in self-coding responses after the word association test (Experiment 1, Chapter 2), there is no alternative but for the experimenter to code the responses. The 'bark – tree' example would be relatively straightforward for the researcher to code as a subordinate response (albeit somewhat time-consuming). However, for an ambiguous word such as 'sink', participants might often respond with 'water'. This meaning is impossible to categorise as either the dominant ('to become submerged') meaning or the subordinate ('water basin') meaning of 'sink' because it relates to both meanings. These responses must therefore be excluded from analyses, narrowing down the pool of potential ambiguous word stimuli. Together, these issues with word association limit the power of experiments using this method.

An existing alternative measure of assessing dominance effects on comprehension is reading times, using eye tracking. This method is suited to assessing online effects of processes relating to ambiguous word comprehension. However, measuring reading times requires the hardware and software for eye tracking, which is not always readily available. Moreover, experiments using this method (and even visual world paradigms using eye tracking) must be carried out in a laboratory setting, despite the increasing popularity of online experiments (e.g. Goodman, Cryder, & Cheema, 2013; Litman, Robinson, & Abberbock, 2017), which allow for easier recruitment and data collection, which is faster and more time-efficient for a researcher, without a significant compromise in the quality of data (Casler, Bickel, & Hackett, 2013).

Recently, Armstrong, Tokowicz, and Plaut (2012) provided an alternative measure of meaning availability. Their 'eDom' task and application (in MATLAB) is based on explicit ratings of the relative frequencies of ambiguous word's dictionary definitions. Participants are provided with multiple possible meanings of an ambiguous word and must rate, as a percentage, the frequency with which they encounter each meaning in everyday life. Armstrong et al. (2012) suggest that their eDom software is a method for measuring dominance norms, and can be used as a means of selecting suitable ambiguous word stimuli for use in language studies. The authors provide evidence to suggest that the eDom method is superior to word association for two reasons. First, they argue that this method is more reliable, since ratings were highly consistent across participants and items (to a similar level of a measure in which participants rated the age of acquisition of each meaning of an ambiguous word; Khanna & Cortese, 2011). Second, they argue that eDom is more efficient than word association, as it requires fewer observations per ambiguous word. Standard norming studies using word association have used approximately 100 participants to generate norms for 100 words (e.g. Twilley et al., 1994) yet, with eDom, Armstrong et al. (2012) suggest that only 16 participants are required to generate norms for 146 words. However, it is likely that eDom is restricted to stimuli selection, as its explicit nature would leave it prone to demand characteristics if an experimental manipulation were involved.

It is important to emphasise that, whilst these drawbacks should not prevent the use of these methods, they do show that the testing method must be carefully selected based on the design of the experiment and the research question. Whilst word association and eye tracking have been, and continue to be, very insightful in many experiments, the limitations mean that an alternative method might provide additional insights into the way in which ambiguous words are interpreted. In this chapter, an alternative test for effects of meaning availability on comprehension is provided: a novel semantic relatedness task using pictures. The semantic relatedness task has been successfully used in a range of experiments in the field of language (e.g.

Blumstein, Milberg, & Shrier, 1982; Gilbert et al., 2018; Stringaris, Medford, Giora, Giampietro, Brammer, & David, 2006; Zwaan & Yaxley, 2003). The clear benefit of semantic relatedness over word association is that responses (reaction times and accuracy) are measured earlier in the time-course of processing compared to word association (Cai et al., 2017). By measuring responses earlier in the process, semantic relatedness can provide a measure of online processing and whether or not particular experimental manipulations have an autonomous (during meaning access) effect on comprehension.

A second benefit is that in semantic relatedness the meanings are probed independently, allowing for availability of the dominant and subordinate meaning to be measured separately. After a subordinate priming manipulation (e.g. 'the woodpecker clung onto the *bark*'), reaction time and accuracy to both the subordinate (tree bark) and dominant (dog bark) meaning pictures will be tested. If subordinate priming is driven by a boost in subordinate meaning availability, responses to the subordinate 'tree' picture are likely to be faster and/or more accurate after hearing 'the woodpecker clung onto the *bark*', compared to the unprimed baseline. However, if subordinate priming is (also) driven by a loss of dominant (unprimed) meaning availability, responses to the dominant 'dog' picture are likely to be slower and/or less accurate after hearing 'the woodpecker clung onto the *bark*', compared to the unprimed baseline (see Chapter 4). Hence, semantic relatedness can separate these underlying effects where word association cannot.

Typically, standard sematic relatedness tests present participants with a word and, on its offset, they must decide whether or not a second word probe is related to it. With this method, the relatedness of the word probes can vary in the degree of relatedness and across different categories of relatedness. For instance, for the trial 'tiger', the related probe could be 'lion' (closely related in the category of 'big cat'), or the related probe could be 'dog' (arguably less closely related but also in the 'animal' category), or the related probe could be 'jungle' (related in the category of 'habitat'). Clearly, just these three probes vary greatly in their degree of and categories/types of relatedness. Across many items in an experiment, this variation could add a great deal of extra complexity. Although words can be, and are frequently, used as probes successfully (e.g. Cai et al., 2017; Gilbert et al., 2018), using pictures as relatedness probes eradicates this extra complexity, since each picture is the visual referent of the meaning of the word. For instance, for 'bark', the dominant probe is a picture of a dog barking and the subordinate probe is a picture of the covering of a tree. The picture probes therefore add a third benefit of this novel semantic relatedness task, both compared to the standard word-probe semantic relatedness method and the word association test. Finally, as with word association, this semantic relatedness test can be easily programmed and deployed in an online experiment, making it time-efficient for a researcher.

Due to the multitude of benefits of the picture probe semantic relatedness task, the present chapter involves the design, development and use of a set of novel picture stimuli. It is important to point out that the use of these pictures is not limited to this method. In fact, this novel picture stimulus set could be used in a variety of methods. For example, the pictures could be used in visual world experiments, where looks to the pictured referent reveal the time-course of disambiguation. Or, the pictures could be used in semantic priming experiments, in which the effect of priming the meaning of an ambiguous rather than the word itself could be investigated. Alternatively, the pictures would be suitable for experiments on negative priming, which often show two pictures but one must be ignored, or even for masked priming experiments. Clearly, these pictures, applied to different methods, can be used in experiments for a range of research questions. For the present chapter, however, they will be used for the semantic relatedness picture task.

The present chapter therefore has three main aims. The first aim is to develop the pictures (including the pretesting of these pictures for quality). The second aim is to validate the picture stimuli for use with the semantic relatedness task by confirming that the task can detect differences (in reaction times and/or accuracy) between picture probes of the dominant and subordinate meanings of ambiguous words. If the task is sensitive to dominance, then it is potentially a suitable alternative to the standard word association method for measuring the availability of word-meanings. After confirming that semantic relatedness can detect dominance, the third aim is to derive dominance norms, from two different measures, on the pictured meanings (dominant and subordinate) in the stimulus set. By collecting word association and eDom norms for each picture, the relationship between word association, eDom, picture quality (from pretest ratings) and semantic relatedness performance will be investigated: whether word association and eDom scores predict RTs and/or error rates in the semantic relatedness picture task. In doing so, these additional measures will provide information on the pictures that can be used by other researchers in language experiments, providing a dominance baseline against which other experiment results can be compared.

Development of picture stimulus set

According to Nishimoto, Miyawaki, Ueda, Une, and Takahashi (2005), psychologists are increasingly using picture stimuli in a range of language experiments (e.g. Carroll & Snowling, 2004; Shook & Marian, 2012; Zwaan, Stanfield, & Yaxley, 2002). However, picture resources are currently limited, particularly for experiments on semantic ambiguity, which often require pictures of both the dominant and subordinate meaning of each ambiguous word. There is currently no source of suitable-quality stimuli that depict the dominant and subordinate meanings of a large enough sample of ambiguous words.

Whilst there are some existing resources specifically picturing ambiguous word-meanings (e.g. Duñabeitia, Crepaldi, Meyer, New, Pliatsikas, Smolka, & Brysbaert, 2018; Nishimoto et al., 2005; Nishimoto, Ueda, Miyawaki, Une, & Takahashi, 2012; Snodgrass & Vanderwart, 1980), the number or quality of the pictures is inadequate. The Snodgrass and Vanderwart (1980) set of 260 normed line drawings is a large set of ambiguous word-meaning pictures, although they were created almost four decades ago, hence the pictures lack the high resolution of more modern standards. Whilst Nishimoto et al. (2005) present a set of 359 normed ambiguous word-meaning pictures that are superior in quality, they are designed for Japanese rather than English. And whilst Duñabeitia et al. (2018) provide a large set of 750 pictures, the number of semantically ambiguous items is limited and the bright colours of drawings might restrict their use particularly in eye tracking studies because the differences in colour across pictures introduce unnecessary visual feature inequalities.

Taking into account the above, there is a clear need for a set of English-based high-quality line drawings in the field of psycholinguistics. Here, a novel set of stimuli is provided for use in language experiments. The following section outlines the development and pretesting (for quality) of this newly developed set of pictures for ambiguous experimental words and pictures, along with unambiguous filler words and pictures for use across a variety of language experiments.

Experimental items

A set of 88 ambiguous words (e.g. 'bark', 'cabinet') were taken from Rodd et al. (2016) and Experiments 2 and 3 (Chapter 2) of the present thesis. From this set of 88 possible stimuli, the author (HNB) evaluated whether the word was suitable for use in the picture semantic relatedness task, that is, that both the dominant and subordinate (second most common) meanings of each word could each be depicted using a single line drawing. Since one or both of the meanings were not deemed "picturable" for 28 words (e.g. the political meaning of 'cabinet' could not be drawn, thus the word 'cabinet' had to be removed), this left a set of 60 words that could have pictures designed for the dominant and subordinate meanings ready for the picture quality pretest. To maximise the number of potential stimuli, an additional 12 ambiguous words were taken from an existing stimulus set (Warren et al., in preparation) for which pictures could be created for the dominant and subordinate meanings[11]. This left a set of 72 experimental ambiguous words, each with two pictures, for the pretest.

[11] These words included non-homographs (word-meanings pronounced the same but spelled differently, e.g. 'night'/'knight') and polysemes as long as the dominant and subordinate meanings were judged by the author as sufficiently distinct that they could be distinguished on the basis of word association responses (e.g. typical associates related to the two related meanings of 'wave', *disturbance in water* or *hand gesture*, were deemed sufficiently distinct, whereas those related to the two meanings of 'passage', *corridor/tunnel* or *journey over time/distance*, were not). The percentage of polysemous and non-homographic out of the total stimulus set will be given in the details of the final stimulus set, after the pretest.

Filler items

Twenty-two unambiguous words were taken from Experiments 2 and 3 (Chapter 2) of the present thesis. An additional 21 unambiguous words were chosen from an existing stimulus set (Warren et al., in preparation) to increase the number of filler items. For each of these 43 filler items only a single line drawing was required as these unambiguous words have only one meaning.

Designing the pictures

For each ambiguous word, one picture was drawn for the dominant meaning and one picture was drawn for the subordinate meaning (the subordinate meaning was the second most common meaning; the third, fourth etc. most common meanings did not have pictures drawn for them), see Figure 7 for an example. For the unambiguous filler words, a single picture was created for each word. The pictures were based on the style of the pictures from Snodgrass and Vanderwart (1980) and Nishimoto et al. (2005), and were also inspired by line drawings from an online Microsoft picture resource (Microsoft Clipart, 2016). These were simple, slightly cartoon-like, black and white line drawings, (hand-drawn for increased control over the complexity of the pictures compared to photos). The author (HNB) drew all pictures using a Bamboo (Wacom, 2016) computer stylus and track pad on Photoshop software (Adobe Systems, 2016). All pictures were drawn in the same size and style: a black and white line drawing, with as few details as possible for the picture to clearly depict the correct meaning. Shading was also avoided where possible to maximise the impact of single lines in each drawing. Where pictures of humans or animals were required (e.g. ‘knight’ – a human on a horse), eyes were drawn to look closed rather than open because eyes are a facial feature that attract attentional gaze in particular (Itier, Villate, & Ryan, 2007). Attracting attention inconsistently across pictures is undesirable, as the aim is for the pictures to be relatively consistent in visual attractiveness.

Figure 7. Example of the newly created picture stimuli for the ambiguous word 'bark' (dog bark/tree bark) used in the semantic relatedness picture test.

A pretest was conducted to confirm that people consistently judged each picture to be a good representation of its intended meaning, as a measure of picture quality. The method and results of this picture quality pretest are outlined below.

Picture Quality Pretest

Method

Participants

In total, 102 native British English speakers participated in the picture quality pretest (70 females; mean age = 24, range = 18-45). All participants reported that they had no language, hearing or vision impairments (other than corrected-to-normal vision) and had lived in the UK for the majority of their lives, speaking English as their first language from birth. The pretest was conducted in two batches: one batch was run online via the Prolific online recruitment platform (Prolific Academic Ltd., www.prolific.ac, 2016) and one batch was run in the laboratory at the end of another experiment, recruited via poster advertisements and the University College London online recruiting website. All participants were paid the standard rate at the time of £6/hour.

Design & Materials

Each ambiguous word was always paired with one of its two corresponding meaning pictures to provide ambiguous items where the picture was supposed to be an accurate depiction of the intended word-meaning. However, although each unambiguous filler words had a corresponding picture, the set of unambiguous words was randomised such that each unambiguous word was paired with a mismatching picture to provide items where the picture was supposed to be an inaccurate depiction of the intended word-meaning. These trials were included to ensure that participants used the full range of the quality rating scale. For example, word ‘a’ was paired with picture ‘b’, and word ‘b’ was paired with picture ‘c’. The author (HNB) ensured that each of these randomised pairings was not inadvertently related in any way.

All ambiguous and unambiguous words and pictures were included in the picture quality pretest. However, the pretest was split into two versions. Version A included all ambiguous words, half paired with the dominant picture and half paired with the subordinate picture, and half of the unambiguous filler words paired with their respective mismatched pictures. In Version B, all experimental ambiguous words were again included but with the alternative meaning picture to Version A (i.e. subordinate meaning picture where it was the dominant meaning picture, and vice versa). Version B also included the remaining half of the unambiguous filler words paired with their respective mismatched pictures. Hence, a given participant contributed responses to half of the total number of ambiguous and unambiguous pictures but, across participants, all pictures received the same number of responses.

Procedure

The picture quality pretest was presented to participants using Qualtrics survey software (Qualtrics Inc., www.qualtrics.com, 2016), regardless of whether they were tested online or in the laboratory. Each participant was randomly assigned to one of the two versions. After giving their informed consent, participants' demographic data were collected and instructions were displayed on-screen. Participants were told that they would see some pictures, each accompanied by a word and a definition of that word, and that their task was to rate how much the picture was related to the defined meaning of the word. They were asked to rate the relatedness of each picture-word pair on a five-point scale (1 – highly unrelated, 2 – somewhat unrelated, 3 – neutral, 4 – somewhat related, 5 – highly related). Trials were presented in a different random order for each participant. Each picture-word pair was displayed on a separate screen, where participants were required to press an on-screen arrow button to progress to the next trial. There were no time restrictions on the task, although participants were encouraged not to deliberate for too long on each trial and were assured that the experimenter was interested in their opinion and that there were no right or wrong answers.

Results

All data from the two testing batches were combined. Ratings were averaged across participants to provide a mean rating per item (picture-word pair). A mean rating of 4 (somewhat related) was the minimum requirement for an ambiguous word picture to be a suitable depiction of the intended meaning. Although this value is an arbitrary threshold, it is stringent and should therefore ensure high quality of the pictures in the final stimulus set. An ambiguous word was removed from the set if at least one of the pictured meanings failed to meet the criterion of a mean rating of 4 or over. Out of the total 144 pictures, only 2 fell below the criterion of a mean rating of 4 or over (the subordinate picture for 'bar' and for 'craft'), hence 'bar' and 'craft' were removed from the stimulus set leaving 70 words and 140 pictures. All filler items were rated less than 3 and were therefore all deemed to be unrelated[12]. This means that the stimulus set for use in Experiment 4 comprised of 70 ambiguous words and their corresponding dominant and subordinate meaning pictures (140 in total) and 43 unambiguous filler words and their corresponding unrelated pictures. Details of the final stimulus set will be provided in Experiment 4 (following the Task and Coding Checks section, which outlines some further item exclusions from the stimulus set).

[12] Only two unrelated picture-word filler pairs were rated above 2 on average – the picture 'jug' paired with the word 'bath' and the picture 'pond' paired with the word 'feather'; mean rating 2.71 and 2.74 respectively. These can be removed if a more stringent threshold is required for others experiments, although this was deemed adequate for Experiment 4.

Experiment 4 – validation of picture semantic relatedness test & collection of word-meaning dominance norms

Using the newly developed picture stimuli in a semantic relatedness task

There are two aims of Experiment 4. The first aim is to validate the picture stimuli for use with the semantic relatedness task by confirming that the task can detect differences (in reaction times and/or accuracy) between picture probes of the dominant and subordinate meanings of ambiguous words. In the present semantic relatedness task, the ambiguous word is presented auditorily and, at word offset, a picture of either the dominant or the subordinate meaning is displayed. The meaning of the picture presented varies by trial and across participants, such that both meanings of each ambiguous word are presented across all participants, but half encounter the dominant and half encounter the subordinate picture. Participants must then judge whether this picture is semantically related to the word (i.e. whether it depicts either the dominant or subordinate meaning, or the picture is not related to the word). The assumption is that where the picture is consistent with the participants' preferred interpretation (i.e. it is the expected and available meaning), reaction times are faster and/or accuracy is increased. Where the picture is inconsistent with the participants' preferred interpretation (i.e. it is the unexpected and less available meaning), reaction times are slower and/or accuracy is reduced. Hence, faster and/or more accurate responses reflect more available meanings and therefore higher dominance at the time of testing.

The second aim is to compare semantic relatedness performance to performance in word association and eDom measures. This will provide two measures of baseline dominance for each picture in the stimulus set. This information is therefore useful to other researchers since different patterns of results can be found with different meaning dominance stimuli (e.g. Armstrong et al., 2012; Rodd et al., 2013, Experiment 1). The additional word association and eDom measures in the present experiment will provide information on the pictures that can be used by other researchers in language experiments, providing a dominance baseline against which other experiment results can be compared.

Method

Participants

In total, ninety-one native British English speakers participated in the present experiment (59 females; mean age = 22, range = 18-45). All participants reported that they had no language, hearing or vision impairments (other than corrected-to-normal vision) and had lived in the UK for the majority of their lives, speaking English as their first language from birth. Participants were recruited through poster advertisements and the University College London online recruiting website, and were paid the standard rate at the time of £8/hour.

Design

The present experiment had a between-subjects design where participants were pseudo-randomly allocated to one of three tasks: the word association test, the eDom test or the picture semantic relatedness test. Since the aim was to investigate whether word association and eDom dominance scores predicted performance in the picture semantic relatedness test, for the main analysis, word association and eDom dominance scores were used as the two independent variables and reaction times and error rates in the semantic relatedness task were the two dependent variables (in separate analyses). The picture quality pretest ratings (the mean participant rating per word-meaning) were also included in analyses as a covariate.

The aim was to equate the three tasks for total participant time. For eDom, the guidance from Armstrong et al. (2012) was followed, which suggests that 16 participants are required to generate norms for up to 146 words. It was estimated that eDom would take each participant approximately 20 minutes to complete, summing to 320 minutes of participant time in total for the eDom task. Since it was estimated that word association and semantic relatedness tests would each take approximately 10 minutes to complete – half the time of eDom – the number of participants required for these tasks was doubled. This resulted in 32 participants for word association and 32 participants for semantic relatedness (16 for each of the 2 versions), again adding up to 320 minutes of participant time in total per task.

Materials

The set of 70 ambiguous words (with dominant and subordinate pictures from the picture quality pretest) formed the stimuli for the present experiment.

Word Association

All 70 experimental ambiguous words were included in the word association test, together with a further 43 unambiguous filler words, which were included to reduce the proportion of ambiguous words in the task with the aim of making the ambiguity less salient. All words were presented auditorily, recorded by a female native British English speaker with a Southern English accent (HNB). This auditory modality was chosen to ensure methodological consistency between this word association test and those used in previous experiments (Rodd et al., 2016; Rodd et al., 2013).

eDom

All 70 experimental ambiguous words were included in the eDom test, although no unambiguous fillers were required since the task was to rate the relative occurrence of each meaning of an ambiguous word. The programme was identical to how it was designed by Armstrong et al. (2012), with six definition entry boxes. However, only two meanings were provided per word (the dominant meaning and the most common subordinate meaning), which corresponded to the two meanings depicted in the picture semantic relatedness task. This was because these were the only meanings that were to be analysed. The ambiguous word was displayed at the top of the eDom screen and each of the two word-meaning definitions was presented in a separate white box below, which was coupled with a 'percent of occurrences' box into which participants could type their ratings. For each word, three further boxes were coloured yellow to indicate that participants could enter their own definitions and percentage of occurrences of the word if they were not included in the two meanings already listed. The sixth box was coloured grey to show that it was to be ignored. See Figure 8 for an example of the eDom software design.

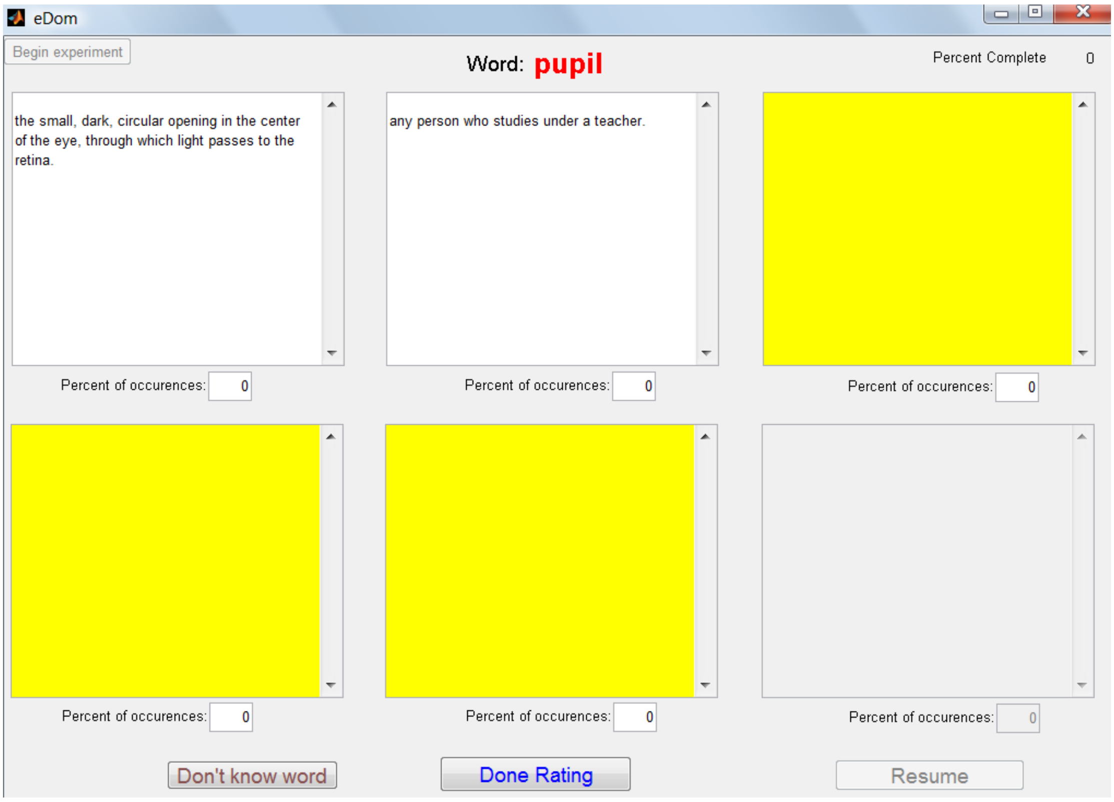

Figure 8. Experiment 4. An example of the eDom screen shown to participants for the ambiguous word 'pupil', taken from online eDom software http://edom.cnbc.cmu.edu (Armstrong et al., 2012).

Picture Semantic Relatedness

There were 140 pictures (a dominant meaning picture and a subordinate meaning picture for each of the 70 ambiguous words) for use in the picture semantic relatedness test. Each participant encountered all 70 ambiguous words but only 70 pictures – either the dominant or the subordinate picture of each word. All words were presented auditorily, recorded by a female native British English speaker with a Southern English accent (HNB).

All 70 experimental trials were 'related' at test (i.e. the picture was related to the auditory word). A further 43 unambiguous filler words paired with unrelated pictures were included to reduce the salience of the ambiguity but, most importantly, to also provide trials in which the picture was *not* related to the auditory word (and therefore required an 'unrelated' response). This meant that approximately one third of trials were dominant and related, one third were subordinate and related and one third were unambiguous and unrelated. Although this meant that all unambiguous trials were unrelated, the dominant trials should have also often appeared to participants as unambiguous (e.g. hearing 'bark' and seeing a picture of a dog), since the dominant meaning is, on average, the most easily disambiguated meaning and therefore participants are unlikely to be aware of the alternative, subordinate meaning.

Procedure

All conditions were conducted in laboratory cubicles on desktop computers using MATLAB (R2015a, 2015; version 8.5.0.197613). Participants wore headphones for the whole experiment in all conditions to ensure that the stimuli could be heard easily (for word association and semantic relatedness only) and to minimise any background noise. Each participant was randomly assigned to one of the three test conditions of the experiment. After giving their informed consent, participants' demographic data were collected and instructions for the experiment were displayed on-screen.

Word Association

A fixation cross on an otherwise blank screen was displayed for 1000ms, followed by the auditory presentation of an ambiguous word. On word offset, a blank textbox replaced the fixation cross and participants were required to type into that box the first word they thought of in relation to the auditory word. They were asked to type '0' if they were unable to make out the word, unable to generate a response or felt uncomfortable giving one. Once they had typed their responses, participants pressed the 'enter' key, which triggered the next trial. Responses longer than 1500ms were followed by an on-screen prompt that lasted for 2000ms to encourage participants to respond faster on subsequent trials. Two practice trials were always presented at the start of the task. Although the presentation order of experimental items was randomised, a further four of the filler items were always presented at the start of the test to allow participants to become accustomed to the task. The mean duration of this task was 12 minutes.

eDom

Comprehensive instructions were presented to participants to explain the task and give examples of the eDom programme. These instructions were those provided in the eDom package (Armstrong et al., 2012) and were obtained from the following website address: http://edom.cnbc.cmu.edu. Participants were presented with an ambiguous word at the top of the screen accompanied with the dominant and subordinate meaning definitions in boxes. Participants were required to rate each definition with regards to their perception of its relative frequency (percentage) of occurrence in English language, in their own experience, where ratings had to sum to 100%. Definitions were presented in a random order (i.e. either the dominant or subordinate meaning first, which varied by-trial and across participants). Participants were also able to add additional meanings of each ambiguous word by entering the definition into a blank box and including that meaning's frequency percentage. Whilst this was encouraged (to keep the method as similar as possible to Armstrong et al., 2012), it was not necessary for successful task completion. Once participants had rated the frequency of all meanings of a given word, they were required to press the 'done rating' button, which prompted the next trial. If participants were not familiar with a word, they were able to press the 'don't know word' button, which also advanced the task to the next trial. The mean duration of this task was 23 minutes.

Picture Semantic Relatedness

There were two versions of the stimuli, which counterbalanced which picture was presented to each participant. This ensured that, across participants, half encountered the dominant picture for a given word and half encountered the subordinate picture of that word. Each participant encountered half the words with the dominant meaning picture and half with the subordinate meaning picture.

For each semantic relatedness test trial, a fixation cross on an otherwise blank screen was displayed for 1000ms, followed by the auditory presentation of an ambiguous word (identical to the start of the word association trials). On word offset, a probe picture was then presented in place of the fixation cross. Participants were asked to respond as quickly and accurately as possible to the probe picture by either pressing the 'r' key for related in meaning to, or the 'u' key for unrelated in meaning to, the auditory word. Response times longer than 1500ms prompted a message displayed for 2000ms encouraging faster responses on subsequent trials. The key press response triggered the next trial. One practice trial was given at the start of the task, with feedback. All trials were randomised with the exception of four filler trials at the start of the task, which allowed participants to become accustomed to the task. The mean duration of this task was 12 minutes.

Task and Coding Checks

The word association responses needed to be coded for each ambiguous word by the first author (HNB). Each response was coded either as being related to (1) the dominant meaning, (2) the primed subordinate meaning, or (3) 'other', which included alternative meanings of the word, responses which were ambiguous/unclear and '0' responses (which participants were instructed to give if they could not think of a response or felt uncomfortable giving a response). For example, for the subordinate meaning of 'glasses' as in the sentence 'she poured the champagne into the glasses', the word association response 'eyes' would indicate the dominant meaning, whereas the response 'drink' would indicate the subordinate meaning. Any uncertainties were discussed with another researcher and if any doubt remained as to which meaning a participant intended, the response was coded as 'other'. For the analyses, 'other' responses (5%) were removed.

All participants were checked for task performance. Out of 91 participants, 8 were removed for not meeting demographic requirements or for showing poor performance on the task they completed (if more than 20% of responses in word association were the auditory word repeated, less than 70% accuracy in the semantic relatedness task, or spent over 30 minutes on the eDom task). This left a total of 83 participants for which data were analysed: 32 for word association, 17 for eDom and 34 for the picture semantic relatedness task.

In addition, all items were checked for performance. If an item was excluded from one task, it was also excluded from the other two so that all items in the analyses had sufficient data across all tasks. This resulted in the exclusion of 5 items, leaving a total of 65 items for analyses (*drill, gear, racket* and *temple* had less than 70% accuracy on average for the subordinate meaning in the semantic relatedness picture task, *sink* had 40% of word association responses that could be coded as either the dominant or subordinate meaning, e.g. 'water').

Details of final picture stimulus set

Due to the 5 excluded items (see Task and Coding Checks section above for details), the final stimulus set proposed in this chapter comprises of 65 experimental ambiguous words, each of which has an accompanying dominant picture and subordinate picture. For these ambiguous words, a word association based pretest (Warren, Vitello, Devlin & Rodd, in preparation) showed that the subordinate meaning was semantically distinct from the dominant meaning and the mean dominance for the subordinate meaning was 26% (range: 0-48%). This included 39 (60%) words that were polysemous (Parks et al., 1998), and 5 (8%) words that were non-homographs, where the two meanings were pronounced in the same way but spelled differently. In addition to the experimental items, there are 43 filler unambiguous words, each of which has a related picture of its meaning as well as an unrelated picture with which it can be paired. Details of the final set of 65 experimental ambiguous words and pictures (dominant and subordinate), as well as the 43 filler unambiguous words and pictures, are available online at: https://osf.io/4fmqu/files/.

Results

There were two stages to the analysis for the present experiment. The first was to validate the picture stimuli for use with the semantic relatedness task by confirming that the task can detect differences (in reaction times and/or accuracy) between picture of the dominant and subordinate meanings. The second was to derive dominance norms, from word association and eDom measures, on the pictured meanings (dominant and subordinate) in the stimulus set. This second stage therefore investigated the relationship between word association dominance scores, eDom dominance scores, picture quality (from picture quality pretest ratings) and semantic relatedness performance and whether word association and eDom scores predicts RTs and/or error rates in the semantic relatedness picture task. The analyses are separated as such below. All analyses are between-item, since responses are being made to the pictures, and these are all *either* dominant or subordinate.

The majority of the analyses below used linear (for RTs) or logistic (for errors) mixed effects modelling, with the lmer and glmer functions respectively from the lme4 package (version 1.1-7; Bates et al., 2014) in R (version 3.3.1; R Core Team, 2016). Mixed effects modelling is the most appropriate form of analysis for the present data since it takes within-subject and within-item dependencies into account within a single model (Jaeger, 2008). The construction of models with different fixed effects was required for the two stages of analyses (details can be found in the relevant sections, below). Once the main model had converged, a model comparison approach (Baayen et al., 2008) was used to test the significance of individual factors. This involved removing (from the fixed effects but not the random effects) one factor at a time and in each case comparing this reduced model to the main model using a likelihood ratio test to assess whether the inclusion of the factor significantly improved the model fit.

Analysis Stage 1: Sensitivity of Semantic Relatedness to Dominance

Reaction Time Analyses

RTs for filler trials and incorrect trials were removed from the data, as were RTs less than 300ms and greater than 1500ms, as these suggested accidental key presses or unusually slow responses (respectively). In addition, RTs less than or greater than 3 standard deviations from a participant's mean RT were discarded.

A model was built with fixed effects of picture dominance category (dominant or subordinate picture) and picture quality (picture quality by-items factor, which had been averaged across participants to provide a single average rating per picture), with only picture dominance category as a random effect (for subjects and items slopes, intercepts and the correlations between these). The picture quality factor was not included in the random effects structure because it was only in the model as a covariate. This model converged for the raw RTs. However, since the assumptions of normality and homoscedasticity were violated, the RT data were inverse transformed and the same model was run on these inverse RTs. This was also compared to the same model where the RTs were log transformed. The inverse transformation showed to be the most suitable, since it did not violate the assumptions of normality and homoscedasticity, and was therefore used as the dependent variable in this model. A model comparison approach was used to test the significance of the picture dominance category factor, which involved removing the factor of interest from the fixed effects structure and comparing this to the model with it in.

Participants were faster when responding to the dominant picture, compared to the subordinate picture (Figure 9).

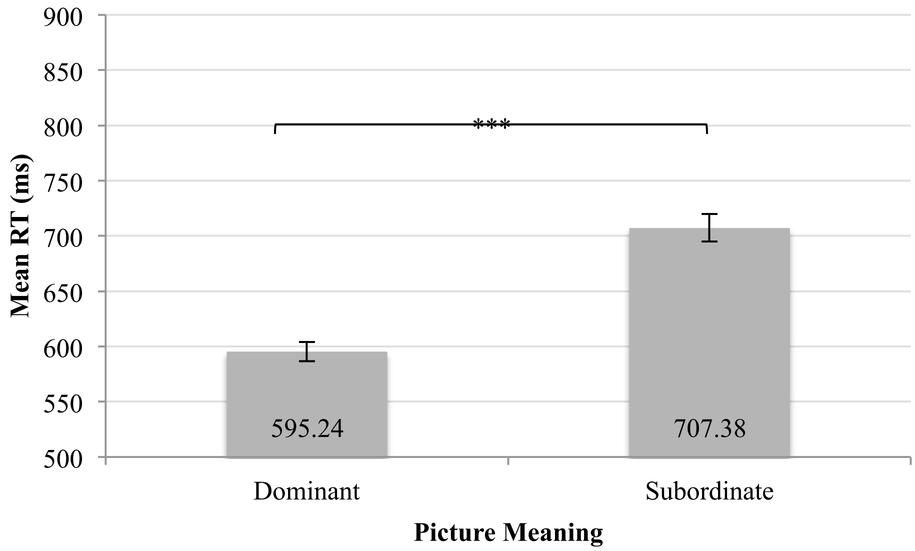

Figure 9. Mean by-items raw reaction times (averaged across participants) for both the subordinate and dominant meaning picture in Experiment 4. Significance level indicated with asterisks (***<.001).

The model comparison revealed a significant main effect of picture dominance category, whereby participants were significantly faster to respond to the dominant meaning pictures than the subordinate meaning pictures (X^2 (1) = 33.851, p < .001). This confirms that reaction times in the semantic relatedness test using pictures are sensitive to the dominance of an ambiguous word-meaning (i.e. whether it is the dominant or the subordinate meaning).

Error Rate Analyses

Filler trials and experimental trials where responses were less than 300ms and greater than 1500ms were discarded, as in the RT analyses (these suggested accidental key presses or unusually slow responses, respectively). Other than the use of the glmer function (appropriate for logistic models), the analysis approach was identical to that of the RTs where a model was built with fixed effects for picture dominance category and picture quality, with only picture dominance category as a random effect.

Participants made fewer errors when responding to the dominant picture, compared to the subordinate picture (Figure 10).

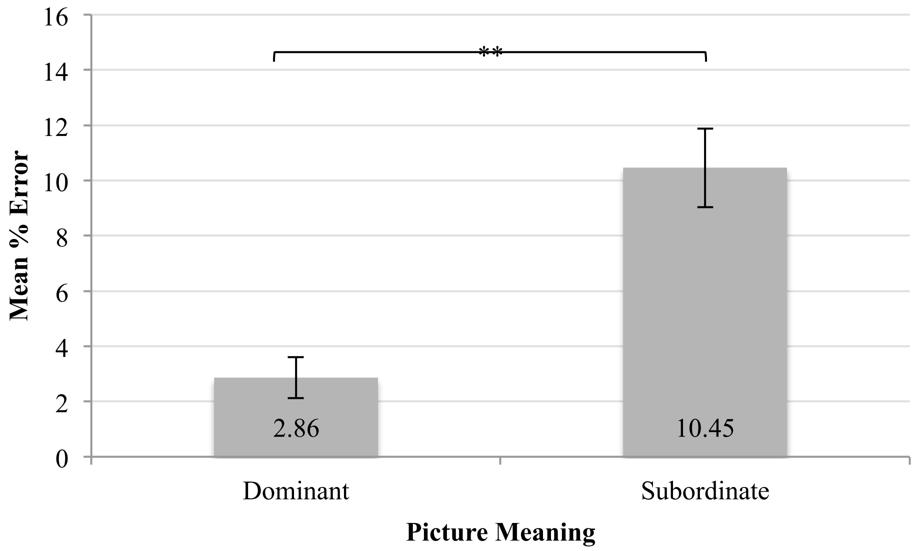

Figure 10. Mean by-items percentage error rate (averaged across participants) for both the subordinate and dominant meaning picture in Experiment 4. Significance level indicated with asterisks (**<.01).

The model comparison approach revealed a significant main effect of picture dominance category whereby participants made fewer errors when responding to the dominant meaning pictures compared to the subordinate meaning pictures (X^2 (1) = 10.588, p = .001). This confirms that accuracy in the semantic relatedness test using pictures is sensitive to the dominance of an ambiguous word-meaning.

Analysis Stage 2: Comparing Different Measures of Word-Meaning Dominance

The data were entered into mixed effects models to investigate whether word association and eDom scores predicted RTs and/or error rates in the semantic relatedness task. Whilst multiple regression has long been the standard approach for this type of analysis, more recently it has been shown that mixed effects analyses are more appropriate, since they take trial-level information into account, unlike standard multiple regression (Jaeger, 2008). Mixed effects analyses are therefore the main analyses in this section and allow us to infer whether word association and eDom performance (as well as picture quality) affect semantic relatedness performance.

Obtaining correlations between variables and measures of determination (e.g. R^2) from mixed effects models is problematic. There is no clear method for calculating model-specific variable correlations or an R^2 measure of variance explained by each predictor (some methods have been proposed, e.g. Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008, although the inference gained from such statistics is questionable and depends heavily on the experimental design). In order to provide transparent statistics that are easily comparable to other research, correlation matrices between variables and multiple regressions, thereby providing R^2 for each predictor, were computed using JASP statistical software (JASP Team, https://jasp-stats.org/, 2017). These correlations, multiple regressions and R^2 results are reported in Appendix C.

Additionally, coded responses for the word association and eDom tests were averaged across subjects to provide a single dominance score per meaning. The data were then split by word-meaning, such that dominant and subordinate meanings were analysed separately. This is because if both the dominant and subordinate meaning of a word were included in a single model, they would be treated as independent items, which they are not. In other words, there was a model for dominant meanings, which included only dominant meaning scores for word association and eDom as predictors of dominant picture semantic relatedness RTs and errors, and a separate model for subordinate meanings, which included only subordinate meaning scores for word association and eDom as predictors of subordinate picture semantic relatedness RTs and errors. It is important to bear in mind when interpreting the present findings that, for both dominant and subordinate meaning categories, a higher dominance score indicates that that particular meaning is more frequent.

The separate models for the dominant and subordinate meanings were constructed with fixed effects for word association and eDom, and picture quality was included as a fixed effect covariate. An intercepts-only random effects structure was used. It did not make theoretical sense to include slopes for word association, eDom and picture quality, as they have been included only as by-item factors.

There were two stages to each set of analyses. For each of dominant meaning RTs, dominant meaning errors, subordinate meaning RTs and subordinate meaning errors analyses, the following two steps were conducted to analyse how word association, eDom and picture quality affected semantic relatedness performance:

1. A model comparison approach was used to test the significance of each factor alone (i.e. each predictor in a separate model) and therefore their individual effects on predicting semantic relatedness performance.
2. A model comparison approach was used to test the significance of each factor in predicting semantic relatedness performance whilst the model accounted for all other factors (i.e. all predictors included in one combined model).

Mixed effects model analyses

The full model converged in all cases for the raw RTs. However, since the assumptions of normality and homoscedasticity were violated, the RT data were inverse and log transformed and the same model was run on these transformed RTs. The inverse transformation showed to be the most suitable transformation for the dominant RTs, with the log transformation the most suitable for the subordinate RTs, since these transformed data no longer violated the assumptions of normality and homoscedasticity. These transformed variables were therefore used as the dependent variables in the respective mixed effects models for RTs. See Table 3 for the results.

The pattern of significance for the three predictors was the same for the models of dominant RTs, subordinate RTs and subordinate errors as the dependent variable. Higher word association dominance scores, higher eDom dominance scores and higher picture quality ratings all significantly predict improved semantic relatedness performance when included in separate models (compared to the null model). However, when all three predictors are entered into the same model, eDom ceases to significantly predict semantic relatedness performance. This suggests that, despite a medium to high correlation between word association and eDom, word association is a stronger predictor of performance in the picture semantic relatedness task. Hence, only higher word association dominance scores (i.e. more frequent meanings as measured by word association) and higher picture quality ratings predict faster RTs to dominant pictures, faster RTs to subordinate pictures and fewer errors to subordinate pictures.

As for the models of dominant errors, neither word association scores, nor eDom scores, nor picture quality ratings seem to significantly predict the accuracy of performance in the semantic relatedness task. This is most likely due to there being so few erroneous responses to dominant pictures in this task (i.e. so little variance in the model).

Broadly speaking, the additional simultaneous regression analyses are consistent with the mixed effects analyses. There are only two differences in the patterns of significance between the two types of analysis, which only concern the picture quality covariate: (1) for the dominant RTs, picture quality is significant for mixed effects analyses but marginal in the regression, and (2) for dominant errors, picture quality is not significant for mixed effects analyses but is significant in the regression. These differences will not affect the conclusions drawn for the word association and eDom predictors of semantic relatedness performance.

Table 3. Experiment 4 mixed effects model analyses for dominant meaning RTs and errors, and subordinate meaning RTs and errors. For each of these four measures, the three predictors (word association (WA), eDom and picture quality rating (PicQuality)) were entered individually in separate mixed effects models and then entered simultaneously into a single, combined mixed effects model. Significance level emphasised with asterisks (* <.05, ** < .01, ***<.001).

Measure	**Predictors**	**In separate models**		**In combined model**	
		X^2 **(1)**	***p***	X^2 **(1)**	***p***
Dominant RTs	WA	23.879	<.001***	13.811	<.001***
	eDom	7.156	.008***	0.028	.868
	PicQuality	8.667	.003***	4.007	.045*
Dominant Errors	WA	3.384	.066	0.853	.356
	eDom	2.385	.123	0.094	.759
	PicQuality	3.060	.080	1.725	.189
Subordinate RTs	WA	22.601	<.001***	17.530	<.001***
	eDom	7.471	.006**	0.007	.936
	PicQuality	13.354	<.001***	16.319	<.001***
Subordinate Errors	WA	11.804	<.001***	8.380	.004**
	eDom	4.500	.034*	0.050	.824
	PicQuality	8.378	.004**	9.318	.002**

Discussion

There were two aims of Experiment 4. The first aim was to validate the picture stimuli for use with the semantic relatedness task by confirming that the task can detect differences (in reaction times and/or accuracy) between picture probes of the dominant and subordinate meanings of ambiguous words. The second aim was to compare word association scores and eDom scores (including a picture quality covariate) as predictors of semantic relatedness performance.

The first stage of analyses confirmed that the semantic relatedness picture task is a successful measure of dominance effects on comprehension, due to the significant differences of speed and accuracy between the dominant and subordinate meaning pictures. On average, responses to dominant meaning pictures were 112ms faster and 7.6% more accurate than to subordinate meaning pictures. This demonstrates that this semantic relatedness task is a suitable test of the difference in availability between an ambiguous word's alternative meanings.

The second stage of analyses compared word association scores and eDom scores (including a picture quality covariate) as predictors of semantic relatedness performance. The data were entered into mixed effects models to investigate whether word association scores, eDom scores and picture quality ratings predicted RTs and/or error rates in the semantic relatedness task. Each predictor was first entered into a separate mixed effects model. Then, all predictors were entered together into the same mixed effects model.

When word-meaning dominance was measured using a word association test, a higher mean dominance predicted significantly faster reaction times to dominant and subordinate pictures. It also predicted significantly fewer errors to the subordinate picture in the semantic relatedness test. The significant effect of word association was not eliminated by either the inclusion of eDom as a predictor or by picture quality as a covariate in the model. This indicates that word association is a strong and reliable predictor of semantic relatedness performance overall.

Similarly, when word-meaning dominance was measured with eDom, a higher mean dominance predicted significantly faster reaction times to dominant and subordinate pictures, as well as significantly fewer errors to the subordinate picture. This significant effect of eDom was, however, eliminated by the inclusion of the word association predictor and the picture quality covariate. Clearly, eDom, as a dominance measure, is related to semantic relatedness performance, but it seems that word association is more strongly related. This relationship might be driven by semantic relatedness and word association tasks being more similar where, unlike eDom, they are both speeded tasks. Additionally, participants are explicitly made aware of the ambiguous nature of the words in eDom but might not be aware of this in the other two tasks, which also makes semantic relatedness and word association more similar. Either way, word association does seem to be the stronger candidate in the present study.

Word-meaning dominance scores from word association and eDom (or picture quality) did not significantly predict mean error rates for the dominant picture in the semantic relatedness test. This might be because the dominant picture is, on average, the most expected meaning, thus participants are unlikely to make incorrect responses to dominant pictures, leaving little variance in the model. It is therefore not surprising that the dominant pictures do not show significant error results and yet the subordinate pictures do – responses to subordinate pictures are more likely to be incorrect, as this meaning is unexpected on average.

There are clear advantages and disadvantages of both word association and eDom methods of collecting dominance norms. Word association and eDom can be relatively easily deployed both in the laboratory and online and set up in a range of programs. Moreover, as argued by Armstrong et al. (2012), eDom requires few participants, since eDom can collect the dominance scores for multiple meanings of an ambiguous word per participant, whereas word association can only collect one meaning per word per participant. As explained in the Introduction, despite the fewer required participants, eDom requires double the amount of time to complete per participant, making the lower participant requirement of eDom less of an advantage.

Overall, the findings indicate that word association is a stronger task than eDom, although eDom remains an effective method for collecting dominance norms. Of course, the use of either of the methods should depend on the particular needs of different experiments. As for the newly developed pictures, the word association and eDom tests have provided two sets of dominance norms for the dominant and subordinate picture for each word, along with the picture quality rating[13]. The present chapter has also shown that the semantic relatedness task, using the newly-developed picture stimuli, is sensitive to dominance effects on comprehension, and so provides an appropriate measure of meaning availability that could be used to measure word-meaning priming. This will be the focus of the experiments in Chapter 4, which will use this task to examine effects of priming on both the primed and unprimed meanings of ambiguous words.

[13] Since pictures with ratings lower than 4 out of were excluded after the pretest (with 5 being the highest picture-meaning relatedness and therefore the best depictions of the meanings), even the lower quality pictures here have still met this stringent minimum threshold.

Chapter 4: Priming and the availability of the unprimed meaning

Introduction

Many studies have shown that context can help comprehenders rapidly select the appropriate meaning of an ambiguous word (e.g. Chen & Boland, 2008, Experiment 2; Colbert-Getz & Cook, 2013; Duffy et al., 1988). For example, compared to 'she sat next to the bank', the additional context provided by 'she sat next to the river on the grassy *bank*' increases the availability of the riverside land meaning of 'bank' to maximise the processing efficiency of the subsequently encountered ambiguous word. Despite the literature showing this context-driven boost in the availability of the encountered meaning, it is still not clear how this increase affects the unencountered, inappropriate meaning.

There are two outcomes for the availability of the inappropriate meaning as a result of an increase in availability of the appropriate meaning. On the one hand, it is possible that an increase in appropriate meaning availability does not affect the level of availability of the alternative (competing) meanings of an ambiguous word. This would mean that the increased availability of the riverside meaning of 'bank' does not decrease the availability of the alternative financial meaning of 'bank' or change the way in which a listener would access that meaning. On the other hand, it is possible that an increase in appropriate meaning availability is associated with a decrease in inappropriate meaning availability. This would mean that the increase in availability of the riverside meaning of 'bank' makes the financial meaning harder to access.

The learning mechanism(s) underlying language comprehension cannot be understood until we know the consequences for the availability of inappropriate meaning. This issue of reciprocal changes in meaning availability has implications for any model of semantic ambiguity resolution, as a model is incomplete without accounting for whether or not the representations of alternative meanings of a word affect one another. Hence, the aim of the present chapter is to investigate whether alternative word-meanings and their availabilities are independent of one another, or whether they are necessarily linked to each other. A word-meaning priming paradigm (Rodd et al., 2013) will be newly applied to this area of research, where word-meanings are encountered in context and then tested with a semantic relatedness test in isolation after a delay. This will allow the measurement of availability of both the primed (recently appropriate and therefore likely to be available) and unprimed (recently inappropriate) meanings of ambiguous words in the absence of immediate context.

Existing literature on semantic ambiguity resolution

There are two predominant classes of models for semantic ambiguity resolution, which make different predictions about whether or not alternative word-meaning availabilities are independent of one another. It is important to note that the models within these classes tend to make predictions about meaning activation with immediate sentential context, rather than meaning availability following a priming manipulation and delay (between prime and test). Nevertheless, it is important to consider what the model classes predict about the activation of multiple word-meanings generally, and whether they have the potential to be extended to incorporate effects of word-meaning priming and therefore effects of recent experience on comprehension.

Models of short-term context effects

The first model class assumes that all possible meanings of an ambiguous word are activated when the word is encountered and that contextual cues only act to make the appropriate meaning more available for selection. According to this account, this increase in availability *never* causes a decrease in the availability of competing meanings. An example of this type of model is the widely accepted 'reordered access model' (Duffy et al., 1988). This model accounts for findings on the subordinate bias effect, where fixation times on biased ambiguous words are longer following subordinate context, compared to balanced ambiguous words

(Rayner et al., 1994). This suggested that subordinate context increases subordinate meaning availability so that it competes with the already-available dominant meaning and therefore slows disambiguation, but all meanings still tend to be available (i.e. exhaustive access; Vu & Kellas, 1999).

Specifically, the reordered access model assumes that all possible word-meanings are always activated in parallel upon encountering the ambiguous word, where the dominant meaning is typically the most available meaning. Crucially, immediately present context serves to "reorder" the pattern of meaning activation, allowing the contextually appropriate meaning to be rapidly selected. This allows comprehension to be guided by contextual cues. For instance, where context is subordinate (e.g. 'the grassy *bank*'), the dominant meaning is activated due to frequency (financial 'bank'; since it is more likely to occur) and the subordinate meaning is activated due to the presence of subordinate context (the riverside meaning), hence the two meanings compete for selection. Importantly, according to this class of model, the correct selection of the subordinate meaning does *not* cause the contextually inappropriate dominant meaning to be inhibited; dominant meaning availability is unaffected by boosted subordinate meaning availability (Chen & Boland, 2008; Duffy et al., 1988; Rayner, Binder, & Duffy, 1999). In summary, this class of model is consistent with the idea that alternative word-meaning availabilities change independently of one another.

Evidence of short-term context effects

The literature provides inconclusive evidence in support of this view. In an eye tracking study using the visual world paradigm, Chen and Boland (2008) found evidence that context can both increase and decrease the level of activation of alternative word-meanings. In their second experiment, participants were presented with sentences in which context biased the interpretation of the ambiguous word towards the subordinate meaning. The ambiguous word appeared at the end of the sentence and, upon its onset, four pictures were displayed: the subordinate meaning referent, a shape competitor of the dominant referent (shown to track lexical access to the dominant meaning without biasing participant responses directly to the dominant meaning; Dahan & Tanenhaus, 2005; Huettig, Gaskell, & Quinlan, 2004), and two fillers. Compared with neutral contexts, subordinate context decreased looks to the dominant meaning shape competitor, although the shape competitor still received more looks than fillers, indicating that the dominant meaning was less available after the subordinate context. Chen and Boland (2008) therefore provide evidence that the activation of the inappropriate dominant meaning can be inhibited by the subordinate context. This finding is incompatible with the reordered access model, which is consistent with the idea that prior disambiguating context should not inhibit activation of the inappropriate meaning.

In contrast, Colbert-Getz and Cook (2013) concluded that they found no evidence that subordinate priming reduces activation of the dominant meaning. Their study involved eye tracking while participants read a set of nine sentences. In the "neutral" context condition, the fifth sentence contained an ambiguous word subsequently disambiguated towards its subordinate meaning, whilst the eighth sentence contained the same word subsequently disambiguated towards its dominant meaning. The "unelaborated" context condition provided one additional sentence of subordinate context in sentence four, whilst the "elaborated" context condition provided four additional sentences of subordinate context in sentences one to four (although no additional repetitions of the ambiguous word itself in either condition). Hence the unelaborated and elaborated conditions provided either weak or strong prior subordinate context, respectively. Colbert-Getz and Cook (2013) found that, compared to the neutral condition, the unelaborated condition did not slow reading times of the later encountered dominant sentence. In contrast, the elaborated condition *did* slow reading times of the later encountered dominant sentence compared to the neutral condition. The authors concluded that the slower reading times of the dominant meaning following strong subordinate priming were evidence of an increase in availability of the subordinate meaning with no reduction in availability of the dominant meaning, and that these findings are consistent with the reordered access model.

Models accommodating longer-term priming effects

The second model class assumes that the activation levels of alternative word-meanings are necessarily linked. When subordinate context increases the availability of the subordinate meaning, the competing dominant meaning must decrease in availability. One model that falls into this second class is the distributed connectionist model developed by Rodd and colleagues (Rodd et al., 2004; Rodd et al., 2013). The nature of the distributed representations means that changing the structure of lexical-semantic representations to make one meaning more readily available will *necessarily* make the other meaning(s) less readily available.

Applying this model to word-meaning priming (i.e. learning from recent experience), there are two ways in which recent experience could strengthen connections. One way is that availability is increased by strengthening connections *between* layers in the network. Here, priming would be driven by changes in the connections between the form-based (phonological or orthographical) representation and the semantic (meaning) representation as a result of experience with the meaning (form-to-meaning mapping; Rodd et al., 2013). Another way is that availability is increased by strengthening connections *within* the semantic layer in the network (Rodd et al., 2016). Here, the semantic units activated for a given meaning become more strongly connected to one another because of the recent experience and result in a more stable semantic representation.

In both cases, the strengthening of connections related to the primed meaning would necessarily weaken the connections related to the unprimed meaning, leaving this meaning less available on a subsequent encounter with the word (compared to if priming from recent experience had not occurred). It is currently unclear which connections (either between or within layers) are affected by priming without running the model simulations, although recent findings have provided some evidence that changes might be made within the semantic layer of the network (since priming effects were shown not to be modality-specific; Gilbert et al., 2018). Either way, this class of model is consistent with the idea that alternative word-meaning availabilities change in relation to one another.

Evidence of longer-term priming effects

Much of the existing literature focuses on disambiguation with immediate context effects, where the availabilities of alternative word-meanings are tested immediately after biasing sentential context and on a trial-by-trial basis (e.g. Chen & Boland 2008; Duffy et al., 1988). Whilst this has been seen as the "window" into the disambiguation process, it is not the only way of testing whether or not meaning availabilities are independent of one another – the lasting effects of recent experience shown with word-meaning priming (Rodd et al., 2013) allow us to examine whether *prior* (i.e. recently encountered but no longer present) context has enduring effects on lexical-semantic representations. When listeners encountered ambiguous words such as 'fans' without any biasing context, they were 30-40% more likely to interpret the words as referring to the subordinate (less common) 'supporter' meaning if they heard that subordinate meaning in a sentence (e.g. 'the footballers were greeted warmly by the adoring *fans*') 20 minutes earlier (Rodd et al., 2013). Hence, just a single subordinate encounter increased the likelihood with which it is later used and therefore presumably the availability of that meaning. These effects of experience with word-meanings are therefore not caused by activation from present context per se, but are the long-term (up to 40 minutes, Rodd et al., 2016, Experiment 2) enduring changes in meaning availability as a result of recent experience.

Until now, word-meaning priming has only shown a positive boost for the primed meaning (Rodd et al., 2016; Rodd et al., 2013; Chapter 2). However, according to Gaskell and Dumay (2003), in order to conclude that changes to lexical-semantic representations have been integrated into the lexicon, one must show that learning new information (i.e. recently encountering the riverside meaning of 'bank' and updating availability to reflect the higher likelihood of its subsequent recurrence) interferes with the availability of, or access to, existing information (i.e. inhibits or interferes with the availability of the alternative financial meaning). Gaskell and Dumay (2003) found that learning the novel word 'cathedruke' as a competitor of the existing word 'cathedral' did not impede performance with 'cathedral' on a lexical decision task immediately after learning, but did impede performance when tested after five days. This is evidence that the newly learned meaning was integrated into the lexicon with more time between learning and test, as it affected the availability of other words competing for selection. This learning and integration does not typically happen immediately, but occurs after a period in which the new information can be consolidated. To draw strong conclusions about whether word-meaning priming is driven by changes to lexical-semantic representations in the lexicon itself, it must be shown that this learning has an impact on competing (i.e. unprimed) meanings of the same ambiguous word.

The benefit for researchers of testing meaning availabilities after a priming manipulation is that immediate context itself and any subsequent learning/updating based on that context are separated, since the availability of representations is measured on the word in isolation. It is not currently known whether this enduring increased availability, as a result of experience, is accompanied by an enduring reduced availability of the unprimed meaning. Yet this issue is theoretically important, since the storage of lexical-semantic representations has implications for language learning, and is a fundamental aspect of any model of semantic ambiguity resolution. With this word-meaning priming method, we can therefore investigate the learning mechanism underlying the construction, maintenance and/or updating of lexical-semantic representations.

Experimental method

Word association tests reveal that people are more likely to interpret an ambiguous word in light of its subordinate meaning when that meaning was encountered up to 30 minutes previously (Rodd et al., 2013). So far in this thesis, one repetition of either the subordinate or dominant meaning has been shown to bias interpretation towards that meaning (Experiments 1, 2, 3). Whilst, three spaced subordinate repetitions *further* bias interpretation in word association (Experiments 2 & 3), three massed repetitions provide no such benefit over one repetition (Experiment 3). As discussed above, it is possible that word-meaning priming is driven by an increase in the availability of the primed meaning coupled with a decrease in the availability of the unprimed meaning. If this were the case, it would suggest that priming of one meaning could interfere with the availability of the alternative, unprimed meaning. This would indicate that the multiple representations of an ambiguous word would be necessarily linked and that priming effects occur due to enduring increased availability of the selected, primed meaning *and* enduring decreased availability of the unselected, unprimed meaning.

However, the standard experimental method used to measure dominance and word-meaning priming – the word association test – is not sufficiently sensitive to distinguish between an increase in one meaning and a decrease in the other meaning, and was not designed to do so. The word association test gives only one data-point per ambiguous word, which means that any differences between the availability of the dominant and subordinate meanings, as a function of subordinate priming, are combined into a single word response. When responses are combined and turned into proportional values, the relative difference of word-meaning availabilities becomes the basis of the word association measure. The relative difference between primed and unprimed meaning availability would be the same with (for instance) a 20% boost in primed meaning availability and a 20% decrease in unprimed meaning availability. As a result, the experiments in the present chapter use a new semantic relatedness measure, with either the dominant or subordinate meaning of an ambiguous word depicted as a probe on each trial, to investigate whether or not priming with one meaning necessarily reduces the availability of the unprimed meaning.

The semantic relatedness picture probes allow for the separation of the two alternative meanings of an ambiguous word; responses can be collected from either the dominant or the subordinate meaning alone. The test provides the participant with a picture of one of the possible interpretations of the word and can therefore measure whether or not the picture meaning is consistent with the participants' initial interpretation of the word. If it is inconsistent, participants might re-disambiguate the word to make sense of it, thus resulting in slower or less accurate responses. Experiment 4 showed that the semantic relatedness test was sensitive to dominance: participant responses were, on average, significantly (112ms) faster and significantly (7.6%) more accurate for the dominant than for the subordinate picture. Where the picture is consistent with the participants' interpretation (i.e. it is the expected and most available meaning), reaction times are faster and accuracy is increased. Where the picture is inconsistent with the participants' interpretation (i.e. it is the unexpected and less available meaning), reaction times are slower and accuracy is reduced.

This detection of dominance means the semantic relatedness test could be at least as successful as the standard word association test at detecting priming effects (used to measure word-meaning priming in Rodd et al. 2016; Rodd et al., 2013; and throughout Chapter 2, Experiments 1, 2 & 3). If the subordinate meaning is primed, its availability is increased and, at test, the participant is more likely to expect this primed subordinate meaning than if priming had not occurred. Hence, responses to the subordinate, prime-consistent picture should be faster and more accurate than an unprimed subordinate picture baseline (replicating the standard word-meaning priming effect shown in word association; Rodd et al., 2016; Rodd et al., 2013; Chapter 2 of the present thesis). This test therefore has the capability to detect word-meaning priming.

If subordinate priming only increases availability of the subordinate meaning (and does not affect dominant meaning availability) and the unprimed dominant, prime-inconsistent picture is presented at test, responses to the dominant picture will be no different than if subordinate priming had not occurred. This could mean one of two things. First, it could mean that word-meaning availabilities update independently, such that priming one meaning and boosting its availability does not reduce availability of the alternative, unprimed meaning, supporting the reordered access model, which predicts this outcome (Duffy et al., 1988). Second, it could mean that the positive boost in availability from word-meaning priming does not reflect changes to the lexical-semantic representations in the lexicon itself, but that word-meaning priming operates via a mechanism that lies outside of the lexicon. Learning new information about word-meanings cannot therefore interfere with existing information stored in the lexicon. If responses to the unprimed meaning are not affected by priming, then further experiments will be required to disentangle these two possible causes.

Conversely, if subordinate priming both increases availability of the subordinate meaning *and* simultaneously reduces availability of the dominant meaning, when the dominant prime-inconsistent meaning is presented, responses will be slower and less accurate compared than if priming had not occurred. This means that this can test whether the multiple meanings of an ambiguous are linked, such that increasing the availability of one meaning necessarily comes at the cost of a reciprocal reduction of availability of an alternative meaning.

The set of five experiments in this chapter combine word-meaning priming and the newly developed semantic relatedness picture test to further investigate whether priming of one meaning reduces the availability of the alternative, unprimed meaning. All five experiments (5, 6, 7, 8 & 9) were run using Gorilla experiment software (Cauldron, www.gorilla.sc, 2017). Whilst participants in Experiment 9 completed the experiment in the laboratory, participants in Experiments 5, 6, 7 and 8 were recruited online and took part online. This was to achieve faster and more cost-effective data collection (online experiments benefit from these factors without a significant compromise in the quality of data; Casler et al., 2013; Gosling, Vazire, Srivastava, & John, 2004). See Chapter 5 for a more detailed discussion of the advantages of online data collection.

Each of the five experiments uses the standard word-meaning priming procedure (Chapter 2) of a prime phase, which varies across experiments in the primed meaning and the number of repetitions of this meaning. However, across all experiments participants subsequently encounter the same filler task puzzle and the same semantic relatedness test phase (where the standard word association test is replaced with the semantic relatedness picture test). Within an experiment, the prime phase only exposes participants to a single primed meaning (i.e. either subordinate or dominant), whereas the test phase exposes participants to either the subordinate or the dominant meaning, which varied across items (with the exception of Experiment 9 for reasons outlined therein). Varying dominance at test allows for the effects of recent experience on meaning availability of both meanings to be examined within a single experiment, and therefore any reciprocal changes in availability of the different meanings. See Table 4 for a summary of the designs of each of the five experiments.

Table 4. A summary of the designs of the five experiments in Chapter 4 using the semantic relatedness picture test (Experiments 5 – 9). Unprimed baseline also included in all experiments.

Expt.	Prime Meaning	No. Prime Repetitions	Prime Repetition Distribution	Prime-Test Delay (mins)	Picture at Test
5	Subordinate	1	---	13	Dominant & subordinate
6	Dominant	1	---	7	Dominant & subordinate
7	Subordinate	3	Spaced	18	Dominant & subordinate
8	Subordinate	3	Massed & spaced	18	Dominant & subordinate
9	Subordinate	1, 3	Massed & spaced	17	Subordinate only

Experiment 5 – one subordinate repetition

Experiment 5 involved a subordinate prime phase, where participants encountered the subordinate meaning of each ambiguous word in a single sentence. At test, each ambiguous word was presented to participants accompanied by either the dominant or subordinate picture. Participants were asked to make a response about whether or not the picture was related in meaning to the word, which resulted in two dependent measures: reaction time and error rate. See Figure 11 for the order of the tasks in the experiment.

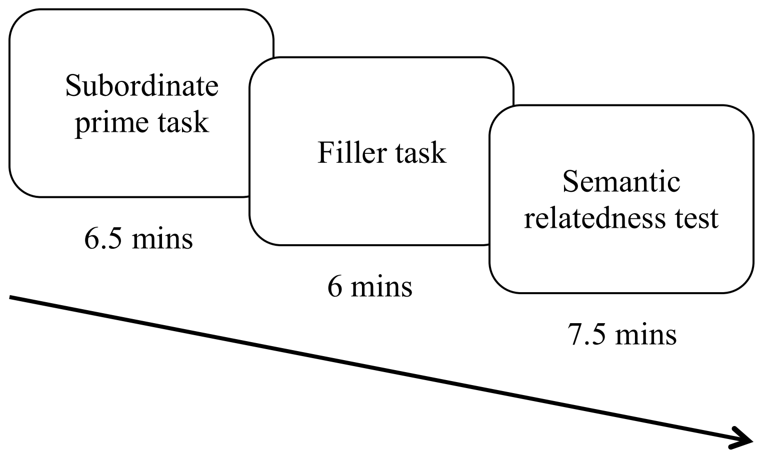

Figure 11. Experiment 5 task order, with the mean duration of each task. The average delay between an ambiguous word being presented in the prime phase and in the test phase (prime-test delay) is 13 minutes.

The present experiment had two aims. The first was to confirm that the newly developed semantic relatedness task (see Chapter 3 for details) was sensitive enough to detect subordinate priming. If this task can detect priming, responses to the subordinate picture should be faster and/or more accurate following an encounter with the subordinate meaning, than if the subordinate meaning had not been encountered. Based on robust word-meaning priming effects (e.g. Rodd et al., 2016; Rodd et al., 2013) and sizeable dominance effects in semantic relatedness (112ms and 7.6% accuracy advantage for the dominant meaning, compared to the subordinate meaning), it was predicted that reaction times and error rates to the subordinate picture would be reduced following subordinate priming.

The second aim was to investigate whether the subordinate prime repetition would reduce the availability of the unprimed, dominant meaning. As discussed previously, the literature provides little guidance as to the outcome of the unprimed meaning as a result of priming, hence predictions based on evidence were not possible. However, there were two possible outcomes for performance on the dominant picture at test. If dominant picture performance is the same regardless of whether or not there was prior subordinate priming, then increasing the availability of the subordinate meaning does not come at the cost of dominant availability. This would show that the representations for the alternative meanings of an ambiguous word are independent. However, if RTs and errors to the dominant picture are increased after subordinate priming, compared to the unprimed baseline, then increasing the availability of the subordinate meaning does come at the cost of reducing dominant availability. This would show that the alternative meanings of an ambiguous word interact and therefore their representations are related to some extent.

Method

Participants

One-hundred-and-twelve native British English speakers participated in the current experiment (60 females; mean age = 29.63, range = 18-44). All participants reported that they had no language, hearing or vision impairments (other than corrected-to-normal vision) and had lived in the UK for the majority of their lives, speaking English as their first language from birth. Additionally, in-built features of Gorilla (gorilla.sc; Cauldron) verified that the participants were in the UK (IP geolocation), and had a minimum internet connection speed of 15Mbps (ensuring adequate speed for the reaction time task), at the time of testing. Participants were recruited via Prolific (Prolific Academic Ltd., www.prolific.ac, 2016) and paid the standard rate at the time of £6/hour.

Design

This experiment had a within-subjects design with two independent variables: subordinate priming and picture meaning, each with two levels. The dependent variables were the reaction times and error rates of responses to the pictures at test.

The first independent variable was subordinate priming (levels: subordinate unprimed and subordinate primed) where participants encountered half of the experimental ambiguous words in the prime phase, each with a single sentence disambiguating the ambiguous word towards its subordinate meaning. At prime, two versions were created (A and B) where the 30 primed words for half of the participants were then the 30 unprimed words for the other half of the participants and vice versa. This ensured that each item appeared in both priming conditions but across different participants. The unprimed half of the experimental ambiguous words were not encountered in the prime phase but were later introduced in the test phase to provide an unprimed baseline.

The second independent variable was semantic relatedness picture meaning (levels: dominant and subordinate; factor referred to as "picture meaning" for brevity) where participants encountered half of the experimental ambiguous words at test paired with a picture of the dominant meaning and half paired with a picture of the subordinate meaning. At test, two subversions were created (1 and 2) where the words paired with the dominant picture for half of the participants were then the words paired with the subordinate picture for half of the participants, and vice versa. Four versions were therefore required at test (A1, A2, B1, B2), where each prime version was coupled with each of the test subversions. This meant that at test, across participants, each word appeared as both a primed and an unprimed trial and a dominant picture and subordinate picture trial. All participants contributed to each of the four conditions (subordinate unprimed - dominant picture, subordinate unprimed - subordinate picture, subordinate primed - dominant picture, and subordinate primed - subordinate picture) but for different ambiguous words. All filler trials were identical across versions.

As is standard with semantic relatedness tests, unrelated responses were not to be analysed, since the reason for 'unrelated' being given as a response is unclear. To maximise the trials for which data could be analysed (i.e. maximise the number of related trials at test), all experimental items were paired with related pictures, which provided two test stimulus types: 30 primed, ambiguous, related trials and 30 unprimed, ambiguous, related trials. In addition, 28 ambiguous and 22 unambiguous sentences were included at prime as fillers. These fillers served two purposes: first, the 22 unambiguous fillers reduced the salience of ambiguity across prime sentences and second, the 28 ambiguous fillers and 10 of the unambiguous fillers provided trials that, at test, could be paired with unrelated pictures and therefore trials that could be removed from analyses without lowering the number of experimental items to analyse. The remaining 12 unambiguous fillers from the prime were paired with related pictures at test so that not all unambiguous trials were unrelated at test (again, these were not analysed as they were fillers). This design allowed items in the test phase to differ across the three dimensions of: 1) priming, 2) ambiguity and 3) word-picture relatedness at test. As a result, knowing whether or not a word was primed/unprimed or ambiguous/unambiguous could not help a participant make a faster or correct response to either the dominant or subordinate picture at test. Finally, the considerable mixture of stimulus types at test would make it highly unlikely that participants would be aware of the inequality between the types at test. See Table 5 for details of stimulus types.

Table 5. Details of ambiguous word stimuli at prime and test phases in Experiment 5.

Prime	**Test**		
Word Qualities	**Stimulus Type**	**Word-Picture Relatedness**	**No. of Stimuli**
Primed, ambiguous	Experimental	Related	30
Primed, ambiguous	Filler	Unrelated	28
Primed, unambiguous	Filler	Related	12
Primed, unambiguous	Filler	Unrelated	10
Unprimed, ambiguous	Experimental	Related	30
Unprimed, ambiguous	Filler	Unrelated	0
Unprimed, unambiguous	Filler	Related	0
Unprimed, unambiguous	Filler	Unrelated	16

Note: the numbers in the cells are not equal and two cells of the stimulus-combination types do not have any stimuli. This is due to constraints with the number of words that had related pictures and the number of filler pictures available. Unprimed words at prime are in grey to emphasise that they were not encountered until the test phase. Although not listed here, one additional example and 5 practice trials were also created for use throughout the experiment.

Materials

All word and sentence stimuli were taken from Rodd et al. (2016, Experiment 2)[14], and were also used in Experiments 2 and 3 of the present thesis. For the present experiment, this included 60 experimental ambiguous words (used in Experiment 4), 28 filler ambiguous words and 38 filler unambiguous words for use either in the prime or test phase of the experiment.

The subordinate prime task used the sentences created for Rodd et al. (2016, Experiment 2), where a sentence disambiguated each of the 60 experimental ambiguous and 28 filler ambiguous words towards its subordinate meaning with prior

[14] These stimuli were created as part of this PhD.

context. Twenty-two of the unambiguous filler words were included in the prime phase, where the unambiguous word was used as a basis for a sentence (using the unambiguous word). As in Experiments 2 and 3, each sentence was coupled with a related or an unrelated probe word (50% sentences with related probes, 50% with unrelated for experimental trials; unrelated probes were not related to any meaning of the ambiguous word). The target ambiguous words did not appear in any other sentences, instructions, or as any of the probe words throughout the experiment.

For the filler task, a Towers of Hanoi puzzle was used. Due to Copyright rules practiced by Aardman Animations Ltd. (creators of the 'Shaun the Sheep' animations, which were used as the filler task in Chapter 2), it was not possible to distribute the animations online and use them as the filler task for the semantic relatedness experiments. The Towers of Hanoi task was chosen as a suitable replacement because it was deemed to be a similar task, since, like the 'Shaun the Sheep' animation, it was engaging and did not involve language. This task required participants to move disks from one "tower" to another whilst maintaining their size order i.e. at the start, the largest disk was at the bottom and the smallest disk was at the top of the tower on the left and must end up in this configuration on the right-most tower. The task started with three disks, although difficulty was progressively increased with an additional disk after each level was completed. This task was provided with Gorilla experiment software (Cauldron, www.gorilla.sc, 2017) and was ideal for use as a filler task between the prime and test, as it provided a time delay that could be specified, did not involve language (and therefore should not interfere with word-meaning priming) and was sufficiently cognitively demanding as to distract from the prime-test manipulation.

For the semantic relatedness picture test there was a total of 126 items all from Experiment 4: 60 ambiguous experimental and 66 filler words, recorded individually by the same speaker as the prime sentences. The 60 experimental ambiguous words each had a picture for both the dominant and subordinate meaning; the presented picture was dictated by the version and was, of course, always related. The 12 filler primed, unambiguous, related words were coupled with a single picture that depicted its meaning. All other trials at test were unrelated; the 28 primed ambiguous, 10 primed unambiguous and 16 unprimed unambiguous words were each coupled with a single picture that was unrelated to its meaning(s). In addition, 6 practice trials were created where an unambiguous word was coupled with an unrelated picture for use at the start of the test task. In total at test, 60% of trials were primed, 67% of trials were ambiguous and 55% of trials were related.

Procedure

The experiment was presented online to participants using Gorilla experiment software (Cauldron, www.gorilla.sc, 2017). Each participant was randomly assigned to one of the four versions of the experiment. After giving their informed consent, participants' demographic data were collected and instructions for the experiment were displayed on-screen.

Subordinate Prime Task

For each trial, participants saw a fixation cross while they heard a sentence. Upon sentence offset, the fixation cross was replaced with the probe word. Participants were asked to respond as quickly and accurately as possible to the probe word by either pressing the 'r' key for related in meaning to, or the 'u' key for unrelated in meaning to, the sentence. Response times longer than 3000ms prompted a message encouraging faster responses on subsequent trials. The key press response triggered a screen with only a central fixation cross for 1000ms followed by the next trial. One practice trial was given at the start of the task, with feedback. The order of sentence presentation was randomised separately for each participant.

Filler Task

The instructions for the Towers of Hanoi task were as follows (a picture example was also shown to participants): 'In The Towers of Hanoi, your objective is to move all of the disks on the left-most tower on to the right-most tower. You can only move one disk at a time and a disk may only be placed on top of a larger disk. The objective is to complete the exercise using the smallest number of moves possible.' Participants were then required to click a button on-screen to begin the task. All participants started with 3 disks and, upon successful completion, automatically progressed to the next level of difficulty. The filler task ended automatically and progressed to the next task after 6 minutes (see Figure 11 for procedure timings).

Semantic Relatedness Picture Test

For each semantic relatedness test trial, participants saw a fixation cross on an otherwise blank screen whilst they heard a word. A probe picture was then presented immediately after word offset in place of the fixation cross. Participants were asked to respond as quickly and accurately as possible to the probe picture by either pressing the 'r' key for related in meaning to, or the 'u' key for unrelated in meaning to, the word. Response times longer than 1500ms prompted a message encouraging faster responses on subsequent trials. The key press response triggered a blank screen with a central fixation cross for 1000ms followed by the next trial. One practice trial was given at the start of the task, with feedback. All trials were randomised with the exception of 5 filler trials at the start, which allowed participants to become accustomed to the task.

Post-Experimental Task

At the end of the experiment, participants were asked two questions to investigate their awareness of the priming manipulation: 'What do you think the aim of the experiments was?' and 'How many words from the word association do you recognise from the tasks earlier in the experiment?' to measure awareness of the priming manipulation and investigate its impact on priming.

Task Checks

All participant responses were checked for prime and test accuracy and for total time spent on the experiment. Participants were excluded if they had less than

70%[15] accuracy on either the prime or test task (or both), or if they spent longer than the time limit of 28 minutes on the experiment (including consent sheet, demographics questions and awareness questions). This was because low accuracy at prime would indicate that participants were not attending to the sentences enough to respond to the probe word accurately, suggesting that they would not have been susceptible to the priming in the sentences. Low accuracy at test would indicate that participants were not concentrating on their responses, rendering their interpretation unreliable. Spending over the maximum time on the experiment would increase the prime-test delay too far beyond the average 13 minutes, which could reduce the likelihood of detecting priming and therefore introduce noise in the data compared to the participants who completed the experiment in the expected time. Whilst 112 participants completed the experiment, only the data from 102 participants were analysed: ten participants were excluded for meeting one or more of the above exclusion criteria. In addition, all items were checked for prime and test accuracy and were excluded if their accuracy, averaged across participants, was less than 70%. This resulted in the exclusion of one item ('iron') for all participants, leaving a total of 59 items in the analyses.

Results

The RT and error data were modelled using mixed effects modelling: linear mixed effects modelling with the lmer function for the RTs and logistic mixed effects modelling with the glmer function for the errors; both functions from the lme4 package (version 1.1-7; Bates et al., 2014) in R (version 3.3.1; R Core Team, 2016). Mixed effects modelling is the most appropriate analysis method for these data since it takes the within-subject and within-item dependencies into account within a single model (Jaeger, 2008). Note that error rate, as opposed to accuracy, is reported for all experiments in the present chapter, since it allows RT and error graphs to be consistent in the direction in which improved performance is shown.

---

[15] This minimum prime/test accuracy requirement is lower than the minimum prime accuracy requirement used in Experiments 1, 2 and 3 because the overall difficulty of the present experiment (which involved semantic relatedness both at prime and test, as well as a more cognitively demanding filler task) was deemed greater than the overall difficulty of these earlier experiments.

For both the RT and error analyses, a full model was built that included a full fixed effects structure (subordinate priming and picture meaning factors as well as the interaction term) and full random effects structure (subjects and items slopes, intercepts and the correlations between these), as recommended to protect against inflated Type I error for suitably powered designs (Barr et al., 2013). A model comparison approach (Baayen et al., 2008) was then used to test the significance of individual factors, which removed (from the fixed effects but not the random effects) one factor at a time and in each case compared this reduced model to the full model using a likelihood ratio test to assess whether the inclusion of the factor significantly improved the model fit. However, when more than one (categorical) factor is present in a model and one is removed for model comparisons, it is not clear whether R fully removes this factor if it has used the in-built automatic coding of factor levels. To ensure that the factor was fully removed when required, each factor was manually deviation-coded (subordinate priming: unprimed = -1/2, primed = 1/2; picture meaning: subordinate = -1/2, dominant = 1/2; the interaction term was specified as a separate coded factor: the multiplication of the subordinate priming and the picture meaning factors). This resulted in a model with three factors: subordinate priming, picture meaning, subordinate priming by picture meaning interaction.

Reaction Time Analyses

As the subject means in Figure 12 indicate, RTs to the subordinate picture at test seem to be faster after subordinate priming, compared to the unprimed baseline, but there seems to be no pattern to suggest that subordinate priming has an impact on RTs to the dominant picture. Overall, RTs seem to be faster for dominant pictures than for subordinate pictures.

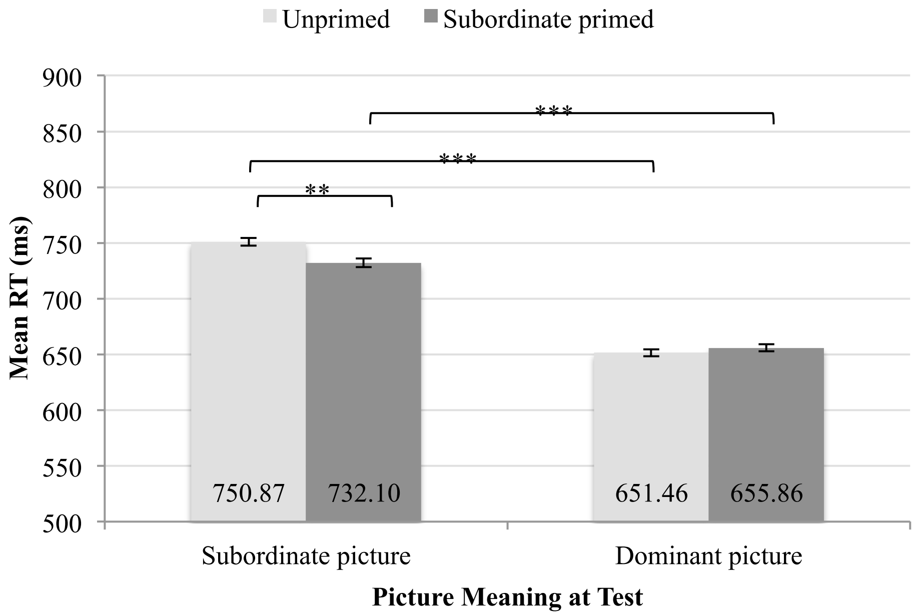

Figure 12. Mean by-subjects reaction times for Experiment 5. Responses for both the subordinate and dominant meaning picture, following either no priming or one subordinate prime repetition. Significance level indicated with asterisks (** <.01, ***<.001) for theoretically important contrasts (for all figures in this chapter). Error bars are adjusted for the within-subjects design[16].

Reaction times (RTs) for filler trials and incorrect trials were removed from the data, as were RTs less than 300ms and greater than 1500ms as these suggest accidental key presses or unusually slow responses (respectively). In addition, RTs less than or greater than 3 standard deviations from a participant's mean RT were discarded. The full random and fixed effect model converged for the raw RTs. However, since the assumptions of normality and homoscedasticity were violated, the RT data were inverse and log transformed and the same full models were run on these transformed RTs separately. The inverse transformation showed to be the most

[16] Whilst mixed effects modelling was used to analyse these data and those in Experiments 6, 7, 8 and 9, it does not provide "interpretable" means, hence the results figures for these experiments show the subject means. For this reason, there may be some slight discrepancies between the results of mixed effects analyses, which account for both item- and subject-specific effects, and the results implied by the subject means in the figures. These do not alter the pattern of results in any case.

suitable, since it did not violate the assumptions of normality and homoscedasticity, and was therefore used as the dependent variable for all RT analyses in the present experiment.

The model comparison approach revealed that there was no significant main effect of subordinate priming, (X^2 (1) = 2.644, p = .104; where there was no significant difference between a model with and a model without this factor). This suggests that overall, across picture meaning conditions, participant RTs to pictures did not change depending on whether or not they had encountered the subordinate meaning of that word in the prime phase. However, comparisons showed a significant main effect of picture meaning at test (X^2 (1) = 21.107, p < .001), showing that across subordinate priming conditions, participants were generally faster at responding to dominant than subordinate pictures. There was also a significant interaction between subordinate priming and picture meaning (X^2 (1) = 11.108, p < .001), indicating that the effect of subordinate priming differed depending on whether they encountered the dominant or subordinate picture at test.

Due to the significant interaction, simple effects analyses were conducted to investigate the nature of the significance. These were implemented by creating subsets of the full inverse RT data set such that, for each subset/simple effect, only one level of one of the factors was included, with the other factor having both levels. For instance, to test the simple effect of subordinate priming just for the subordinate picture (i.e. not including the dominant picture condition), a subset of the picture meaning factor was created that only included the subordinate picture trials. Then a model was created which only included the subordinate priming factor[17] but with the subsetted data for the subordinate picture condition. As with the analyses above, a model comparison approach was used to test the significance of the single factor (subordinate priming). This removed the single fixed effect factor and compared this null model to the model with it included using a likelihood ratio test to assess whether its inclusion significantly improved the model fit to the subordinate picture data. This

[17] Note that this means that simple effects models had different random effects structures to models for the main effects and interaction, since the simple effects models each included only one factor as a fixed effect (and therefore in the random effects).

test therefore allowed the analysis of whether there was an effect of subordinate priming on subordinate picture test trials alone.

A set of four simple effects analyses was conducted, with Bonferroni-adjusted p values[18] for *post-hoc* comparisons. First, the effect of subordinate priming for the subordinate picture at test was significant (X^2 (1) = 9.448, p = .008), showing that one subordinate repetition at prime was sufficient to speed RTs to the subordinate picture at test. Second, the effect of subordinate priming for the dominant picture at test was not significant (X^2 (1) = 1.381, p = .959), showing that one subordinate repetition did *not* slow RTs to the dominant meaning picture at test. This suggests that subordinate priming did not interfere with availability of the alternative, dominant, meaning. Third, there was a significant effect of picture meaning for the subordinate unprimed condition (X^2 (1) = 24.025, p < .001), indicating faster RTs to the dominant picture than to the subordinate picture for unprimed trials. Fourth, there was a significant effect of picture meaning for the subordinate primed condition (X^2 (1) = 15.905, p < .001), indicating faster RTs to the dominant picture than to the subordinate picture even after subordinate priming.

Error Rate Analyses

The subject means in Figure 13 indicate that subordinate priming slightly reduces percentage error for both the subordinate and dominant picture, although the percentage error for the dominant picture is overall lower than for the subordinate picture.

[18] Bonferroni-adjustment: the p values reported for the simple effects analyses have been multiplied by the total number of familywise simple effects in the experiment (Experiment 5 in this case) to control for the familywise error rate.

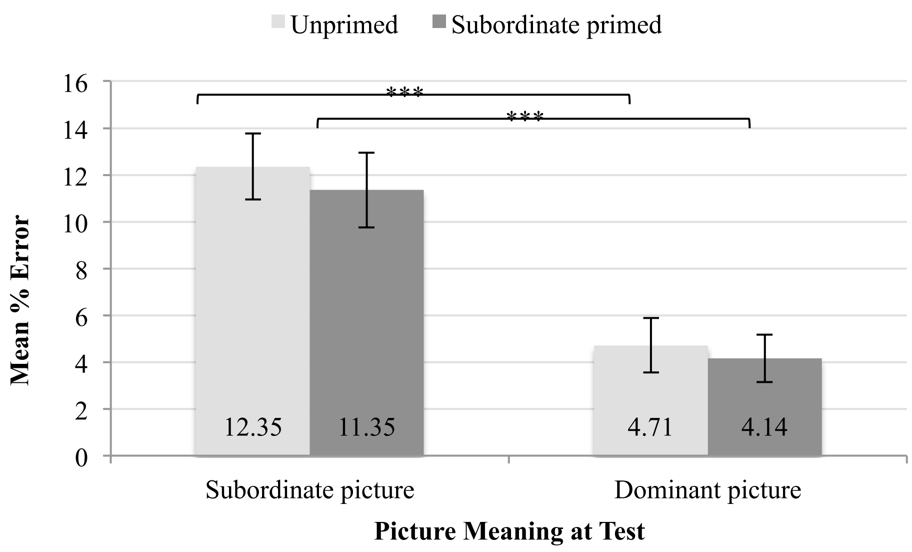

Figure 13. Mean by-subjects percentage error for Experiment 5. Responses for both the subordinate and dominant meaning picture, following either no priming or one subordinate prime repetition. Significance level indicated with asterisks (***<.001). Error bars are adjusted for the within-subjects design.

Whilst filler trials and trials where responses were than 300ms and greater than 1500ms were discarded, as in the RT analyses (as these suggested accidental key presses or unusually slow responses, respectively), incorrect trials were necessary for error analyses. Other than the use of the glmer function (for logistic models), the analysis approach was identical to that of the RTs where a model with a maximal random effects structure model was built with fixed effects for the subordinate priming and picture meaning factors as well as the interaction term. Again, a model comparison approach was used to test the significance of individual factors.

The full model failed to converge across all tests of main effects and interactions (most likely due to the complex random effects structure). Hence the recommended protocol for dealing with non-convergence (Barr et al., 2013) was followed, where the random effects structure is simplified by removing one random effect term at a time (correlations removed first, then intercepts, then slopes; the subject or item term that explained the least variance was removed first) until all of these nested models also converged. The second fullest model converged (correlations between subject and items slopes and intercepts removed) for all nested models.

The model comparisons revealed no significant main effect of subordinate priming (X^2 (1) = 0.128, p = .721), suggesting that, across picture meaning conditions, participant errors to pictures were unaffected by whether or not they had encountered the subordinate meaning of that word in the prime phase. There was a significant main effect of picture meaning (X^2 (1) = 22.627, p < .001), showing that, across subordinate priming conditions, there were fewer errors when responding to dominant pictures, compared to subordinate pictures. However, there was no significant interaction between subordinate priming and picture meaning conditions (X^2 (1) = 0.185, p = .667), indicating that the effect of subordinate priming was the same regardless of whether the participant responded to the dominant or subordinate picture presented at test. Thus, subordinate priming did not significantly affect error rates overall.

The four simple effects analyses confirmed the pattern from the main effects. As with the RT analyses, the Bonferroni-adjusted p values are reported. First, the effect of subordinate priming for the subordinate picture at test was not significant for the errors (X^2 (1) = 0.551, p = .999), suggesting that one subordinate repetition at prime was *not* sufficient to reduce error rates to the subordinate picture at test. Second, the effect of subordinate priming for the dominant picture at test was not significant (X^2 (1) = 0.399, p = .999), showing that one subordinate repetition did not increase error rates to the dominant picture. Again, this suggests that subordinate priming did not interfere with availability of the alternative, dominant, meaning. Third, there was a significant effect of picture meaning for the subordinate unprimed condition (X^2 (1) = 14.700, p < .001), indicating fewer errors when responding to the dominant picture than to the subordinate picture for unprimed trials. Fourth, there was a significant effect of picture meaning for the subordinate primed condition (X^2 (1) = 19.799, p < .001), indicating fewer errors to the dominant picture than to the subordinate picture even after subordinate priming.

Awareness Analyses

There were two awareness measures: awareness of experimental aim and awareness estimate. One experimenter (HNB) coded the responses to the awareness of experimental aim question. If the participant demonstrated some, or full, correct awareness of the experimental aim (e.g. 'to see if the original sentences influenced my later associations'), their response was coded as aware, whereas if they demonstrated little/incorrect or no awareness of the aim (e.g. 'how large or small people's semantic fields are'), their response was coded as unaware, hence these data were dichotomous. Ninety-eight participants were unaware of the aim (for subordinate picture test condition only, primed and unprimed levels combined[19]: RT *mean* = 742.83ms, *SD* = 98.96ms, percentage error *mean* = 11.91%, *SD* = 8.67%) and only 4 participants were fully/partially aware of the aim (for dominant picture test condition only: RT *mean* = 712.65ms, *SD* = 76.23ms, percentage error *mean* = 11.09%, *SD* = 7.17%), hence there was an insufficient number of participants (only 3.9%) in the "aware" category to run an analysis to examine whether priming interacts with awareness of aim. Additionally, it was not possible to analyse the awareness estimate data since the sliding scale was erroneously capped at a total of only 72 words as opposed to the correct total of 131 words. This meant that it was not possible for participants to give an accurate estimate and therefore their responses on this task were not meaningful.

Discussion

Importantly, the results replicate the results from Experiment 4 in showing that, on unprimed trials, performance on the dominant meaning is significantly faster (by 99ms) and significantly more accurate (by 8%) than the subordinate meaning. Aside from this, there were two aims of Experiment 5. First, it was necessary to ensure that the semantic relatedness picture test could detect subordinate word-

[19] Only the prime-test congruent meaning condition (subordinate picture) was included, i.e. only the picture meaning that was consistent with the primed meaning. This was because we were interested only in whether or not participants were aware of the repetition of a meaning from the prime phase in the test phase, therefore it is irrelevant for the dominant picture for this experiment, since this dominant meaning was not primed.

meaning priming. The results show that the test was indeed sensitive to the word-meaning priming; participants were significantly faster to respond to the subordinate picture if they had encountered that subordinate meaning in the prime phase, compared to the unprimed baseline[20]. Hence, the subordinate meaning was significantly more available following a single encounter with the subordinate meaning earlier in the experiment than if the subordinate meaning had not been encountered earlier. Whilst subordinate priming numerically reduced participant error rates to subordinate pictures as well, this difference was not significant. Regardless, measuring the speed of a semantic relatedness response was sufficient to observe priming effects. This suggests that the semantic relatedness test is a suitable alternative to word association for measuring word-meaning priming.

The second aim was to investigate whether a priming-driven increase in subordinate meaning availability caused a decrease in the availability of the unprimed dominant meaning. Compared to the unprimed baseline, neither the RTs nor the error rates for the dominant picture were significantly increased following subordinate priming. This suggests that an increase in the availability of one meaning does not necessitate a decrease in the availability of the alternative meaning. This finding seems to support the conclusion from Colbert-Getz and Cook (2013) and the model class including the reordered access model (Duffy et al., 1988), where the availabilities of different word-meanings are independent and do not operate based on reciprocal changes. However, it is inconsistent with findings by Chen and Boland (2008), and what would be suggested by the current view of the connectionist model (e.g. Rodd et al., 2004; Rodd et al., 2013). The current findings seem to suggest that changes to connection strengths might be more independent and specific to the encountered meaning rather than to all available meanings of the word.

However, the null effect of subordinate priming on dominant meaning availability may have arisen for two reasons. Either because (a) there really is no effect of priming on the unprimed meaning, or (b) listeners have so much more

[20] Note that it was the simple effects that were crucial; the absence of a significant main effect for subordinate priming was simply because the priming effects on the subordinate and dominant pictures cancelled each other out.

experience with dominant meanings that one subordinate encounter does not have an observable impact on dominant availability at test. These two possibilities are confounded in the present experiment; it is always the dominant meaning that is the unprimed meaning. Thus any impact of priming on unprimed (dominant) meaning availability is tested on the strongest, most available (dominant) meaning. Since this meaning is most available, it is unlikely that a single subordinate encounter would interfere with dominant meaning availability.

The present experiment was designed in this way because subordinate priming has consistently been a successful manipulation (e.g. Chapter 2), since its lower availability leaves it more susceptible to a boost in availability from recent experience. However, the two possible explanations for the null effect of priming on the unprimed meaning availability need to be further investigated to tease them apart. If a dominant meaning priming manipulation is used (shown to be an effective manipulation in Experiment 1) instead of the subordinate meaning, then the subordinate meaning becomes the unprimed meaning and is the weaker meaning. A recent experience with the dominant meaning might have a more observable impact on subordinate meaning availability at test, since the subordinate meaning is less available, potentially leaving it susceptible to interference from the competing (primed) dominant meaning. Experiment 6 will therefore investigate whether priming with the dominant meaning increases dominant availability and decreases subordinate availability.

Experiment 6 – one dominant repetition

The previous experiment (Experiment 5) involved a subordinate prime phase and, after a 6-minute filler task to create a 13-minute prime-test delay, each ambiguous word was presented to participants accompanied by either the dominant or subordinate picture. For Experiment 6, two elements of that original design were changed. First, the subordinate prime phase was replaced with a dominant prime phase, in which participants encountered the dominant meaning of ambiguous once in context. Second, the filler task was removed to reduce the prime-test delay, since Experiment 1 found a significant dominant priming effect that was of a comparable magnitude to subordinate priming but at a shorter delay (9-minute average prime-test delay for dominant priming compared to 30 minutes for subordinate priming; the removal of the filler task for Experiment 6 reduced the present prime-test delay from 13 minutes to 7 minutes, comparable to that of Experiment 1). This suggested that dominant priming effects might have been weaker than subordinate priming effects at equivalent, longer, delays. Hence, the removal of the filler task aimed to maximise any dominant priming effects to increase the chances of these effects interfering with unprimed subordinate meaning availability. See Figure 14 for the order of the tasks in the experiment.

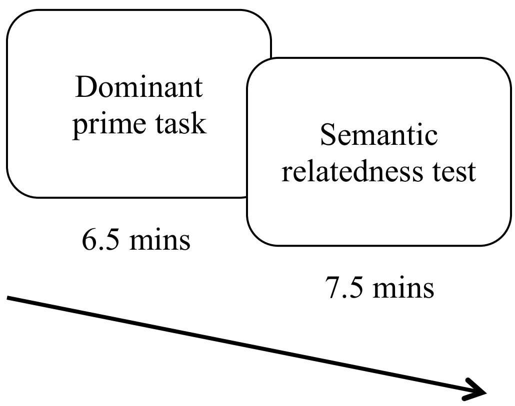

Figure 14. Experiment 6 tasks, with the mean duration of each task. The average delay between an ambiguous word being presented in the prime phase and in the test phase (prime-test delay) is 7 minutes.

The aim was to investigate whether the dominant prime repetition reduced the availability of the unprimed, subordinate meaning. If performance to the subordinate picture is the same regardless of whether or not there was prior dominant priming, then increasing the availability of the dominant meaning does not come at the cost of subordinate availability. This would support the findings from Experiment 5 and would suggest that the representations for the alternative meanings of an ambiguous word are independent. However, if RTs and errors to the subordinate picture are increased after dominant priming, compared to the unprimed baseline, then increasing the availability of the dominant meaning does come at the cost of reducing subordinate availability. This would show that the alternative meanings of an ambiguous word can interact and it would seem that only the less available meanings are susceptible to a reduction in availability.

Method

Participants

One-hundred-and-seventeen native British English speakers participated in the current experiment (68 females; mean age = 28.83, range = 18-43). All participants met the demographic requirements and were recruited and paid as outlined in Experiment 5.

Design

The design was the same as in Experiment 5, except that subordinate priming was replaced with dominant priming (two levels: dominant unprimed and dominant primed). Since this was set up in exactly the same way as the subordinate priming factor, see Experiment 5 Design for details.

Materials

All prime and test words were identical to Experiment 5. However, since the subordinate prime task was replaced with a dominant prime task for the present experiment, new sentences were created for the 60 experimental ambiguous and 28 filler ambiguous words, where each sentence disambiguated the ambiguous word towards its dominant meaning with prior context. Each sentence was coupled with a related or an unrelated probe word (50% sentences with related probes, 50% with unrelated for experimental trials; unrelated probes were not related to any meaning of the ambiguous word). All unambiguous prime sentences were unaltered. All dominant sentences were newly recorded by a female native British English speaker with a Southern English accent (HNB).

See Experiment 5 Materials for details on the semantic relatedness picture test, as it is the same in this experiment.

Procedure

See Experiment 5 Procedure for details, since it was the same overall with the exception of the filler task, which was omitted for the present experiment (and the subordinate prime task simply replaced by the dominant prime task).

Task Checks

Whilst 117 participants completed the experiment, only the data from 104 participants were analysed: thirteen participants were excluded for meeting one or more of the exclusion criteria outlined in Experiment 5. None of the items were excluded as they all exceeded the 70% accuracy requirement, leaving the total 60 items for analyses.

Results

Reaction Time Analyses

As the subject means in Figure 15 indicate, RTs to the dominant picture at test appear to be faster after dominant priming, compared to the unprimed baseline, but there seems to be no pattern to suggest that dominant priming influences RTs to the subordinate picture. Overall, RTs seem to be faster for dominant pictures than for subordinate pictures.

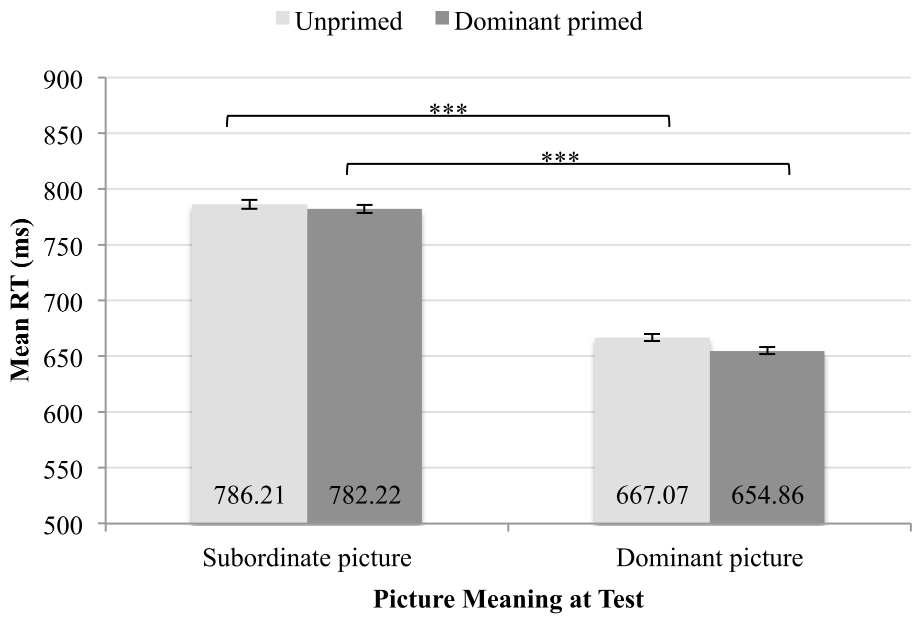

Figure 15. Mean by-subjects reaction times for Experiment 6. Responses for both the subordinate and dominant meaning picture, following either no priming or one dominant prime repetition. Significance level indicated with asterisks (***<.001). Error bars are adjusted for the within-subjects design.

The RTs were trimmed, transformed, modelled and analysed as in the RT analyses in Experiment 5. The fixed effects for the present full model were dominant priming, picture meaning and the interaction term (all coded as in Experiment 5). The full model failed to converge across all tests of main effects and interactions hence the random effects structure was simplified by removing one random effect term at a time until convergence was reached for all nested models. The second fullest model converged (correlations between subject and items slopes and intercepts removed) for all nested models.

The model comparison approach revealed a significant main effect of dominant priming (X^2 (1) = 4.075, p = .044), suggesting that across picture meaning conditions, participants were faster to respond to pictures if they had encountered the dominant meaning of that word in the prime phase. There was also a significant main effect of picture meaning (X^2 (1) = 37.506, p < .001), whereby participants were faster at responding to dominant pictures than subordinate pictures, regardless of dominant priming condition. However, there was no significant interaction between dominant priming and picture meaning (X^2 (1) = 1.139, p = .286), indicating that the effect of dominant priming did not differ depending on whether participants were responding to a dominant or a subordinate picture at test.

A set of four simple effects analyses was conducted, with Bonferroni-adjusted p values, as with Experiment 5. First, the effect of dominant priming for the subordinate picture at test was not significant (X^2 (1) = 0.165, p = .999), showing that dominant priming did not alter the speed of responses to the subordinate picture, compared to trials without dominant priming. This suggests that dominant priming did not interfere with availability of the alternative, subordinate, meaning. Second, the effect of dominant priming for the dominant picture at test was not significant either (X^2 (1) = 4.542, p = .132), where dominant priming did not speed responses to the dominant picture, suggesting that the dominant priming manipulation was unsuccessful for RTs (although note that this *was significant* before Bonferroni correction, p = .033). There were significant simple effects of picture meaning for both the dominant unprimed (X^2 (1) = 32.984, p < .001) and dominant primed conditions (X^2 (1) = 37.975, p < .001), again indicating faster RTs to the dominant picture than to the subordinate picture both without and with dominant priming, respectively.

Error Rate Analyses

The subject means in Figure 16 indicate that dominant priming slightly increases percentage error for both the subordinate and dominant picture, although the percentage error for the dominant picture is overall considerably lower than for the subordinate picture.

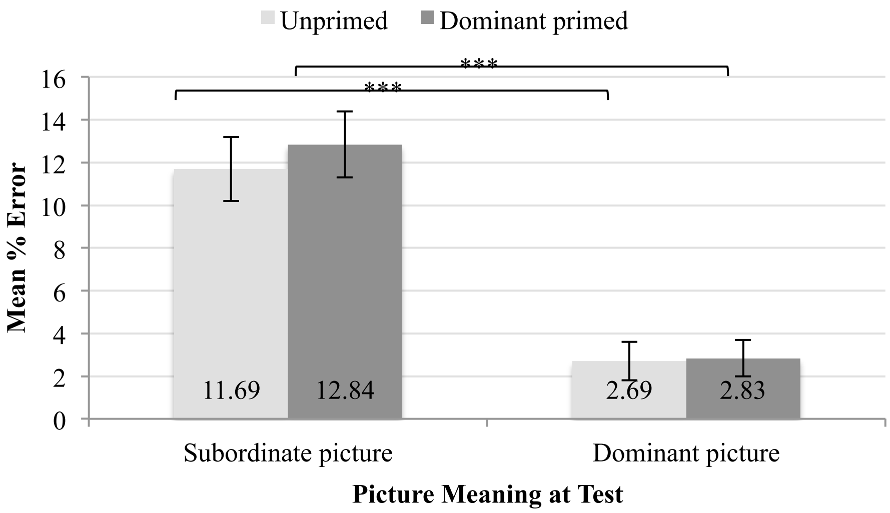

Figure 16. Mean by-subjects percentage error for Experiment 6. Responses for both the subordinate and dominant meaning picture, following either no priming or one dominant prime repetition. Significance level indicated with asterisks (***<.001). Error bars are adjusted for the within-subjects design.

Note that since picture meaning simple effects (i.e. dominant vs. subordinate picture after no priming; dominant vs. subordinate picture after priming) are significant at $p < .001$ in all the analyses reported in this chapter, significance bars for these effects will not be shown on figures from here onwards so as not to complicate the figures.

The errors were trimmed, modelled and analysed using the same method as the error analyses in Experiment 5. The fixed effects for the present full model were dominant priming, picture meaning and the interaction term (all coded as in Experiment 5). The full model failed to converge across all tests of main effects and interactions hence the random effects structure was simplified by removing one random effect term at a time until convergence was reached for all nested models. The intercepts-only model (simplest random effects structure) converged for all models.

The model comparisons revealed no significant main effect of dominant priming (X^2 (1) = 0.566, p = .452), suggesting that, across picture meaning conditions, participant errors to pictures were unaffected by whether or not they had encountered the dominant meaning of that word in the prime phase. There was a significant main effect dominant picture meaning (X^2 (1) = 239.590, p < .001), showing that, across dominant priming conditions, there were fewer errors when responding to dominant pictures, compared to subordinate pictures. However, there was no significant interaction between dominant priming and picture meaning conditions (X^2 (1) = 0.048, p = .826), indicating that the effect of priming was the same regardless of whether the participant responded to the dominant or subordinate picture presented at test. Thus, dominant priming did not significantly affect overall error rates.

The four simple effects analyses confirmed the pattern shown in the main effects analyses. Bonferroni-adjusted p values are reported. First, the effect of dominant priming for the subordinate picture at test was not significant for the errors (X^2 (1) = 1.202, p = .999), suggesting that one dominant repetition at prime was *not* sufficient to increase error rates to the subordinate picture at test. This indicates that dominant priming did not interfere with availability of the alternative, subordinate, meaning. Second, the effect of dominant priming for the dominant picture at test was not significant (X^2 (1) = 0.082, p = .999), showing that one dominant repetition did not reduce error rates to the dominant picture. Again, there were significant simple effects of picture meaning for both the dominant unprimed (X^2 (1) = 109.570, p < .001) and dominant primed conditions (X^2 (1) = 127.640, p < .001), showing fewer errors to the dominant picture than to the subordinate picture both without and with dominant priming, respectively.

Awareness Analyses

As with Experiment 5, there were two awareness measures: awareness of experimental aim and awareness estimate, both of which require analysis with linear mixed effects modelling for RTs and logistic mixed effects modelling for error rates to investigate their effect on priming. One experimenter (HNB) coded the responses to the awareness of experimental aim question (see Experiment 5 for coding scheme). One-hundred participants were unaware of the aim (for dominant picture test condition only, primed and unprimed levels combined[21]: RT *mean* = 664.43ms, *SD* = 77.20ms, percentage error *mean* = 2.72%, *SD* = 3.43%) and only 4 participants were fully/partially aware of the aim (for dominant picture test condition only: RT *mean* = 576.96ms, *SD* = 25.29ms, percentage error *mean* = 4.17%, *SD* = 5.00%), hence there was an insufficient number of participants (only 3.8%) in the "aware" category to run an analysis to examine whether priming interacts with awareness of aim. The awareness estimate data were continuous, indicating participants' estimates of the percentage of ambiguous words in the word association test that had been presented earlier in the experiment as a less explicit measure of awareness, (word estimate *median* = 25, *range* = 0-131, skewed distribution). These estimate data were rescaled (divided by 100) and centred.

The awareness estimate factor was included as a fixed effect in a mixed effects model along with the fixed factor of dominant priming, which indicated whether an item was unprimed or dominant primed. The random effects structure was constructed with subjects and items intercepts and slopes for priming. The model did not require slopes for the awareness estimate factor, as it is a single value for a participant across all items. However, only the intercepts-only model converged for both RTs and error. The interaction between the awareness estimate and dominant priming factor was the crucial test, since a significant interaction would indicate that priming varied as a function of awareness. As before, a model comparison approach

[21] Only the prime-test congruent meaning condition (dominant picture) was included, i.e. only the picture meaning that was consistent with the primed meaning. This was because we were interested only in whether or not participants were aware of the repetition of a meaning from the prime phase in the test phase, therefore it is irrelevant for the subordinate picture in this experiment since this subordinate meaning was not primed.

was used to determine the significance of this interaction, where a model with priming, awareness and their interaction was compared to a model with both fixed effects without the interaction term. This showed that the interaction was not significant for RTs (X^2 (1) = 2.445, p = .118), or for errors (X^2 (1) = 1.988, p = .159), indicating that participants' awareness of how many test words were repeated from the prime phase did not influence dominant meaning priming effects.

Discussion

Importantly, the findings replicated those of Experiment 4 in showing that unprimed RTs to the dominant meaning are significantly faster and more accurate than responses to the subordinate meaning. This once again demonstrates the reliability of the semantic relatedness test. Moreover, the mix of trial types at test (unprimed/primed, ambiguous/unambiguous, related/unrelated) reduced awareness of the priming manipulation to the degree that only 3.8% of participants were aware of it. Additionally, that participants' awareness of how many test words were repeated from the prime phase did not significantly influence dominant meaning priming effects is reassuring. Based on these data, it seems that the semantic relatedness test minimises participant awareness of the prime-test manipulation compared to the word association test (whilst more participants were aware of the prime-test link with word association, awareness still did not affect priming).

The main aim of the present experiment was to investigate whether dominant priming, by increasing availability of the dominant meaning, would decrease availability of the unprimed, subordinate meaning. The results show that RTs and errors to the subordinate pictures at test were not significantly greater following an encounter with the dominant meaning, compared to when the dominant meaning had not been encountered previously. However, despite the significant main effect of priming, the benefit of the dominant priming manipulation failed to reach significance after correction across the simple effects analyses. This means, at a corrected level at least, that the dominant priming manipulation was not successful in significantly reducing RTs and errors to the dominant picture. Hence, any effect whereby dominant priming might increase RTs and errors to the subordinate meaning, and therefore reduce its availability, could not be tested sufficiently within the present experiment, since dominant priming did not significantly increase availability of the dominant meaning.

It is likely that the dominant priming manipulation lost its significance at a corrected level due to the effect of dominant priming being weaker than that of subordinate priming. In Experiment 1, participants encountered the subordinate meaning and then, after a filler task they encountered the dominant meaning before completing the word association test. Whilst the magnitude of priming was comparable for both meanings (4% absolute increase in the proportion of subordinate responses after one repetition of the subordinate meaning, 8% absolute decrease in the proportion of subordinate responses after one repetition of the dominant meaning), the dominant meaning had been encountered much more recently than the subordinate meaning (on average 9 minutes before test, compared to the 30 minutes before test for the subordinate meaning). This suggested that subordinate priming was stronger than dominant priming, since subordinate priming was able to endure a longer delay and maintain the same magnitude of priming as the dominant meaning. As explained by Rodd et al. (2013), it is not surprising that the subordinate meaning is more susceptible to priming; due to its lower likelihood of being available, it can easily accommodate a boost in availability. In contrast, the dominant meaning is already highly available, providing less room for a boost in availability, and is therefore a weaker priming manipulation. If this is the case, the 7-minute prime-test delay in the present experiment might be too long for any dominant availability boost to be sustained at a level that can reach corrected significance.

There are two implications of the unsuccessful dominant priming manipulation. First, it might be that priming with the dominant meaning does not produce a significant change in performance if dominant availability is already so high that priming cannot boost it further. Second, it might be that testing with the dominant meaning does not produce a significant change in performance because it is the most familiar meaning and therefore might not be sensitive to effects of priming at test (i.e. there is a distinction between not being able to prime the dominant meaning and not being able to show dominant priming with the dominant meaning). If the dominant picture is presented to participants at test, after dominant priming, then it is the most recently encountered, available meaning but also the most generally encountered, available meaning. For the response to this picture to be speeded, compared to the unprimed baseline, it would need to be extremely sensitive to show an effect of recent experience with the dominant meaning in addition to all the general experience with that dominant meaning. Hence, it might be that, at a corrected level, the significant dominant priming effect disappears because it is too small to be observed with the current statistical power.

Taken together, Experiments 5 and 6 show that there is no effect of priming on the unprimed meaning from one repetition, regardless of whether it is the subordinate or dominant meaning that is primed. However, before concluding (on null findings) that such a priming effect does not exist, we will try to adapt the priming manipulation to boost the priming effect to give us the best possible chance of observing any interference effects on the unprimed meaning. One way to boost the priming effect is to increase the number of prime repetitions.

Previous research has shown that a stronger priming manipulation is one in which word-meanings are repeated multiple times. For instance, Experiments 2 and 3 of this thesis (Chapter 2) showed that three spaced subordinate repetitions are consistently and significantly superior to one repetition in boosting word-meaning priming effects. Additionally, Colbert-Getz and Cook (2013) found that four subordinate meaning repetitions increased priming compared to one (although note that this was four repetitions of the context rather than of the ambiguous word itself, which is only effective in methods such as theirs with a prime-test delay of a few minutes; Rodd et al., 2013). Moreover, two repetitions of the same word-meaning in a sentence improves later recall compared to a repetition of each of two different word-meanings (Thios, 1972), again showing a benefit of repetitions for priming effects. As a result, Experiment 7 will include a stronger subordinate priming manipulation involving *three* subordinate meaning repetitions.

Experiment 7 – three spaced subordinate repetitions

The design of Experiment 5 was revisited for the present experiment, since it involved a one repetition subordinate priming manipulation that could easily be extended to multiple repetitions (since it seems that subordinate priming is a stronger effect than dominant priming), and a filler task creating a longer prime-test delay than in Experiment 6. The design was altered in two ways for the present experiment. The first difference was that the subordinate prime condition was replaced with a *repeated* (spaced) subordinate prime condition for the present experiment, which involved two additional subordinate prime sentences per ambiguous word to provide the stronger priming manipulation[22] (a total of three subordinate sentence repetitions, as opposed to the single subordinate sentences presented in Experiment 5). These three sentences were spaced across three prime blocks, with one of these three sentences per block. This spacing is the same as the spaced conditions in Experiments 2 and 3. The second difference was that unambiguous filler sentences were not included in the prime phase for the present experiment. This was to reduce the length of the prime phase (to reduce the likelihood of participant fatigue during the task), since the two additional subordinate prime sentences per ambiguous word substantially increased its length. See Figure 17 for a brief overview of the procedure.

[22] There is no single subordinate repetition condition, only unprimed versus three subordinate repetitions.

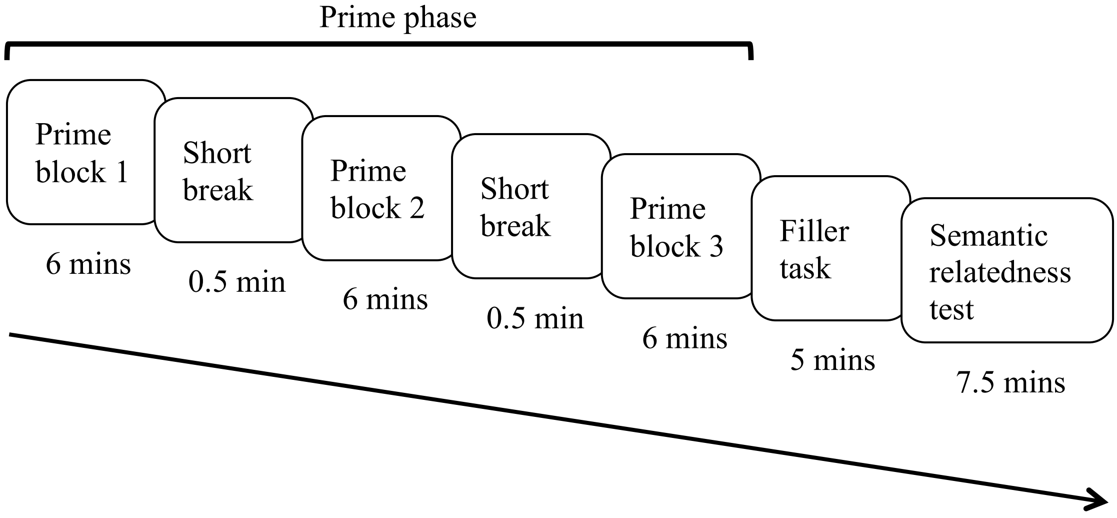

Figure 17. Experiment 7 task order, with the mean duration of each task. The mean prime-test delay is 18 minutes (rounded down from 18.25 for ease of reference).

Method

Participants

One-hundred-and-sixteen native British English speakers participated in the current experiment (68 females; mean age = 29.13, range = 18-44). All participants met the demographic requirements and were recruited and paid as outlined in Experiment 5.

Design

The design was the same as in Experiment 5, except that the single repetition subordinate priming factor was replaced with the spaced subordinate priming factor in the present experiment (two levels: subordinate unprimed and spaced subordinate primed). Since this was set up in exactly the same way as the subordinate priming factor, see Experiment 5 Design for details. Also see Experiment 5 Design for details on the picture meaning independent variable, since this was identical in the present experiment.

Whilst the 28 ambiguous sentences were still included at prime as fillers, all unambiguous filler sentences from the Experiment 5 prime phase were removed for the present experiment. This was because the unambiguous filler words only had one sentence per item, whereas the ambiguous filler words and experimental words each had three sentences, which could make the difference between primed ambiguous and unambiguous words obvious to participants. However, the remaining 28 filler ambiguous sentences still served the purpose of providing trials that, at test, could be paired with unrelated pictures and therefore trials that could be removed from analyses without lowering the number of experimental items to analyse.

The inclusion of unambiguous test trials (both related and unrelated) was to reduce the salience of the prime-test link. This design meant that fillers in the test phase differed across the three dimensions of: 1) priming, 2) ambiguity and 3) word-picture relatedness at test. As a result, knowing whether or not a word was primed/unprimed or ambiguous/unambiguous could not help a participant make a faster or correct response to either the dominant or subordinate picture at test. Finally, the mixture of stimulus types at test (primed/unprimed, ambiguous/unambiguous, related/unrelated, as well as dominant picture/subordinate picture) would make it highly unlikely that for participants would be aware of the inequality between the types at test. Moreover, despite the fact that all the primed sentences were ambiguous, the participants would not necessarily have been aware that the subordinate primed sentences were all ambiguous, since they were disambiguated with prior context (for instance, it is not necessarily obvious that the words were ambiguous and the subordinate meaning of it is always used: ‘it was the hottest day on *record*’). See Table 6 for details of the stimulus types.

Table 6. Details of ambiguous word stimuli at prime and test phases in Experiment 7.

Prime	Test		
Word Qualities	**Stimulus Type**	**Word-Picture Relatedness**	**No. of Stimuli**
Primed, ambiguous	Experimental	Related	30
Primed, ambiguous	Filler	Unrelated	28
Primed, unambiguous	*Filler*	*Related*	*0*
Primed, unambiguous	*Filler*	*Unrelated*	*0*
Unprimed, ambiguous	Experimental	Related	30
Unprimed, ambiguous	Filler	Unrelated	0
Unprimed, unambiguous	*Filler*	*Related*	*12*
Unprimed, unambiguous	*Filler*	*Unrelated*	*26*

Note: Unprimed words at prime are in grey to emphasise that they were not encountered until the test phase. Italics show differences from Experiment 5.

Materials

All word items were the same as in Experiments 5 and 6: sixty experimental ambiguous and 28 filler ambiguous words, but each ambiguous word had three different sentences (and probe words) per ambiguous word, each disambiguating towards the same subordinate meaning but with varying context. See Table 7 for an example of three subordinate prime sentences.

Table 7. An example of the three sentences and probe words for the ambiguous word 'glasses' in Experiment 7.

Number	Sentence (ambiguous word in italics)	Probe
1.	The cupboard stored the mugs and *glasses*	Prefer (unrelated)
2.	She poured the champagne into the *glasses*	Fizz (related)
3.	The waiter set out the plates, cutlery and *glasses*	Table (related)

The Towers of Hanoi task was again used as the filler between prime and test. See Experiment 5 Materials for details.

For the semantic relatedness picture test there was a total of 126 items as in Experiment 5: 60 ambiguous experimental words with related pictures, 28 ambiguous filler words with unrelated pictures, and 38 unambiguous filler words – 12 related and 26 unrelated (these 38 fillers were all unprimed in the present experiment, unlike in Experiment 5). In total at test, 44% of trials were primed, 64% of trials were ambiguous and 55% of trials were related.

Procedure

The procedure was very similar to that of Experiment 5; there were only two differences. The first difference was that the present experiment required that the subordinate prime task be split into three blocks (with a 30 second break between blocks), in order to space the three sentence repetitions, as in Experiments 2 and 3. The break between blocks was automatic and instructed participants to wait for 30 seconds, after which the next block would start automatically. Since there were more sentences in the present experiment (making the prime task longer) than in previous, the block design also provided the opportunity for participants to rest between each block to help maintain concentration levels. The second difference was that the Towers of Hanoi filler task time was reduced from 6 minutes to 5 minutes for the present experiment. This was in an effort to reduce the overall prime-test delay to be more similar to Experiment 5, since the present prime task was longer (due to the three sentences per ambiguous word).

Task Checks

Whilst 116 participants completed the experiment, only the data from 99 participants were analysed: eleven participants were excluded for meeting one or more of the exclusion criteria as outlined in Experiment 5 (where the maximum time allowed for the present experiment was 35 minutes). A further 6 participants were removed due to technical issues that disrupted one or more of the tasks during the experiment. In addition, all items were checked for prime and test accuracy and were excluded if their accuracy, averaged across participants, was less than 70%. This resulted in the exclusion of one item (‘gear) for all participants, leaving a total of 59 items in the analyses.

Results

Reaction Time Analyses

The subject means in Figure 18 suggest that RTs to the subordinate picture are faster after repeated, spaced subordinate priming, compared to the unprimed baseline. On the contrary, it seems that RTs to the dominant picture are *slower* after the spaced subordinate priming. Overall, RTs seem to be faster for dominant pictures than for subordinate pictures.

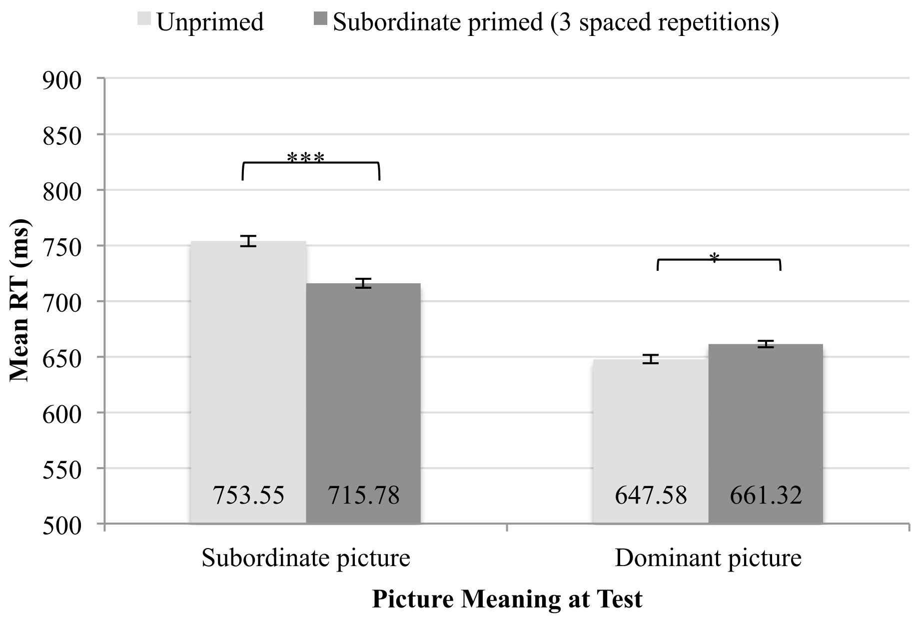

Figure 18. Mean by-subjects reaction times for Experiment 7. Responses for both the subordinate and dominant meaning picture, following either no priming or three spaced subordinate prime repetitions. Significance level indicated with asterisks (* <.05, ***<.001). Error bars are adjusted for the within-subjects design.

Note that, in Figure 18, only the simple effects are shown for the theoretically important contrasts. The main effect of responses to the dominant picture being significantly faster than to the subordinate picture regardless of priming condition is not indicated with asterisks because, as explained in Experiment 6 Results, this main effect is significant for all analyses in the present chapter. It is therefore omitted here and in all subsequent figures so as not to complicate them.

The RTs were trimmed, modelled and analysed using the same method as the RT analyses in Experiments 5 and 6 (although the log transformation was more suitable for the present data). The fixed effects for the present full model were therefore spaced subordinate priming (three spaced repetitions), picture meaning and the interaction term (all coded as in Experiment 5). The full model failed to converge across all tests of main effects and interactions but the second fullest model converged (correlations between subject and items slopes and intercepts removed) for all models.

The model comparison approach indicated a significant main effect of spaced subordinate priming, (X^2 (1) = 10.545, p = .001). This suggests that across picture meaning conditions, participants were faster to respond to pictures after they had encountered three subordinate repetitions of that word in the prime phase. Comparisons also showed a significant main effect of picture meaning at test (X^2 (1) = 17.747, p < .001), showing once more that across subordinate priming conditions participants were generally faster at responding to dominant than subordinate pictures. There was also a significant interaction between spaced subordinate priming and picture meaning (X^2 (1) = 25.979, p < .001), indicating that the speed benefit from the spaced subordinate priming differed depending on whether they encountered the dominant or subordinate picture at test.

A set of four simple effects analyses was conducted, with Bonferroni-adjusted p values, as with Experiments 5 and 6. First, the effect of spaced subordinate priming for the subordinate picture at test was significant (X^2 (1) = 26.795, p < .001), showing that three spaced subordinate priming significantly speeded responses to the subordinate picture, compared to trials without subordinate priming. Second, the effect of spaced subordinate priming for the dominant picture at test was also significant (X^2 (1) = 6.319, p = .048), where subordinate priming *slowed* responses to the dominant picture (even at this Bonferroni-corrected p value; uncorrected value p = .012). This suggests that repeated spaced subordinate priming interfered with the availability of the alternative, dominant, meaning. There were also significant simple effects of picture meaning for both the subordinate unprimed (X^2 (1) = 26.987, p < .001) and spaced subordinate primed conditions (X^2 (1) = 8.430, p = .015), again indicating faster RTs to the dominant picture than to the subordinate picture both without and even with three spaced subordinate prime repetitions, respectively.

Error Rate Analyses

As the means in Figure 19 indicate, it seems that repeated, spaced subordinate priming might reduce percentage error for the subordinate picture, but have no effect on the dominant picture. The percentage error for the dominant picture is overall lower than for the subordinate picture.

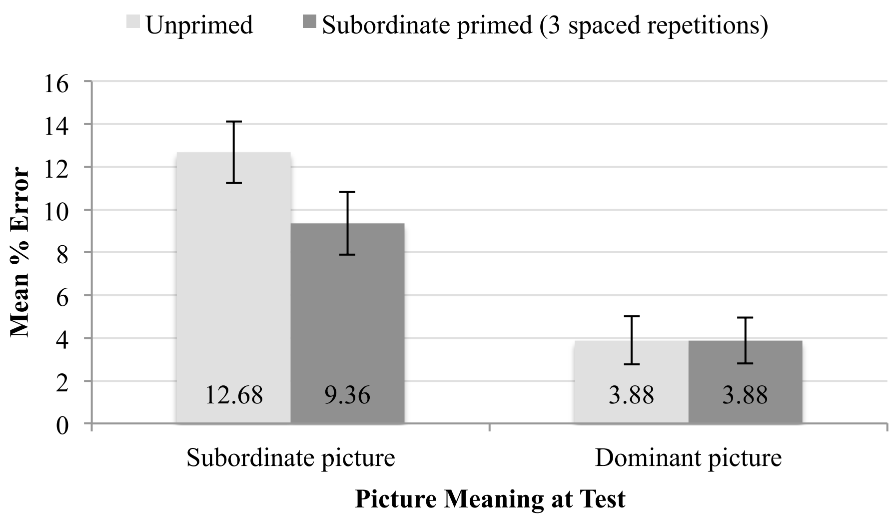

Figure 19. Mean by-subjects percentage error for Experiment 7. Responses for both the subordinate and dominant meaning picture, following either no priming or three spaced subordinate prime repetitions. Significant main effect of picture meaning not represented with asterisks for reason explained earlier. Error bars are adjusted for the within-subjects design.

The errors were trimmed, modelled and analysed using the same method as the error analyses in Experiments 5 and 6. The fixed effects for the present full model were spaced subordinate priming, picture meaning and the interaction term (all coded as in Experiment 5). The full random effects structure model converged for all models.

The model comparisons showed no significant main effect of spaced subordinate priming (X^2 (1) = 0.783, p = .376), suggesting that, across picture meaning conditions, errors were not influenced by whether or not participants had encountered the subordinate meaning of that word in the prime phase. There was a significant main effect of picture meaning (X^2 (1) = 16.127, p < .001), showing that, across subordinate priming conditions, there were fewer errors when responding to dominant pictures, compared to subordinate pictures. However, there was no significant interaction between spaced subordinate priming and picture meaning conditions (X^2 (1) = 0.017, p = .898), indicating that the effect of priming was the same regardless of whether the participant responded to the dominant or subordinate picture presented at test. Spaced subordinate priming therefore did not significantly affect error rates overall.

As with Experiment 5, the four simple effects analyses confirmed the pattern from the main effects. Bonferroni-adjusted p values are reported. First, the numerical effect of spaced subordinate priming reducing errors to the subordinate picture was not significant (X^2 (1) = 1.522, p = .869), suggesting that even three subordinate repetitions at prime were insufficient to significantly reduce error rates to the subordinate picture at test. Second, the effect of subordinate priming for the dominant picture at test was not significant (X^2 (1) = 0.283, p = .999), showing that three subordinate repetitions did not increase error rates to the dominant picture. This suggests that, whilst three spaced subordinate repetitions can *slow* responses to the dominant picture, they cannot significantly increase the number of errors made when participants respond to the dominant picture. Hence, spaced subordinate priming did not interfere with the availability of the alternative, dominant, meaning in terms of error rates. Third, there was a significant effect of picture meaning for the subordinate unprimed condition (X^2 (1) = 11.897, p = .002), indicating fewer errors when responding to the dominant picture than to the subordinate picture for unprimed trials. Fourth, there was a significant effect of picture meaning for the spaced subordinate primed condition (X^2 (1) = 13.685, $p < .001$), indicating fewer errors to the dominant picture than to the subordinate picture even after three subordinate repetitions in the prime phase.

Awareness Analyses

There were two awareness measures: awareness of experimental aim (coded as in Experiments 5 and 6) and awareness estimate (rescaled and analysed as in Experiment 6). For the awareness of experimental aim question, 96 participants were unaware of the aim (for subordinate picture test condition only[23]: RT *mean* = 737.65ms, *SD* = 93.35ms, percentage error *mean* = 10.90%, *SD* = 8.07%) and only 3 participants were fully/partially aware of the aim (for subordinate picture test condition only: RT *mean* = 636.67ms, *SD* = 99.39, percentage error *mean* = 15.15%, *SD* = 7.28%). Hence, as with Experiments 5 and 6, there was an insufficient number of participants (only 3%) in the "aware" category to run an analysis to examine whether priming interacts with awareness of aim.

The awareness estimate data were continuous, indicating participants' estimates of the percentage of ambiguous words in the word association test that had been presented earlier in the experiment as a less explicit measure of awareness, (word estimate *median* = 40, *range* = 0-131, skewed distribution). These estimate data were rescaled (divided by 100) and centred. As with Experiment 6, the model with priming slopes did not converge, leaving an intercepts-only model. A model comparison showed that the interaction between spaced subordinate priming and awareness estimate was not significant for RTs or errors (X^2 (1) = 1.244, $p = .265$; (X^2 (1) = .002, $p = .965$, respectively), again indicating that participants' awareness of how many test words were repeated from the prime phase did not influence priming effects.

[23] As with Experiment 5, only the prime-test congruent meaning condition (subordinate picture) was included.

Discussion

The aim of the present experiment was to investigate whether strong subordinate priming, by increasing availability of the subordinate meaning, would decrease availability of the unprimed, dominant meaning. For the stronger priming manipulation, participants encountered the subordinate meaning three times, spaced throughout the prime phase in different sentences (or not at all; unprimed baseline).

First, the present experiment replicated results from Experiment 6 in that participants' awareness of how many test words were repeated from the prime phase did not influence subordinate meaning priming effects. Also like in Experiment 6, the percentage of participants aware of the prime-test manipulation was too small for analyses with this measure (3%). It seems that the semantic relatedness test minimises participant awareness of the prime-test manipulation compared to the word association test.

Second, the results demonstrated that the semantic relatedness picture test successfully detected subordinate priming, supporting the findings from Experiment 5. The availability of the subordinate meaning was significantly increased following three spaced encounters with the subordinate meaning earlier in the experiment. Participants were significantly faster at responding to subordinate pictures following subordinate priming, although they were not significantly more likely to make a correct response, despite a numerical effect (a 3% absolute increase in accuracy after three spaced repetitions, compared to a 1% absolute increase in accuracy after one repetition in Experiment 5). Third, and crucially, participants were significantly *slower* at responding to the dominant meaning picture following three earlier spaced encounters with the alternative subordinate meaning compared to responses to the dominant picture when priming had not occurred. This suggests that the availability of the dominant meaning significantly decreased as a result of the increased availability of the subordinate meaning.

The present results rule out the possibility discussed in Experiment 5 that an increase in the availability of one meaning *never* affects the availability of the alternative meaning. Combining the present results with those of Experiment 5, it seems that whilst a single subordinate meaning encounter is only sufficient to boost the availability of that meaning after a delay, multiple spaced subordinate encounters are sufficient to reciprocally reduce the availability of the alternative, dominant meaning after a delay. This suggests that the connection strengths concerning multiple meanings of an ambiguous *can* change depending on one another. It might be that only stronger priming, such as with three spaced repetitions, reduces unprimed meaning availability (i.e. there is no reduction after one repetition), or it might be that both one and three spaced repetitions reduce unprimed meaning availability but to different extents, where the reduction is only large enough to be detected after the stronger spaced repetition priming. Alternatively, it might be that the time between each spaced repetition provides an opportunity for the primed meaning to be consolidated and, as a result, interfere with the unprimed meaning.

These findings extend those of Chen and Boland (2008) to a longer prime-test delay, which found that on a trial-by-trial basis (i.e. at a much shorter delay than the present experiments) that a single subordinate sentence could interfere with and reduce access to the unprimed, dominant meaning. The present findings are inconsistent with the class of model in which meaning availabilities update independently of one another. They are therefore consistent with the class of model that assumes that the availabilities of alternative word-meanings are linked. When subordinate context increases the availability of the subordinate meaning, the competing dominant meaning can decrease in availability, although the circumstances in which unprimed meaning availability is and is not decreased are currently unclear.

Interestingly, responses to the dominant meaning are consistently faster and more accurate than to the subordinate meaning, even after three subordinate prime repetitions. This shows that despite both increasing subordinate availability and decreasing dominant availability, the dominant meaning remains the most available, highlighting its strong influence on disambiguation. This is not surprising given that lexical-semantic representations seem to be developed over a lifetime and reflect both long-term experiences over months or years, as well as more recent experience over minutes or hours (Rodd et al., 2016). This finding of the dominant meaning still being the most available meaning after strong subordinate priming is consistent with the reordered access model (Duffy et al., 1988), which would predict that access to meanings is exhaustive even in the presence of strong context.

In summary, the present experiment revealed that three spaced subordinate repetitions increased the availability of the subordinate meaning and decreased the availability of the unprimed, dominant meaning. However, Experiment 5 suggested that dominant availability was not decreased following one subordinate repetition. Relating this finding back to the word association repetition experiment (Experiment 3), which found that massed repetitions were no better than one repetition but spaced repetitions were, it would be interesting to reintroduce the “massed versus spaced” comparison (from Chapter 2) with the semantic relatedness picture test to investigate whether the superiority of spaced repetitions over massed repetitions replicates. Since the semantic relatedness test gives additional data per item (i.e. RT and error rate rather than the single word response from word association) and can test the effects of priming on the alternative meanings of words independently, it might give an insight into the cause of the massed versus spaced difference. As a result, Experiment 8 will extend the present experiment with the addition of a massed repetition condition.

Experiment 8 – three massed & three spaced subordinate repetitions

There were two aims of the present experiment: (1) to investigate whether the effect from Experiment 7 of three spaced subordinate repetitions reducing dominant meaning availability replicated, and (2) to investigate whether three spaced subordinate repetitions boost priming compared to three massed subordinate repetitions in a semantic relatedness task (i.e. whether the findings of Experiment 3 are consistent across tests). Consistent with the priming patterns found in Experiment 3, it was predicted that massed repetitions would lead to significantly faster RTs to the subordinate picture compared to the unprimed baseline, but significantly slower RTs to the subordinate picture compared to spaced repetitions,. No predictions were made for the effect of massed subordinate repetitions on the dominant picture, since there was no clear steer from previous findings.

The present experiment was therefore an extension of the design of Experiment 7 and differed in only one way: the addition of a massed repetition condition. This resulted in three prime levels (unprimed, three massed, three spaced). A one repetition condition was not included since a limit on the number of items available meant that a fourth condition would substantially reduce power.

Method

Participants

To compensate for the addition of the third experimental condition of three massed repetitions compared to Experiment 7 (which would mean that participants give fewer data points per condition, due to a limit on the number of stimuli in the experiment), the target for the number of participants recruited for Experiment 8 was approximately 50% more than for Experiment 7. Therefore, one-hundred-and-eighty native British English speakers participated in the current experiment (125 females; mean age = 31.96, range = 18-44). All participants met the demographic requirements and were recruited and paid as outlined in Experiment 5.

Design

This experiment involved a within-subjects design with two independent variables (repeated subordinate priming – massed versus spaced – and picture meaning). The dependent variables were the reaction times and error rates of responses to the pictures at test.

The first independent variable was repeated subordinate priming (three levels: subordinate unprimed, subordinate primed with three massed repetitions, subordinate primed with three spaced repetitions). Participants encountered two thirds of the ambiguous words in the prime phase, (one third massed, one third spaced). The remaining third of the ambiguous words were not encountered in the prime phase but were later introduced in the test phase to become an unprimed baseline. At prime, three versions were created so that the 20 massed primed words for a third of the participants were the 20 spaced primed words for another third and the 20 unprimed words for the remaining third of the participants, and vice versa.

The 20 spaced experimental items per version were naturally distributed over the prime phase, with one sentence per prime block (sentence 1 in block, sentence 2 in block 2, sentence 3 in block 3). This summed to 20 *sentences* per block that were in the spaced condition. The 20 massed experimental items per version were divided into 3 groups, allocating 7 items to block 1, 7 items to block 2 and 6 items to block 3. When a massed item was presented, sentence 1 was immediately succeeded by sentence 2 and sentence 2 immediately succeeded by sentence 3, as in Experiments 2 and 3. This summed to 21 massed sentences in each of blocks 1 and 2, and 18 massed sentences in block 3. The filler primed ambiguous items were distributed in the same way – spaced items spread over blocks such that a spaced item appeared in each block but with a different sentence; massed items divided such that one third of the items, and therefore the three different sentences for each of the items in that third, appeared in any given block.

The second independent variable was semantic relatedness picture meaning (levels: dominant and subordinate) where participants encountered half of the experimental ambiguous words at test paired with the dominant meaning picture and the other half paired with the subordinate meaning. The test therefore required each of the three prime versions to be split further into two versions, where the items were rotated so that, at test, each word appeared as both an unprimed, massed and spaced trial and paired with a dominant picture and a subordinate picture at a test trial. This meant that all participants contributed to each of the six conditions but for different ambiguous words. Across different versions a given ambiguous word therefore appeared in each of the six conditions but for different participants. All filler trials were identical across versions at prime and at test.

As with Experiments 5, 6 and 7, all experimental items were paired with related pictures at test. In addition, 28 ambiguous sentences were included as fillers at prime (10 unprimed, 9 massed, 9 spaced). These fillers served the purpose of providing trials that, at test, could be paired with unrelated pictures and therefore trials that could be removed from analyses without lowering the number of experimental items to analyse. The 38 unprimed unambiguous trials at test (both related and unrelated) were the same as in Experiment 7. See Table 8 for details on stimulus types.

Table 8. Details of ambiguous word stimuli at prime and test phases in Experiment 8.

Prime		Test		
Word Qualities	**Repetition Spacing**	**Stimulus Type**	**Word-Picture Relatedness**	**No. of Stimuli**
Primed, ambiguous	*3 Massed*	*Experimental*	*Related*	*20*
	3 Spaced	*Experimental*	*Related*	*20*
Primed, ambiguous	*3 Massed*	*Filler*	*Unrelated*	*9*
	3 Spaced	*Filler*	*Unrelated*	*9*
Primed, unambiguous		Filler	Related	0
Primed, unambiguous		Filler	Unrelated	0
Unprimed, ambiguous		*Experimental*	*Related*	*20*
Unprimed, ambiguous		*Filler*	*Unrelated*	*10*
Unprimed, unambiguous		Filler	Related	12
Unprimed, unambiguous		Filler	Unrelated	26

Note: Unprimed words are in grey to emphasise that they were not encountered until the test phase. Italics show differences from Experiment 7.

Materials

See Experiment 7 for details, since the materials in the present experiment were identical.

Procedure

See Experiment 7 for details, since the procedure in the present experiment was identical.

Task Checks

Whilst 180 participants completed the experiment, only the data from 168 participants were analysed: twelve participants were excluded for meeting one or more of the exclusion criteria outlined in Experiment 5 (where 35 minutes was the maximum time allowed for the present experiment). None of the items were excluded as they all exceeded the 70% accuracy requirement, leaving the total 60 items for analyses.

Results

Whilst the present experiment is simply an extended version of Experiment 7, due to the addition of a third subordinate priming condition (massed repetitions), there are 6 conditions rather than the 4 used in Experiment 7. As the number of stimuli was limited to 60, this potentially leaves the present experiment somewhat lacking in power. In order to address this, a different procedure of selecting a random effects structure for the mixed effects models will be followed. Matuschek et al. (2017) provide evidence to suggest that whilst the maximal random effects structure (Barr et al., 2013) is desirable for suitably-powered designs because it minimises Type I error, it can in fact reduce power unnecessarily. Matuschek and colleagues show that, for less well-powered designs, the use of a *minimal* random effects structure achieves higher power without inflating Type I error, provided that a model selection criterion is used to select a random effect structure supported by the data. This can improve the balance between these two key elements of Type I error and power in statistical analyses.

Briefly, the Matuschek et al. (2017) approach starts with the maximal random effects structure model (with full fixed effects) and gradually reduces random effects complexity until a further reduction would result in a significant *loss* of goodness-of-fit (as measured with a likelihood ratio test). This approach was followed for all RT and error analyses for the present experiment. Where a simpler random effects model is reported, it did not result in a significant loss of goodness-of-fit compared to the more complex model. For tests and main effects, interactions and simple effects, the same model comparison approach used in Experiments 5, 6 and 7 was used here, where a model without the fixed factor of interest was compared to the full fixed effect model using a likelihood ratio test.

Reaction Time Analyses

The means in Figure 20 indicate that RTs to the subordinate picture are faster after both massed and spaced repeated subordinate priming, compared to the unprimed baseline. On the contrary, it seems that RTs to the dominant picture are again slower after spaced subordinate priming, although perhaps not after massed subordinate priming. Overall, RTs seem to be faster for dominant pictures than for subordinate pictures.

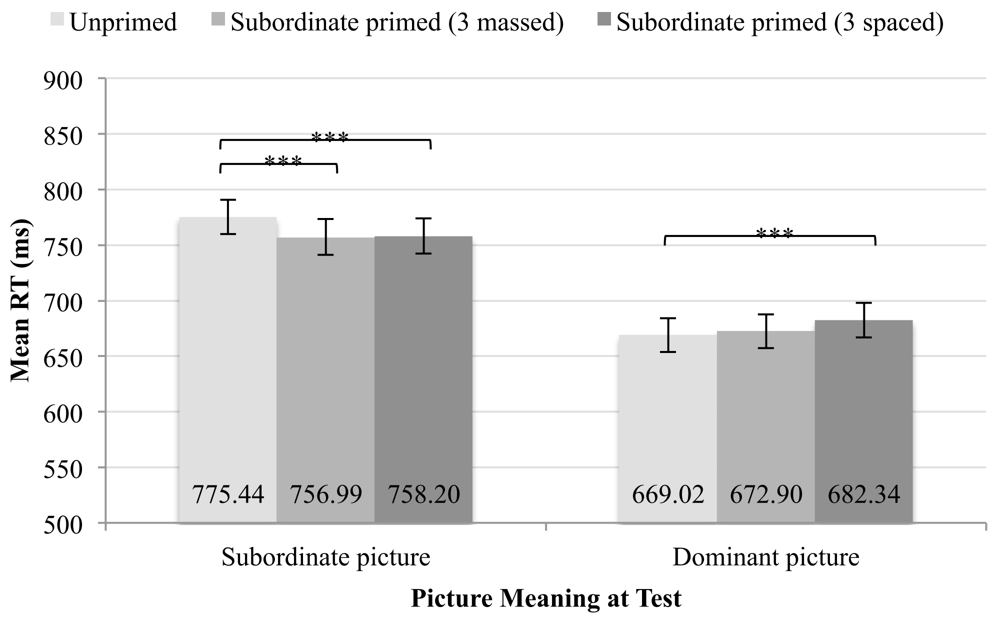

Figure 20. Mean by-subjects reaction times for Experiment 8. Responses for both the subordinate and dominant meaning picture, following either no priming, three massed, or three spaced subordinate prime repetitions. Significance level indicated with asterisks (***<.001). Error bars are adjusted for the within-subjects design.

As with Experiments 5, 6 and 7, the data were manually coded and trimmed (RTs were inverse transformed). As the subordinate prime repetitions factor had three levels, Helmert contrasts were used to code for this factor. Both factors were deviation coded (subordinate priming contrast 1: unprimed = -2/3, three massed repetitions = 1/3, three spaced repetitions = 1/3; subordinate priming contrast 2: unprimed = 0, three massed repetitions = -1/2, three spaced repetitions = 1/2; picture meaning: subordinate = -1/2, dominant = 1/2; the interaction term was specified as two separate coded factors: the multiplication of the subordinate priming contrast 1 and the picture meaning factor and the subordinate priming contrast 2 and the picture meaning factor). A model was then built with these five fixed effect coefficients.

The model reduction approach revealed that each simpler model was not significantly worse than the former more complex model, therefore the intercepts-only (simplest random effect structure) model was used. Regarding the method for the usual model comparison approach, although the subordinate prime and interactions factors were each split into two by the contrast codes, the two factors for each were either left in the model as a whole or removed as a whole for tests of the subordinate priming main effect and tests of the interaction, respectively. In each case, a model without the fixed factor of interest was compared to the full fixed effect model using a likelihood ratio test.

The main effect of repeated subordinate priming was not significant ($X^2(2) = 2.090$, $p = .352$) suggesting that, across picture meaning conditions, participant RTs to pictures were no faster after they had encountered the subordinate meaning of that word in the prime phase. The main effect of picture meaning was significant ($X^2(1) = 749.930$, $p < .001$), showing that, across subordinate priming conditions, participants were generally faster at responding to dominant than subordinate pictures. There was also a significant interaction between repeated subordinate priming and picture meaning ($X^2(2) = 23.648$, $p < .001$), showing that the effect of subordinate priming differed depending on whether they encountered the dominant or subordinate picture at test.

Simple effects analyses were conducted to further investigate the main effects and interaction. The manually-coded factors were not used for these since simple effects involving all three levels for the subordinate priming required use of the glht (general linear hypothesis testing) function in the multcomp package (version 1.4-1;

Hothorn et al., 2008), which cannot be computed using manually-coded factors. This glht function compares each of the three levels of the subordinate priming factor (at just one level of the picture meaning factor, since it is a simple effect) at a time in a single, Tukey-corrected step and outputs β, *SE* and *z* values. Where the simple effects did not require all three levels of the subordinate priming factor, the usual model comparison approach was implemented (X^2 values with Bonferroni-corrected *p* values reported). This does not have a negative impact on the results, since the simple effects require only one factor per model and manual coding is only necessary when two or more factors are included in a model (see Experiment 5 Results for details).

The simple effect of all three subordinate priming levels at only the level of the subordinate picture at test revealed a significant difference between unprimed and massed subordinate priming (β = 0.041, *SE* = 0.010, *z* = 4.189, *p* < .001), a significant difference between unprimed and spaced subordinate priming (β = 0.044, *SE* = 0.010, *z* = 4.585, *p* < .001), although no significant difference between massed and spaced subordinate priming (β = -0.004, *SE* = 0.010, *z* = -0.391, *p* = .919). These results show that, compared to the unprimed baseline, participants are faster to respond to the subordinate picture at test if they had encountered three subordinate repetitions at prime, with no difference in RTs between massed and spaced repetitions.

The simple effect of all three subordinate priming levels at only the level of the dominant picture at test revealed no significant difference between unprimed and massed subordinate priming (β = -0.019, *SE* = 0.010, *z* = -1.821, *p* = .163), a significant difference between unprimed and spaced subordinate priming (β = -0.037, *SE* = 0.010, *z* = -3.658, *p* < .001), and again no significant difference between massed and spaced subordinate priming (β = 0.019, *SE* = 0.010, *z* = 1.828, *p* = .160). These simple effects show that participants are significantly *slower* to respond to the dominant meaning after three spaced subordinate repetitions, but not after three massed subordinate repetitions.

Three further simple effects were conducted to investigate the effect of different picture meanings at each level of subordinate priming. These revealed a significant difference between subordinate and dominant pictures at each level of priming (unprimed: $X^2(1) = 368.110$, $p < .001$; massed subordinate primed: $X^2(1) = 215.380$, $p < .001$; spaced subordinate primed: $X^2(1) = 174.620$, $p < .001$). Overall, these analyses show that participants are faster to respond to dominant pictures than subordinate pictures regardless of the presence or spacing of the three subordinate priming repetitions.

Error Rate Analyses

The means in Figure 21 indicate that there are fewer errors for the subordinate picture at test after both massed and spaced repeated subordinate priming, compared to the unprimed baseline. However, it also suggests that spaced repetitions might also reduce errors for the dominant picture, compared to the unprimed and massed conditions. Again, there seem to be fewer errors overall for dominant pictures compared to subordinate pictures.

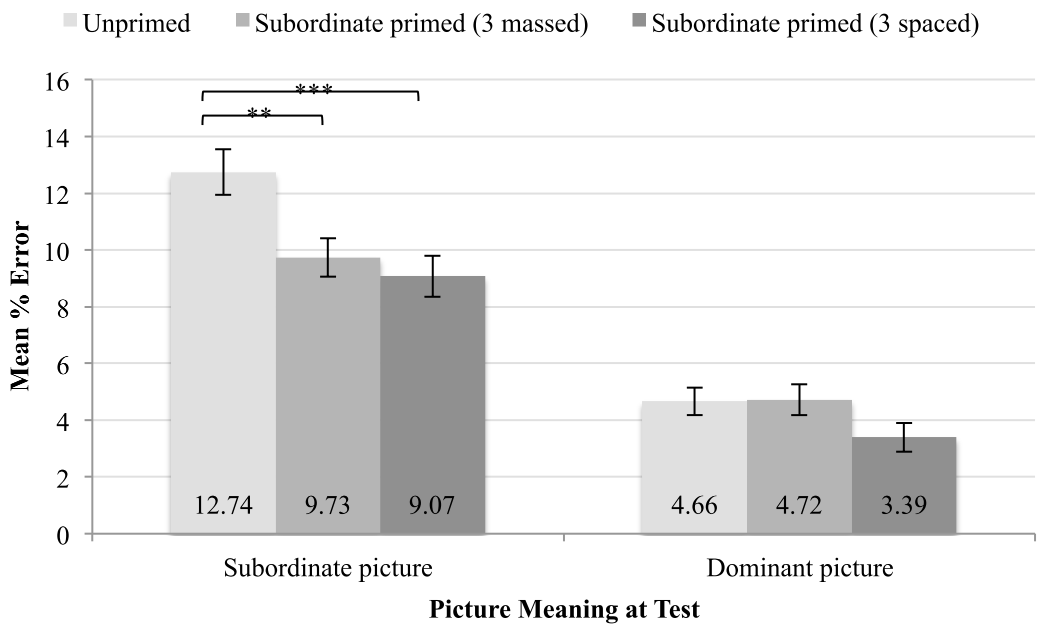

Figure 21. Mean by-subjects percentage error for Experiment 8. Responses for both the subordinate and dominant meaning picture, following no priming, three massed or three spaced subordinate prime repetitions. Significance level indicated with asterisks (** <.01, ***<.001). Error bars are adjusted for the within-subjects design.

The errors were trimmed using the same method as the error analyses in Experiments 5, 6 and 7, yet coded and analysed using the same method as the RT analyses for the present experiment. The model reduction approach showed each simpler model *not* to be a significantly worse fit than the former more complex model, therefore the intercepts-only model was used.

The model comparison approach indicated a significant main effect of repeated subordinate priming (X^2 (2) = 14.194, p <.001) suggesting that, overall, participants made fewer errors to pictures when they had encountered the subordinate meaning of that word in the prime phase. In addition, there was a significant main effect of picture meaning (X^2 (1) = 153.590, p < .001), which suggests that, across subordinate priming conditions, participants made fewer errors when responding to dominant than subordinate pictures. However, there was no significant interaction between repeated subordinate priming and picture meaning (X^2 (2) = 2.632, p = .268), indicating that the effect of subordinate priming did not differ depending on whether participants encountered the dominant or subordinate picture at test. Therefore, subordinate priming did not affect error rates overall.

The simple effect of all three subordinate priming levels at only the level of the subordinate picture at test revealed a significant difference between unprimed and massed subordinate priming (β = 0.390, *SE* = 0.122, z = 3.207, p = .004), a significant difference between unprimed and spaced subordinate priming (β = 0.535, *SE* = 0.123, z = 4.345, p < .001), although no significant difference between massed and spaced subordinate priming (β = -0.145, *SE* = 0.130, z = -1.111, p = .507). These results show that, compared to the unprimed baseline, participants make fewer errors when responding to the subordinate picture at test if they had encountered three subordinate repetitions at prime, with no difference in error rates between massed and spaced repetitions.

The simple effect of all three subordinate priming levels at only the level of the dominant picture at test revealed no significant difference between unprimed and massed subordinate priming (β = -0.021, *SE* = 0.176, z = -0.117, p = .993), no significant difference between unprimed and spaced subordinate priming (: β = 0.354, *SE* = 0.190, *z* = 1.867, *p* = .148), and again no significant difference between massed and spaced subordinate priming (β = -0.375, *SE* = 0.189, *z* = -1.984, *p* = .116). These simple effects show that repeated subordinate priming, regardless of spacing, did not increase the number of errors participants made when responding to the dominant picture.

Three further simple effects were conducted to investigate the effect of different picture meanings at each level of subordinate priming. As with the RTs, these revealed a significant difference between subordinate and dominant pictures at each level of priming (unprimed: X^2 (1) = 75.846, $p < .001$; massed subordinate primed: X^2 (1) = 34.142, $p < .001$; spaced subordinate primed: X^2 (1) = 46.685, $p < .001$). Overall, these analyses show that participants make fewer erroneous responses when responding to dominant pictures than subordinate pictures regardless of the presence or spacing of the three subordinate priming repetitions.

Awareness Analyses

There were two awareness measures: awareness of experimental aim (coded in the same way as Experiments 5, 6 and 7) and awareness estimate (rescaled and analysed as in Experiments 6 and 7). For the awareness of experimental aim question, 157 participants were unaware of the aim (for subordinate picture test condition only[24]: RT *mean* = 765.63ms, *SD* = 102.89ms, percentage error *mean* = 10.62%, *SD* = 8.19%) and only 11 participants were fully/partially aware of the aim (for subordinate picture test condition only: RT *mean* = 710.26ms, *SD* = 86.80ms, percentage error *mean* = 8.83%, *SD* = 4.99%). Hence, as with Experiments 5, 6 and 7, there was an insufficient number of participants (only 6.5%) in the "aware" category to run an analysis to examine whether priming interacts with awareness of aim.

The awareness estimate data were continuous, indicating participants' estimates of the percentage of ambiguous words in the word association test that had

[24] As with Experiment 5, only the prime-test congruent meaning condition (subordinate picture) was included.

been presented earlier in the experiment as a less explicit measure of awareness, (word estimate *median* = 43.5, *range* = 0-131, skewed distribution). These estimate data were rescaled (divided by 100) and centred. Subordinate priming levels (three massed and three spaced) were combined such that the priming factor indicated whether a trial was unprimed or primed. The model with priming slopes did not converge for the RT data (leaving an intercepts-only model) but did converge for the error data. A model comparison showed that the interaction between repeated subordinate priming and awareness estimate was not significant for RTs or errors (X^2 (1) = 0.129, p = .719; X^2 (1) = 1.651, p = .199, respectively), again indicating that participants' awareness of how many test words were repeated from the prime phase did not influence priming effects.

Discussion

There were two aims of the present experiment. The first was to investigate whether the effect from Experiment 7 of three spaced subordinate repetitions reducing dominant meaning availability replicated. The second was to investigate whether, in a semantic relatedness task, three spaced subordinate repetitions boosted priming compared to three massed subordinate repetitions, and whether three massed repetitions primed compared to the unprimed baseline (i.e. whether the findings of Experiment 3 are consistent across tests). In the prime phase, participants encountered the subordinate meaning in three different sentences either in immediate succession within a prime block (massed), three times distributed across prime blocks (spaced), or they did not encounter the meaning at all (unprimed baseline). They were then tested with the semantic relatedness task after a filler task.

The present experiment replicated findings on the link between priming and awareness from Experiments 6 and 7 - participants' awareness of how many test words were repeated from the prime phase did not affect subordinate word-meaning priming. Additionally, there were too few participants aware of the prime-test manipulation for analyses with this measure (6.5%). It seems that the semantic relatedness test consistently minimises participant awareness of the experimental aim.

The findings also replicated Experiment 7 in that participants were significantly faster when responding to the subordinate picture following three spaced subordinate repetitions, compared to the unprimed condition. Participants also made significantly fewer errors to the subordinate picture after three spaced repetitions. The present results also directly replicated those of Experiment 7, where participants were significantly *slower*, when responding to dominant pictures following spaced subordinate priming (compared to the unprimed condition). Once again, this shows that it is possible for an increase in the availability of the primed meaning to cause a decrease in the availability of the unprimed meaning. This is further support for the class of model in which meaning availabilities are linked to one another, such as with the distributed connectionist model (Rodd et al., 2004; Rodd et al., 2013)

The present experiment also introduced a massed repetition condition for the first time with the semantic relatedness test. Experiment 3 showed with a word association test that spaced repetitions significantly boosted word-meaning priming compared to massed repetitions. Based on these findings, it was predicted that massed repetitions would lead to significantly faster RTs for the subordinate picture, compared to the unprimed baseline, but significantly slower RTs (i.e. a smaller word-meaning priming effect) than spaced repetitions.

The data partially supported these predictions. Whilst three massed repetitions significantly speeded RTs to the primed subordinate picture compared to the unprimed baseline (consistent with the prediction), there was no additional benefit of spacing on the responses to the primed meaning: massed repetitions did *not* produce significantly slower RTs than the spaced repetitions (inconsistent with the prediction). These findings show that, when measured by the semantic relatedness test, massed and spaced repetitions do not differ significantly in their word-meaning priming effects, at least for the picture that is related to the primed meaning. This suggests that the number of repetitions, not the spacing of the repetitions, might be crucial for producing a repetition effect in the semantic relatedness experiments, at least when the primed meaning is measured. This seems to directly contradict the findings of

Experiment 3, in which spaced repetitions boosted subordinate priming compared to massed repetitions.

The effect of three massed subordinate repetitions on the unprimed, dominant meaning was inconclusive. This condition did not significantly differ from either the unprimed baseline or the three spaced repetition condition. Unlike spaced subordinate repetitions, massed repetitions did not significantly reduce availability to the unprimed meaning. However, as spaced repetitions did not significantly reduce unprimed meaning availability compared to massed repetitions, it is not possible to draw conclusions about the massed repetition condition from this experiment. Despite this non-significant difference, these findings hint that, whilst three spaced repetitions interfere with availability of the unprimed meaning, the lack of spacing between the three massed subordinate repetitions might render them insufficient to have the same effect. It does seem, however, that priming might benefit from spaced repetitions on two levels: faster RTs for the subordinate meaning, as well as slower RTs for the dominant meaning.

Due to the unexpected findings in the present experiment that massed repetitions were not significantly different from spaced repetitions for boosting primed meaning availability but seemed to be inferior (a trend) to spaced repetitions for reducing availability of the unprimed meaning, Experiment 9 will include a one repetition condition to compare with the massed and spaced conditions. This will allow for the investigation of whether massed repetitions are superior to one repetition for boosting primed meaning availability and reducing unprimed meaning availability. This direct comparison of one repetition, three massed repetitions and three spaced repetitions will help to clarify any differences between these three conditions.

The next experiment will therefore be an exact replication of the priming manipulation used in Experiment 3, but the word association test will be replaced with the semantic relatedness picture test. In doing so, this experiment aims to investigate whether massed repetitions lead to significantly more subordinate word-meaning priming than one repetition. Whilst this was *not* the case for Experiment 3 (massed repetitions did not prime significantly more than one but did prime significantly less than spaced), the findings of Experiment 8 suggest that massed repetitions *might* lead to more priming than one repetition when tested with the semantic relatedness task.

Experiment 9 – one, three massed & three spaced subordinate repetitions

The present experiment included an additional prime condition to Experiment 8 in which only one subordinate sentence was presented to participants. This meant that the four subordinate prime conditions were the same as those of Experiment 3: unprimed, one repetition, three massed repetitions and three spaced repetitions. See Figure 22 for the order of the tasks in the experiment.

The aim was to investigate whether three subordinate massed repetitions improve performance to the subordinate picture compared to one subordinate repetition. Due to the limit of available items (to compensate for the additional prime condition) the dominant picture condition was omitted from the semantic relatedness test, leaving only a subordinate picture test condition.

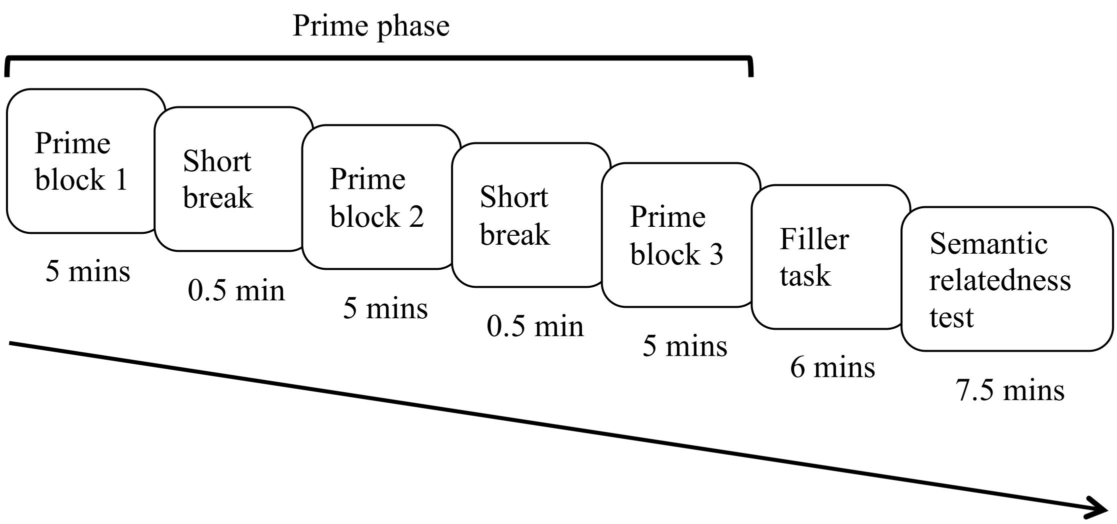

Figure 22. Experiment 9 task order, including prime phase elements, filler task and test, with the mean duration of each task. The mean prime-test delay is 17 minutes (rounded down from 17.25 for ease of reference).

Method

Participants

Two-hundred-and-four native British English speakers participated in the current experiment (93 females; mean age = 29.85, range = 18-45). All participants met the demographic requirements outlined in Experiment 5 but, for this experiment, participants were recruited via the University College London online recruitment system or advertisements on the university campus and paid the standard rate at the time of £8/hour.

Design

This experiment had a within-subjects design with one independent variable, subordinate priming, which had four levels: unprimed, one repetition, three massed repetitions and three spaced repetitions. The dependent variables were reaction time and accuracy in the semantic relatedness test, but crucially involved only the *subordinate* meaning pictures as relatedness probes for experimental trials (unlike Experiments 5-8, dominant pictures were not used in the test phase for the present experiment to reduce the number of conditions and therefore maximise power).

As a result of the four prime conditions, there were four prime versions, which ensured that each item appeared in each of the four priming conditions but for different participants. In each version, 15 of the total 60 ambiguous words were included in each of the four conditions at prime. The massed and spaced sentences were distributed within/across blocks, respectively, in the same way as in Experiment 8. The 15 spaced experimental items per version were naturally distributed over the prime phase, with one sentence per prime block (sentence 1 in block, sentence 2 in block 2, sentence 3 in block 3). The 15 massed experimental items per version were divided into 3 groups, allocating 5 items to each block (presented in immediate succession).

At the time of running the present experiment[25], a lack of filler stimuli prevented the possibility of filler sentences being primed and then at test paired with unrelated pictures so that the number of primed, *experimental*, ambiguous, *related* trials at test could be maximised. With a maximum of 60 experimental ambiguous words, and due to this impossibility of filler prime sentences, half of the experimental ambiguous words had to be paired with unrelated pictures at test and were therefore discarded for analyses. This resulted in "analysable" data from only 30 experimental words per participant, therefore 7 or 8 per prime condition and therefore only 3-4 for the primed dominant picture condition and 3-4 for the primed subordinate picture condition at test. This would have resulted in a sixteen-version experiment. Clearly, this design would not have been suitably powered.

To maximise power, the solution was to have only *one* semantic relatedness test condition, that is, only test with the subordinate picture. Whilst this could not test availability of the unprimed meaning, it could test whether the effect of massed repetitions being equivalent to spaced repetitions for semantic relatedness tests (as opposed to them being inferior to spaced repetitions in word association tests) was replicable, and how massed repetitions compared to one repetition. This design therefore required only two test versions (rather than the four required if there had been two semantic relatedness test picture conditions). These two test versions were created such that for one version, the half of the experimental ambiguous words that were paired with a related picture probe for one participant were then paired with an unrelated picture probe for another participant, and vice versa. This design meant that there were "analysable" data from 7 or 8 items per version, double that of the full sixteen-version design but still half that of Experiments 5, 6 and 7. Consequently, even with only the subordinate picture test condition, this design required double the number of participants as in Experiments 5, 6 and 7 to maximise power.

[25] Experiment 9 was conducted before Experiments 5-8 but for ease of narrative it is included in the present chapter after those experiments.

Materials

All experimental prime stimuli were the same as in Experiments 5, 6, 7 and 8: sixty experimental ambiguous (although no filler ambiguous words as explained in the Design section above), sentences and probe words, where there were three different sentences per ambiguous word (as in Experiments 7 & 8), each disambiguating towards the same subordinate meaning but with varying context.

The Towers of Hanoi task was again used as the filler between prime and test. See Experiment 5 Materials for details.

For the semantic relatedness picture test there was a total of 72 items: 60 ambiguous experimental words, 30 with related pictures and 30 with unrelated pictures, and 12 unambiguous filler words – 6 related and 6 unrelated (these fillers were all unprimed in the present experiment, unlike in Experiment 5). In total at test, 62.5% of trials were primed, 83% of trials were ambiguous and 50% of trials were related.

Procedure

The procedure was very similar to that of Experiment 5; there was only one difference. The present experiment required that the subordinate prime task be split into three blocks (with a 30 second break between blocks), in order to space the three sentence repetitions, as was the case for Experiments 2, 3, 7 and 8. The break between blocks was automatic and instructed participants to wait for 30 seconds, after which the next prime block would start automatically.

Task Checks

Whilst 204 participants completed the experiment, only the data from 185 participants were analysed: nine participants were excluded for meeting one or more of the exclusion criteria outlined in Experiment 5 (with 35 minutes being the maximum time allowed for completion). A further 10 participants were removed due to technical issues that disrupted one or more of the tasks during the experiment. In addition, all items were checked for prime and test accuracy and were excluded if their accuracy, averaged across participants, was less than 70%. This resulted in the exclusion of one item ('iron') for all participants, leaving a total of 59 items in the analyses.

Results

Both the RT and error analyses for the present experiment were analysed using the method from Experiment 8 (Matuschek et al., 2017) where the simplest possible random effects structure that is suitable for the data is used. This Matuschek et al. (2017) approach to analysis was appropriate because this experiment was also slightly underpowered, since half of the ambiguous words were 'unrelated' picture trials at test, and were therefore discarded for all analyses. See Experiment 8 Results for more details. However, the fixed factor (subordinate priming; all pictures are subordinate at test for this experiment) was not manually deviation-coded in the present experiment for two reasons: (1) having four levels of one factor requires the glht function (version 1.4-1; Hothorn et al., 2008) for multiple, corrected, pairwise comparisons and this is not compatible with manual coding, and (2) manual coding of factors is strictly only required when there are two or more factors in a model when the model comparison approach is being used, as explained in Experiment 5; since the present analysis will have only one factor per model (subordinate priming), manual coding was not necessary[26]. As with Experiments 5, 6, 7 and 8, the RT and error data were trimmed (and RTs were inverse transformed).

[26] Since statistics are not reported directly from the model summaries in R for *any* of the analyses in this chapter (they are all reported from model comparisons or the glht comparison function), results across experiments can be compared despite the differences in manual and automatic coding – these differences will not affect the interpretation of the results.

Reaction Time Analyses

The means in Figure 23 suggest that, compared to the unprimed baseline, one subordinate repetition speeds RTs to the subordinate picture. In turn, both three massed and three spaced subordinate repetitions speed RTs compared to one repetition, with massed repetitions providing the largest boost in the speed of responses.

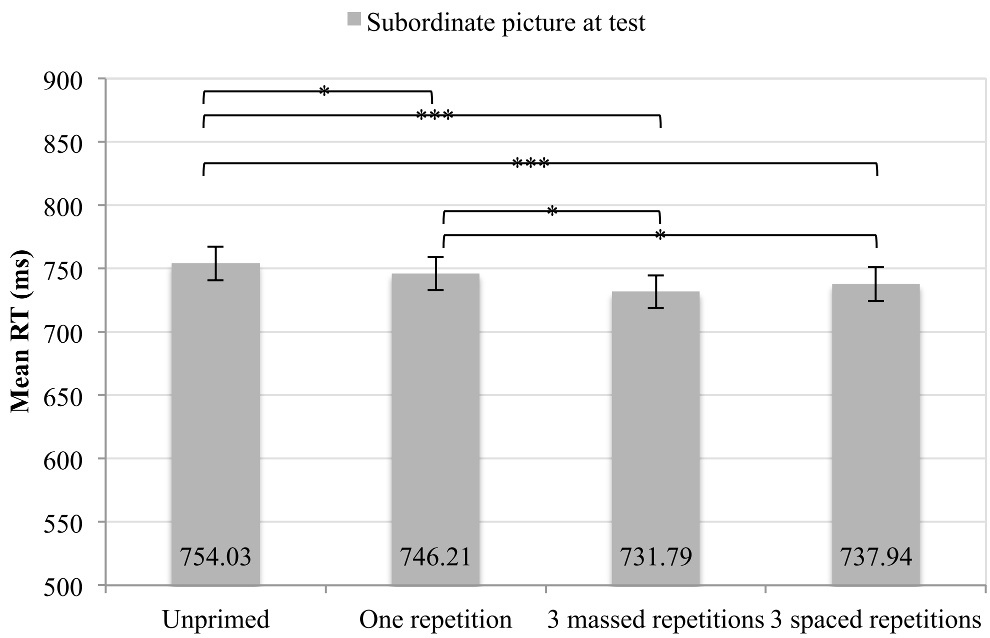

Figure 23. Mean by-subjects reaction times for Experiment 9. Responses for the subordinate meaning picture, following either no priming, or one, three massed, or three spaced subordinate prime repetitions. Significance level indicated with asterisks (*<.05, ***<.001). Error bars are adjusted for the within-subjects design.

The RT data were trimmed and inverse transformed as in Experiments 5, 6 and 8. The model reduction approach confirmed that each simpler model was not significantly worse than the former, more complex model, therefore the intercepts-only model was used. The model comparison showed that the main effect of subordinate priming was significant (X^2 (3) = 39.610, p < .001) suggesting that participant RTs to pictures were faster after they had encountered the subordinate meaning of that word in the prime phase.

Tukey-corrected pairwise comparisons were conducted to investigate the significant main effect. These revealed that, compared to unprimed RTs, participant RTs to subordinate meanings were significantly faster following one subordinate repetition ($\beta = 0.026$, $SE = 0.010$, $z = 2.645$, $p = .041$), three massed repetitions ($\beta = 0.054$, $SE = 0.010$, $z = 5.515$, $p < .001$) and three spaced repetitions ($\beta = 0.052$, $SE = 0.010$, $z = 5.233$, $p < .001$). In turn, compared to one repetition, RTs were significantly faster following three massed and three spaced repetitions ($\beta = 0.028$, $SE = 0.010$, $z = 2.893$, $p = .020$; $\beta = -0.025$, $SE = 0.010$, $z = -2.611$, $p = .045$; respectively). However, there was no significant difference between RTs from three massed and three spaced repetitions ($\beta = 0.003$, $SE = 0.010$, $z = 0.284$, $p = .992$). These results suggest that three massed and three spaced subordinate prime repetitions are both successful in speeding responses to the subordinate picture at test. Whilst one subordinate repetition also achieves this, it is to a significantly lesser extent than both three massed and three spaced repetitions.

Error Rate Analyses

The means in Figure 24 for percentage error replicate the pattern of the results for the RTs for the present experiment. The figure suggests that, compared to the unprimed baseline, one subordinate repetition reduces errors to the subordinate picture. In turn, both three massed and three spaced subordinate repetitions reduce errors compared to one repetition, with massed repetitions providing the largest reduction in error responses.

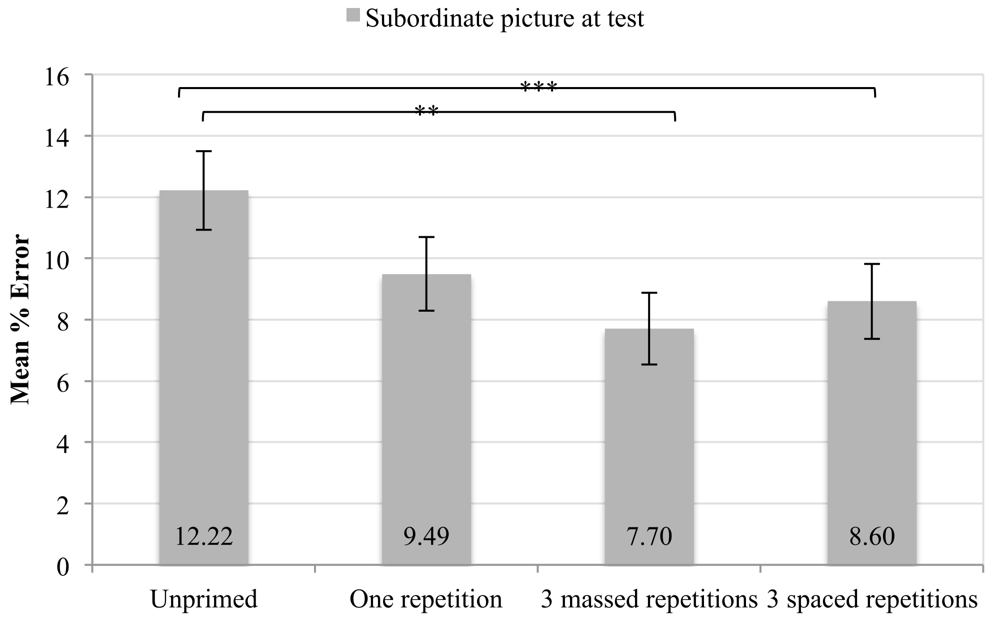

Figure 24. Mean by-subjects percentage error for Experiment 9. Responses for the subordinate meaning picture, following either no priming, or one, three massed, or three spaced subordinate prime repetitions. Significance level indicated with asterisks (**<.01, ***<.001). Error bars are adjusted for the within-subjects design.

The model reduction approach showed each simpler model not to be a significantly worse fit than the former, more complex one, therefore the intercepts-only model was used. The model comparison approach revealed a main effect of subordinate priming (X^2 (3) = 19.240, p < .001), suggesting that participants made fewer errors when responding to the subordinate pictures at test if they had encountered the subordinate meaning of that word in the prime phase.

Tukey-corrected pairwise comparisons were conducted to investigate the significant main effect. These showed that, compared to unprimed responses, participants made marginally fewer errors to the subordinate picture following one subordinate repetition (β = 0.337, *SE* = 0.133, z = 2.545, p = .053), and significantly fewer errors following three massed repetitions (β = 0.575, *SE* = 0.139, z = 4.142, p < .001) and three spaced repetitions (β = 0.478, *SE* = 0.136, z = 3.522, p = .002).

However, three massed and three spaced repetitions did not significantly reduce error any more than one repetition (β = 0.238, *SE* = 0.145, *z* = 1.645, *p* = .353; β = -0.140, *SE* = 0.141, *z* = -0.993, *p* = .753, respectively) and there was no significant difference between errors following massed and spaced repetitions (β = 0.098, *SE* = 0.147, *z* = 0.663, *p* = .911). These results suggest that three massed and three spaced subordinate prime repetitions are similarly successful in reducing the number of erroneous semantic relatedness responses to the subordinate picture of the same word-meaning at test. Whilst there is a marginal trend to suggest that one subordinate repetition might also reduce errors, it is not to a significantly lesser extent than both three massed and three spaced repetitions.

Awareness Analyses

There were two awareness measures: awareness of experimental aim (coded in the same way as Experiments 5, 6, 7, 8) and awareness estimate (rescaled and analysed as in Experiments 6, 7, 8). For the awareness of experimental aim question, the awareness data from 2 participants were missing, leaving a total of 183 participants. Of those, 173 participants were unaware of the aim (for subordinate picture test condition only, since the present experiment tested only with the subordinate picture: RT *mean* = 742.60ms, *SD* = 87.16ms, percentage error *mean* = 9.62%, *SD* = 7.08%) and only 10 participants were fully/partially aware of the aim (for subordinate picture only: RT *mean* = 740.70ms, *SD* = 53.45ms, percentage error *mean* = 7.60%, *SD* = 4.55%). Hence, as with Experiments 5, 6, 7 and 8, there was an insufficient number of participants (only 5.5%) in the “aware” category to run an analysis to examine whether priming interacts with awareness of the aim of the experiment.

The awareness estimate data were continuous, indicating participants’ estimates of the percentage of ambiguous words in the word association test that had been presented earlier in the experiment as a less explicit measure of awareness, (word estimate *median* = 24, *range* = 0-72, skewed distribution). These estimate data were rescaled (divided by 100) and centred. Subordinate priming levels (one, three massed and three spaced) were combined such that the priming factor indicated whether a trial was unprimed or primed. The model with priming slopes did not converge for the RT or error data, leaving intercepts-only models. A model comparison showed that the interaction between subordinate priming and awareness estimate was not significant for RTs or errors (X^2 (1) = 2.635, p = .105; X^2 (1) = 0.282, p = .595, respectively), again indicating that participants' awareness of how many test words were repeated from the prime phase did not influence priming.

Discussion

The aim of the present experiment was to investigate whether three subordinate massed repetitions improved performance to the subordinate picture compared to one subordinate repetition. Participants encountered the subordinate meaning either once, three times massed within, or three times spaced across, blocks in the prime phase. After a filler task, they responded to only the subordinate picture of the ambiguous words (on experimental trials) in the semantic relatedness test.

The present experiment replicated findings on the link between priming and awareness from Experiments 6, 7 and 8 - awareness of the number of test words repeated from the prime phase had no significant effect on subordinate priming. Additionally, there were too few participants aware of the prime-test manipulation for analyses with this measure (5.5%). This is especially reassuring given the repeated nature of the prime stimuli.

The main results show that, compared to the unprimed condition, participants were significantly faster following all levels of priming (one repetition, three massed, three spaced) and made significantly fewer errors to the subordinate picture following three massed and three spaced repetitions. Crucially, participants were significantly faster and made significantly fewer errors in the massed repetition compared to the one repetition condition, indicating that three massed repetitions boosted priming compared to one repetition. Whilst three massed and three spaced repetitions significantly speeded responses compared to one repetition, there was no significant difference between massed and spaced repetitions. This is consistent with the findings of Experiment 8, and suggests that the temporal spacing of repetitions is *not* crucial for a repetition benefit when interpretations of ambiguous words are tested using a semantic relatedness picture test. This seems at odds with the finding from using the same priming manipulation but with a word association test (Experiment 3), where massed repetitions did not significantly boost priming compared to a single repetition and primed significantly less than spaced repetitions (and therefore suggested that temporal spacing *was* crucial for producing a repetition benefit).

Interestingly, testing with only the subordinate pictures, and with a high percentage of ambiguous words, in the semantic relatedness task did not seem to alter the pattern of results between massed and spaced repetitions compared to Experiment 8, which used both subordinate and dominant pictures. This suggests that participants did not adopt a different response strategy when they were only presented with the less common (and therefore more unexpected) meaning of each ambiguous word, and mostly ambiguous words, at test. This finding is reassuring, as it indicates that the semantic relatedness test is reliable and not largely sensitive to changes in stimulus type, which demonstrates the merits of this measure of word-meaning priming.

General Discussion

The effect of priming on the availability of the primed meaning

The overall aim of the five experiments in this chapter was to explore whether priming, in increasing the availability of the primed meaning, reduces the availability of the unprimed meaning. Understanding the relationship between lexical-semantic representations of an ambiguous word would give an insight into the nature of these representations and the mechanism(s) underlying word-meaning priming. Changes to meaning availability following priming were determined by measuring reaction times and error rates to either the subordinate or dominant picture of a given word, independently.

This chapter contributes to the existing findings on word-meaning priming. Experiments 5, 7, 8 and 9 showed that word-meaning priming can speed (and in some cases improve the accuracy of) the correct interpretation of an ambiguous word in a *constrained* context at test. Specifically, a single encounter with the subordinate meaning biases the interpretation of that word when it is encountered 13 to 18 minutes later in a semantic relatedness test (Experiments 5 & 9). Compared to the unprimed baseline, RTs to the picture related to the primed subordinate meaning were significantly faster after subordinate priming, indicating that the earlier encounter with the subordinate meaning increased the availability of this meaning. This is in addition to the findings from Experiments 1, 2 and 3 of Chapter 2, which showed that word-meaning priming can, on a proportion of trials, “flip” the interpretation of an ambiguous word (towards the primed meaning) in a *neutral* test context. This thesis therefore provides a total of 7 experiments that have consistently replicated the word-meaning priming effect. Both word association and semantic relatedness tests consistently show a significant shift in responses towards the subordinate meaning after one recent encounter with this subordinate meaning. It is reassuring that these different measures are consistent in this way, and support existing findings on word-meaning priming effects (Gilbert et al., 2018; Rodd et al., 2016; Rodd et al., 2013). Together, these results indicate that word-meaning priming effects are not limited to a single priming measure or to neutral test contexts.

Furthermore, both three massed and three spaced subordinate prime repetitions also significantly speeded responses to the picture of the primed subordinate meaning, compared to the unprimed baseline (Experiments 7, 8 & 9), with no significant difference between massed and spaced repetitions (Experiments 8 & 9). Surprisingly, for the primed meanings (Experiments 8 & 9), these experiments did not straightforwardly replicate the spacing benefit seen in Chapter 2. We will return to this aspect of the data later.

It is reassuring that awareness of the experimental aim is not a critical factor for priming to occur. Across semantic relatedness experiments, either so few participants were aware of the prime-test link that awareness data could not be analysed, or awareness analyses showed no significant interaction between the magnitude of priming and participants' awareness of the experimental manipulation. This was the case when the magnitude of priming was modelled in terms of reaction times and error rates. This lack of awareness is especially reassuring given that the repetition of ambiguous words (each in a different sentence) in Experiments 7, 8 and 9 may have increased the salience of this experimental manipulation. In fact, very few participants reported being aware of the priming manipulation link between the prime and test phase of the experiment. Instead, since the prime and test both used a semantic relatedness task (prime: sentences with word probes; test: words with picture probes), many participants believed that the prime and test were comparing the quality of semantic relatedness performance in two different styles of the test (for example, 'to see whether people associate related or unrelated words quicker by reading a word or looking at a picture'). This trend suggests that the similarity of tasks in the prime and test phases might even reduce awareness of the priming manipulation compared to word association. Combined with the finding from Experiments 1, 2 and 3 in Chapter 2, and from Rodd et al. (2016), it is clear that word-meaning priming is not driven only by *conscious* attempts to recall previous experience of word-meanings in the experiments. Word-meaning priming is clearly a robust effect.

The effect of priming on the availability of the unprimed meaning

The major finding from the five present experiments is that priming one meaning, and therefore increasing its availability, *can*, but does not always, reduce the availability of the unprimed meaning. Experiment 5 showed that a single encounter with the subordinate meaning increases its availability without significantly reducing the availability of the unprimed, dominant meaning. After hearing the subordinate meaning in context once in the prime phase, participants were significantly faster when responding to the primed meaning, but were not significantly slower in their responses to the unprimed meaning. Experiments 7 and 8 showed that three spaced encounters with the subordinate meaning also significantly increased its availability. Here, however, the priming *did* significantly reduce the availability of the unprimed meaning. After encountering the subordinate meaning in three spaced sentences in the prime phase, participants not only responded significantly faster (and significantly more accurately in Experiment 8) to the primed, subordinate meaning, but they also responded significantly more slowly to the unprimed, dominant meaning. This shows that priming can make the meaning that has *not* been encountered recently less likely to be available.

Whilst Experiment 6 showed no significant impact of priming with the dominant meaning on the availability of the unprimed meaning, it also showed no significant effect of the dominant meaning at all (its significance did not withstand statistical correction for multiple comparisons). Since this dominant priming manipulation did not improve performance on the dominant meaning, it is very unlikely that this priming should have an effect on the unprimed subordinate meaning. Therefore Experiment 6 was inconclusive with respect to the effect on the unprimed meaning. Regardless, it seems that subordinate priming is a more reliable manipulation, as is clear from the replication of word-meaning priming effects across all subordinate priming experiments in the present chapter (Experiments 5, 7, 8 & 9). As discussed by Rodd et al. (2013) and in Chapter 2, the success of subordinate priming is probably due to the lower existing meaning availability of the subordinate meaning, which gives it a greater potential to be boosted (by recent experience) than an already highly available, dominant meaning (Rodd et al., 2013).

Additionally, care must be taken when interpreting the null results of Experiment 5 in particular. It might be that interference with the unprimed meaning does occur after only a single prime repetition but it is difficult to detect such a small effect. More research should be carried out to fully establish the effect of a single recent experience with one word-meaning on competing word-meanings. For now, however, the finding that an increase in the availability of one meaning *can* have a negative impact on the availability of a competing meaning is extremely important for the field of language comprehension, as it has potential implications for all models of semantic ambiguity resolution.

It is not immediately clear *why* learning new information should make existing information less available than if the new information had not been encountered. It seems counterintuitive from an efficiency of communication point of view. For example, a novice rower would begin to learn that the word 'square' in rowing (used to describe a position of an oar where the blade is perpendicular to the water), in addition to their existing knowledge of its 'four-sided shape' meaning. It would certainly be useful from a communication point of view for the rowing meaning of 'square' to increase in availability with increasing experience with it (similar to word-meaning priming). This is because the increase in rowing experience does make it more likely that the rowing meaning of 'square' will be encountered. A more available rowing meaning therefore makes comprehension more efficient as it reflects the linguistic environment. However, since these rowing experiences are likely to be temporally spaced (rowers might row once, or several times, per week), and these spaced repetitions should decrease the availability of the unencountered meaning (Experiments 7 & 8), it is probable that the existing shape meaning of 'square' would decrease in availability. It is not clear why this reduction in availability would be advantageous, since the non-rowing shape meaning of 'square' is no less likely to be encountered in everyday life because of the new rowing experience. It therefore seems counterintuitive that learning the rowing meaning could reduce the availability of the non-rowing meaning. This raises the possibility that language comprehension processes are not always as efficient as we might assume. However, it is possible that whilst the shape meaning of 'square' is no less likely in absolute terms, it is less likely in relative terms (less likely as a proportion of the overall number of times where ‘square’ is encountered, due to the newly learnt meaning). The following section will explore possible explanations for this potentially counterintuitive finding.

Potential mechanisms

There are two possible ways in which multiple spaced prime repetitions could speed responses to the primed meaning whilst also slowing responses to the unprimed meaning. One possibility is that priming directly changes the underlying lexical-semantic representation of the unprimed meaning. The increase in the primed meaning has a direct effect on the availability of the unprimed meaning at the time of, or as a direct consequence of, priming. For example, at the same time as increasing the connection strengths for the *primed* meaning (either connections between the form and meaning layers, or connections within the meaning layer, of the network), priming may also have decreased the connections for the *unprimed* meaning. This type of ‘unlearning’ would reduce the availability of the unprimed meaning. However, the current data do not necessarily require that ‘unlearning’ occurs. Although such an effect *may* underlie some (or all) of the effect on the unprimed meaning, current discussion of similar effects within word-form learning suggests a different, more plausible explanation: that the unprimed meaning becomes less readily available due to increased *competition* from the primed meaning.

The present results are analogous to the competition effects found in the word-form processing literature. Gaskell and Dumay (2003) showed that learning the novel word ‘cathedruke’ slowed recognition times of its overlapping competitor ‘cathedral’ when tested several days after training. Their findings demonstrate that the availability of the newly learned linguistic information can interfere with access to existing (related) information. It is not that ‘cathedral’ has been unlearned, but that it becomes more difficult to access when ‘cathedruke’ becomes a competitor. A similar explanation can account for the present findings: the increased availability of the primed meaning, which arises as a consequence of recent experience, could interfere with the alternative unprimed meaning of the word. Here the emphasis is on the change in *access* to the unprimed representation *at the time of testing*. This account does not assume any 'unlearning' of the unprimed meaning.

Davis and Gaskell (2009) proposed an account of word-form learning based on principles from complementary learning systems (CLS) theories of memory (e.g. McClelland, McNaughton, & O'Reilly, 1995). As summarised by Tamminen and Gaskell (2013b), according to the account, newly learned words are initially stored as episodic memory representations that are independent from existing knowledge (i.e. long-term lexical representations in the lexicon). These episodic memory representations only become stable representations, fully integrated into the mental lexicon, after either (a) spaced learning, i.e. repeated new exposures to these words over time, without sleep (Lindsay & Gaskell, 2013), or (b) memory consolidation of the word encounter "offline", such as during sleep (Dumay & Gaskell, 2007; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). Hence, it is possible to know whether a word has been integrated into the lexicon if it engages, and therefore interferes, with long-term lexical knowledge. Whilst Davis and Gaskell's (2009) CLS account specifically considers word-form representations rather than lexical-semantic representations (the novel words were learned without a meaning), it provides a relevant framework for understanding lexical-semantics. Their account considers how new information about a new word is integrated into the lexicon, which already contains overlapping phonological competitors. Similarly, the present word-meaning priming experiments investigate how new information about an *existing* word is integrated into the lexicon where meaning competitors already exist.

In summary, this CLS account provides a framework for understanding the current results. It seems from Experiments 7 and 8 that three spaced repetitions of a particular word-meaning might produce a consolidated change in the lexicon that results in an observable interference effect on the unprimed meaning. This account will be discussed in more detail later.

Implications for models of semantic ambiguity

At the beginning of this chapter, two classes of model were outlined to provide a basis for understanding how priming might affect the availability of the unprimed meaning. The first possibility was that, in line with the principles of the reordered access model (Duffy et al., 1988), the representations of the different meanings of an ambiguous word are completely independent and do not compete or interfere with each other. Successful priming would increase the availability of the primed meaning but it would *never* change the availability of the unprimed meaning. However, Experiments 7 and 8 demonstrated that spaced priming *does* significantly reduce the availability of the unprimed meaning. After encountering the subordinate meaning in three temporally spaced sentences in the prime phase, participants not only responded faster to the primed, subordinate meaning, but they responded more slowly to the unprimed, dominant meaning (compared to the unprimed baseline)[27]. Whilst the reordered access model was developed to explain immediate context effects on comprehension rather than effects of recent experience, this finding adds to an increasing body of evidence suggesting that this specific aspect of the model is not correct (e.g. Chen & Boland, 2008; Monsell & Hirsh, 1998; Wheeldon & Monsell, 1994).

Conversely, these data are broadly consistent with distributed connectionist models in which the availabilities of alternative word-meanings are necessarily coupled (e.g. Rodd et al., 2004). It naturally emerges from the properties of the model that any increase in the availability of the primed meaning would necessarily decrease the availability of the unprimed meaning, to some extent, due to competition/interference between the two meanings (Rodd et al., 2004). Therefore, current distributed connectionist models that include competition between word-meanings provide a straightforward explanation of the effects on unprimed meanings seen in Experiments 7 and 8.

[27] Note that Experiments 7 and 8 of the present chapter indicated that, on average, participants responded faster and more accurately to the dominant meaning than to the subordinate meaning, even after three massed or three spaced subordinate encounters in the prime phase. This suggests that even strong subordinate context did not lead to the selective access of the subordinate meaning, providing support for an initial stage of exhaustive access in comprehension.

Effects of spacing on primed and unprimed meanings

The most surprising aspect of the data presented in this chapter is the failure to find a spacing benefit for the primed meaning (Experiments 8 & 9). The different pattern of results seen for the two tasks (semantic relatedness in this chapter, word association in Chapter 2) suggest that perhaps these two tasks are tapping into two different types of information that is being learned from the prime sentences. This discrepancy might be best understood within the learning framework set out in the CLS model of novel word-learning (Davis and Gaskell, 2009). A key aspect of this account is that newly learned words are initially stored as episodic memory representations that can only integrate into the lexicon after consolidation, facilitated by spaced learning or sleep. Applying the account to word-meaning priming, it would make two clear predictions. One: all types of word encounters, regardless of repetitions or spacing, would initially be stored as episodic representations. Two: episodic representations of word encounters can be consolidated into the lexicon given sufficient temporal spacing between those encounters. This means that two types of information can be learned from experience with language: episodic and consolidated.

Specifically, the account would predict that the spaced repeated subordinate meaning would initially be stored as an episodic memory representation but that the spacing between each meaning allows time for each encounter with the subordinate meaning to be consolidated. This consolidation would integrate the recent experience with the subordinate meaning into the lexicon to produce a lasting effect of increased availability of the subordinate meaning. Since the availability of the existing subordinate meaning in the lexicon is increased, this could interfere with the availability of the competing dominant meaning in a similar way to how ‘cathedruke’, only when consolidated, interfered with access to ‘cathedral’. Consolidation has been shown to be beneficial for the retention and integration of linguistic information in adulthood (Bakker, Takashima, van Hell, Janzen, & McQueen, 2014; Kurdziel, Mantua, & Spencer, 2017), showing that consolidation is an important process involved in the continual learning from the linguistic environment.

Importantly, under the CLS-based account, competition effects only arise when the newly acquired knowledge is consolidated into the lexicon. This is consistent with our observation that 'interference effects' on the unprimed meaning only occur after three *spaced* repetitions of the prime. This account predicts that the three massed repetitions may not produce significant changes for the *unprimed* meaning, since massed repetitions do not allow for consolidation and subsequent integration into the lexicon (where interference can occur). This is consistent with the absence of a significant effect on the unprimed meaning in this massed condition (Experiment 8), but these data are somewhat equivocal as this relies on a null finding.

The CLS framework also provides a potential explanation for the absence of a spacing benefit for the *primed* meaning in the semantic relatedness experiments (Experiments 8 & 9). The responses to the primed, subordinate meaning revealed that three massed and three spaced repetitions significantly boosted availability of the subordinate meaning compared to the unprimed baseline (Experiment 8) and compared to one repetition (Experiment 9), with no significant difference between massed and spaced repetitions (Experiments 8 and 9). These data suggest that three massed and three spaced repetitions are comparable in their effects on the primed meaning. This result is somewhat surprising given the spacing benefit seen in word-meaning priming effects when tested with word association (Chapter 2). The most likely explanation for this absence of a spacing benefit on the primed meaning is that, unlike word association, these priming effects are being driven by changes to the unconsolidated episodic representations.

It is not yet entirely clear why semantic relatedness might tap into episodic memories more than word association, but it is likely to be due to the presence of context provided by the picture probes in the semantic relatedness test. These picture probes are likely to trigger recall of information from the prime phase but this only happens for the primed meaning, as for the unprimed meaning there is no relevant episode to be recalled (it was not encountered in the prime phase). This means that for the primed meaning, the nature of the semantic relatedness test might increase reliance on the available episodic memory representations. Conversely, on unprimed trials, in the absence of available episodic memory representations, perhaps the reliance on consolidated lexical-semantic representations is increased. In summary, this would mean that, for semantic relatedness, responses to the primed meaning are driven by episodic memory representations but responses to the unprimed meaning are driven by consolidated memory representations.

Specifically, for experimental trials, the primed ambiguous word was always related to the picture probe that followed it. However, the meaning of the picture was either consistent or inconsistent with the meaning encountered in the prime phase. On related, consistent (primed, subordinate picture) trials, the context provided by the picture could act as a cue to trigger recall of this meaning from when it had been encountered earlier in the prime phase. A participant can make a correct response on these trials purely by accessing episodic memory representations, as there would be an available episodic representation of this meaning. For instance, when being presented at test with the word 'bark' and a picture of tree bark (the subordinate meaning), participants should be able to recall the tree meaning of bark from the sentence 'the woodpecker clung onto the *bark*' that they encountered during priming. Recalling this information helps the participant to correctly identify the picture as related in meaning to the word 'bark' and therefore respond successfully in the semantic relatedness test. This focus on the primed meaning at test taps straight into these episodic representations. Here, priming from one repetition when testing with the primed meaning is less effective than from three massed or spaced, as there is only one episodic memory representation to guide disambiguation, as opposed to three. Yet, priming from three massed and three spaced repetitions when testing with the primed meaning should have equal effects. Since responses to the primed meaning can be generated through episodic memories, massed and spaced repetition priming benefit equally from their three episodic representations learned at prime. Spacing is irrelevant presumably because consolidated representations are not required for task success. This is exactly the pattern of results found in the present chapter.

However, on related, inconsistent (unprimed, dominant picture) trials, the context provided by the picture cannot act as a cue to trigger recall of this meaning from the prime phase, as it had not been encountered. A participant cannot make a correct response on these trials purely by accessing episodic memory representations because no such recent memory exists. To make a correct response, participants have to access consolidated word-meaning representations in the lexicon. Here, only spaced repetitions can reduce the availability of the unprimed, dominant meaning representation (through increasing the availability of the primed, subordinate meaning representation via consolidation). Even though consolidated lexical-semantic representations are accessed to complete the task on the unprimed meaning, regardless of the priming condition, only spaced repetitions can slow access to the unprimed meaning; massed repetitions do not affect access to the unprimed meaning. This is also exactly the pattern of results found in the present chapter.

In summary, the most likely explanation for the absence of a spacing benefit on the primed meaning in semantic relatedness experiments (Experiments 8 & 9) is that responses to the primed meaning are driven more by unconsolidated episodic memory representations (producing a general repetition benefit but not a spacing benefit), whereas responses to the unprimed meaning are primarily driven by consolidated, integrated memory representations (producing a spacing benefit). This CLS account distinction between context-based episodic representations and consolidated representations (Davis & Gaskell, 2009) has also been proposed as an explanation for recent findings on novel-word learning in German (Geukes, Gaskell, & Zwitserlood, 2015), with the time-course differing for these two types of learning (Weighall, Henderson, Barr, Cairney, & Gaskell, 2017).

This distinction between episodic representations and consolidated representations can also account for the different word-meaning priming patterns found with the word association test (Chapter 2). Word association showed significant word-meaning priming after one repetition. Since consolidation is unlikely after just one repetition, this indicates that the one repetition priming effect must be driven by episodic memory representations of the context provided in the prime phase. However, since word association also shows an additional priming boost from only spaced repetitions, this must reflect consolidation of the information learned about the primed meaning in the prime phase. This suggests that word association is also sensitive to both episodic and consolidated representations. It seems likely that the different emphases on these two sources of information in word association and semantic relatedness tests (caused by the presence or absence of contextual cues at test) leads to their different priming patterns for massed and spaced repetitions. Future research should aim to tease apart the individual contributions of underlying episodic and/or consolidated representations in word-meaning priming, ideally with a delay between prime and test that involves a period of sleep, since sleep would minimise any effect of episodic representations but maximise any effect of consolidated representations.

Whilst further research is required, the present results do indicate that although word-meaning priming in word association *can* reflect a direct modulation of the lexical-semantic network (from potentially consolidated representations after spaced learning), word-meaning priming can also operate outside of the lexicon through episodic representations. This is inconsistent with claims from earlier research on word-meaning priming. When altering the voice between prime and test phases (and therefore reducing useful episodic retrieval cues at test) did not reduce the priming effect from one repetition (Rodd et al., 2013, Experiment 2), it was concluded that word-meaning priming is not driven purely by episodic memory of the prime phase. With likely episodic effects in the present results, it now seems possible that the change in the sound of the voice simply had no impact on priming (perhaps because it is a lower level feature of language that might not affect the higher-level episodic representations of word-meanings) and that episodic representations can still affect priming.

The explanation of the different repetition priming effects seen with word association (Chapter 2) and semantic relatedness (Chapter 4) relies on the assumption that it was the change of task at test that was critical in explaining the different patterns of results. It is, however, important to rule out some alternative explanations from the literature for these differences in repetition effects. Since the experiments with these two tests involved the same prime phases (identical prime task, identical design and style of stimuli[28]), the difference must arise from either the filler task or

[28] All 60 ambiguous words in the semantic relatedness experiments were also used in the word association experiments, although an additional 28 ambiguous words were also used in the word association experiments.

the test task, both of which differed across experiments. There are two further theories relating to these factors that could explain the difference in spacing effects, although, as will be explained below, these are unlikely.

1. Potential effects of the filler task on priming

One possibility is that the filler task affected the way in which participants learned from the prime phase. The Towers of Hanoi task was chosen to replace the 'Shaun the Sheep' animation[29] because it was deemed to share some of the key characteristics that made it a suitable filler task. Arguably, the puzzle is a more active task than watching the animation. The passive animation might therefore have facilitated "wakeful rest": a period during which cognitive engagement is minimal, making memory consolidation more likely (Dewar, Alber, Butler, Cowan, & Della Sala, 2012). However, currently, no plausible mechanism exists to explain how wakeful rest during the animation filler would increase learning only from the spaced repetitions in the word association experiments. Any effects on learning from the filler tasks should cause an overall increase or decrease in priming across repetition conditions rather than alter the pattern of priming (i.e. change learning from one, three massed *and* three spaced repetitions overall, not selectively boost learning from spaced repetitions only).

2. Similarity between prime and test tasks in the semantic relatedness test experiments

All priming experiments in the present thesis used the same semantic relatedness prime task. Those in Chapter 4 used a similar semantic relatedness task for the test (in both prime and test, participants were presented with a stimulus and asked to decide if a second stimulus was related in meaning), whereas those in Chapter 2 used a word association task for the test. Transfer appropriate processing

[29] Due to Copyright rules practiced by Aardman Animations (creators of 'Shaun the Sheep' animations), it was not possible to distribute the 'Shaun the Sheep' animations online for use as the filler task for the semantic relatedness experiments.

(TAP; e.g. Morris, Bransford, & Franks, 1977) would predict that engaging in a semantic relatedness task at prime and again at test would boost the observed priming effect, compared to a word association task at test, because the semantic relatedness test reinstates the cognitive processes involved at prime, thereby helping the retrieval of learned cues. Indeed, repetition priming can be decreased, or even eliminated, when different tasks are used for prime and test (Franks, Bilbrey, Lien, & McNamara, 2000). However, again, this explanation can currently only explain an overall increase or decrease in priming across repetition conditions rather than a change in the pattern of priming itself.

Summary

In summary, it is unlikely the difference in filler tasks and the difference in the similarity of prime and test tasks across the experiments are the (main) causes of the different spacing effects. It is more likely that the difference in spacing effects is driven by the two types of test tapping into different episodic and consolidated representations learned at prime. Moreover, these findings indicate that more than one mechanism might drive word-meaning priming. It seems that there are at least the following two mechanisms or process: (1) an episodic-based process that is unstable and presumably shorter-lived, which only affects processing of the selected, appropriate meaning, and (2) a consolidation-based process that is stable and presumably longer-lived, which allows integration of information into the lexicon, and can interact with both the selected, appropriate meaning and unselected, inappropriate meaning. These mechanisms could mean that priming effects from spaced representations are more enduring than from massed repetitions. Massed repetitions are likely to have limited effects at longer delays, since they have not been consolidated, are not integrated into the lexicon, and are more resistant to forgetting (Brashers-Krug et al., 1996).

It seems likely that temporally separate encounters are more informative about the general distribution of the use of a particular word-meaning. Hence, retuning representations based on these encounters is likely to be an adaptive strategy that prevents them from being overly sensitive to temporally close encounters that might be more reflective of a single situation. It might be that comprehenders have a more stable set of representations that are learned slowly and cumulatively with experience that forms the basis of lexical-semantic knowledge, as well as a less stable type of information, which reflects the current and most recent of experiences, providing a rapidly-learned and more short-lived set of information that can guide interpretation in a particular situation without altering underlying representations. This is arguably an advantageous learning mechanism, whereby comprehenders update the multiple representations of an ambiguous word in relation to their everyday experience with language. This suggests that, throughout adulthood, all experience with language provides an opportunity from to which learn and update knowledge to continually maximise the efficiency of comprehension.

Conclusions

The present chapter has aimed to determine whether priming, in increasing the availability of the primed meaning, reduces the availability of the unprimed meaning. Using a word-meaning priming manipulation and semantic relatedness test, the experiments revealed that increasing the availability of the primed meaning *can* reduce availability of the unprimed meaning, but only when encounters are repeated and temporally spaced. Currently, the most likely explanation seems to be that, unlike one and three massed repetitions, the time between each of the three spaced encounters allows for the increased availability of the primed meaning to consolidate in the lexicon and interfere with competing meanings, and semantic relatedness is sensitive to these interference effects.

The observed data produced a complex picture indicating that different tests of word-meaning priming reveal different repetition priming patterns. It seems that word-meaning priming is not a simple, unitary process, but that it might be driven by multiple underlying mechanisms. Based on the present findings, it seems more likely that two types of information (episodic and consolidated) are learned during priming and that word association and semantic relatedness are differentially sensitive to these two types of information. It also seems likely that consolidation is an important process for retuning and updating lexical-semantic representations. It might be that the distinction between unconsolidated episodic memories and consolidated, integrated information about word-meanings provides the basis of a framework for further research into the effects of recent experience on ambiguous word interpretation.

Chapter 5: Concluding remarks

The aim of this thesis was to investigate how lexical-semantic representations can be retuned on the basis of recent experience. Word-meaning priming was used as a tool to examine how listeners learn from recent experience to inform their subsequent comprehension. Experiments 1, 2 and 3 (Chapter 2) investigated how multiple recent encounters with a particular word-meaning affected the subsequent interpretation of that word, measured with word association. Multiple subordinate repetitions provided an additional priming boost compared to one repetition when encounters were spaced, although not when massed. One repetition of the dominant meaning reduced, but did not eliminate, the effect of prior subordinate meaning priming. Then, using a newly-developed picture semantic relatedness test (Experiment 4, Chapter 3), Experiments 5 – 9 (Chapter 4) investigated whether word-meaning priming reflects increased availability of the primed meaning alone or the combination of increased availability of the primed meaning and decreased availability of the unprimed meaning. Together, these experiments showed that increasing the availability of the primed meaning *can* reduce availability of the unprimed meaning, but only when prime repetitions were repeated and temporally spaced.

The most likely account of these findings, which can accommodate the different patterns revealed from word association and semantic relatedness tests (Davis & Gaskell, 2009), suggests that information from three spaced repetitions can be consolidated and integrated into the lexicon, but that information from three massed repetitions is unlikely to boost consolidation over that of one repetition. This framework would suggest that word-meaning priming is driven by two underlying mechanisms: one based on episodic memory cues and one based on consolidated lexical-semantic representations. However, it important to emphasise that more work must be done to fully understand the learning mechanisms underlying these word-meaning priming effects. Whilst it is likely that the distributed connectionist model (Rodd et al., 2004; Rodd et al., 2013) can accommodate these word-meaning priming effects, we must not make assumptions or educated guesses about how the model could behave. Hence, running the model simulations is an important next-step in the field of semantic ambiguity resolution if we are to fully understand the mechanisms underlying learning from recent experience.

Another area that should be investigated with further research is how listeners learn from multiple repetitions with longer temporal spaces between repetitions. The temporal spacing between each spaced repetition in this thesis was relatively short, approximately 5 minutes, therefore it would be interesting to extend this and examine whether listeners learn more (or indeed less) from longer spacing of intervals. This leads onto another question about the bridge between learning from recent experience and learning over long time-frames. This thesis focuses on very recent experience in an experiment setting (up to approximately 30 minutes), but similar research has focused on learning from long-term experience in everyday life (Rodd et al., 2016, Experiments 3 & 4). It is not clear how a temporary boost in subordinate meaning availability from recent experience could translate into the potential of, with enough experience, the subordinate meaning becoming the dominant, preferred, interpretation of the word. Investigating this would give an insight into how lexical-semantic representations are developed in childhood, to how lexical-semantic representations are updated with longer-lasting effects in adulthood.

As well as an insight into the mechanisms of word-meaning priming, the experiments in this thesis have provided some valuable reminders about scientific practice. First, the findings show how we must be careful of making assumptions about the precise phenomenon being measured, without considering the processes that occur. For repetition effects on word-meaning priming in this thesis, if only the word association experiments had been conducted, it would have been concluded that the spacing, but not the number, of repetitions was crucial for a repetition benefit. If only the semantic relatedness experiments had been conducted, it would have been concluded that the number, but not the spacing, of repetitions was crucial for a repetition benefit. Although these different patterns can be accounted for by certain explanations (e.g. Davis & Gaskell, 2009), they highlight that we must not draw strong conclusions about a phenomenon being measured when it is only being measured with one test. This shows how the experimental measure can affect conclusions, as the two tests revealed two different findings. As a result, we must not assume that two different ways of measuring how recent experience affects comprehension are measuring the same representations, processes or learning mechanisms. Such measures should not be used interchangeably, but should be carefully selected based on whether the processes involved are likely to reveal meaningful findings about semantic ambiguity resolution. Clearly, there might need to be some compromise between the ideal task for measuring an effect and the most easily implemented task, but researchers should at least avoid generalising findings from one test of comprehension to comprehension as a whole.

A second lesson learned from the present experiments is about the merits of online experiments. In this thesis, Experiments 1, 2, 3 and 9 were all conducted in the laboratory, whereas Experiments 5, 6, 7, and 8 were all conducted online. It is reassuring that in this thesis the priming effects were similar regardless of the testing environment across three different elements. First, subordinate meaning priming effects were significantly different from the unprimed baseline across all subordinate priming experiments (Experiments 1, 2, 3, 5, 7, 8 & 9). Second, the 112ms reaction time difference between dominant and subordinate meaning pictures in the semantic relatedness test found in the laboratory (Experiment 4) was numerically similar online in all experiments that included both meanings at test (Experiments 5, 6, 7 & 8; ranging from a 99ms difference to a 119ms difference). Third, the pattern of repetition effects from online experiments using the semantic relatedness test (Experiment 8) replicated in the laboratory (Experiment 9), where three massed and three spaced repetitions both boosted subordinate word-meaning priming compared to the unprimed baseline.

The reliability of online data collection, compared to lab-based data collection, has been discussed in detail in psychology (see Woods, Velasco, Levitan, Wan, & Spence, 2015). Some have suggested that collecting data online is less reliable than in the lab and could involve unsuitable participants (e.g. Kraut, Patterson, Lundmark, Kiesler, Mukophadhyay, & Scherlis, 1998; McKenna & Bargh, 2000). Indeed, it is not straightforward to verify the suitability of participants, and their internet connection speeds might vary greatly. Arguably, however, there are several disadvantages to lab-based experiments. Typically, they rely on the university student population participating for course credit, meaning that the sample is biased towards this demographic (Gosling et al., 2004). Furthermore, recruitment is often slow (due to the limited pool of participants), more expensive (universities often require payments of above minimum wage, and require access to testing space, which is not always readily available). Moreover, particularly when multiple participants are tested at once, data is collected across multiple computers at different times, hence this is not that far from the situation with online experiments.

As for the advantages of online experiments, they seem to outweigh the potential disadvantages for experiments of this sort. Online experiment software such as Gorilla (Cauldron, www.gorilla.sc, 2017) have in-built internet connection speed tests, IP geolocation checks and internet connection speed should not affect response times within a trial. In Gorilla, the response time of the data is limited only by the refresh rate of the display and by the latency of the input device, both of which can also be the case with experiments conducted on a computer in the laboratory (refresh rate limits affect online and lab-based experiments alike, as can input device latency when multiple lab-based and unchecked computers are used for testing). Additionally, using online recruitment platforms such a Prolific (Prolific Academic Ltd., www.prolific.ac, 2016) allow access to a participant pool diverse in age, background education and so on, as well as a high number of these participants. This allows for quick and inexpensive recruitment, and therefore more data for the same cost as lab-based experiments, without a significant compromise in the quality of data (Casler et al., 2013; Gosling et al., 2004). Such recruitment websites provide a large amount of demographic information for each participant and the opportunity to market an experiment at only those who meet the criteria for participation. The online nature means that people can participate at their own convenience (increasing the chances of alert and motivated participants; Gosling & Mason, 2015). The increasing popularity of online experiments (Goodman et al., 2013; Litman et al., 2017) is therefore unsurprising and, as long as used cautiously, online data collection should be considered a valuable tool for psychologists.

The findings from this thesis are relevant to everyday life and contribute novel and important information to the field of semantic ambiguity resolution. First, experience-based changes to lexical-semantic representations are not solely based on the most recent encounter with a word-meaning, nor does the effect occur with the same magnitude across repeated encounters. Rather, word-meaning interpretation appears to reflect the accumulation of recent experiences with word-meanings. Second, massed repetition priming boosts the availability of the primed meaning, whereas spaced repetition priming both boosts the availability of the primed meaning and reduces the availability of the unprimed meaning. This demonstrates that the multiple lexical-semantic representations of an ambiguous word can, at least to some extent, affect one another. Learning from experience is not a straightforward process purely based on the number of exposures to a particular meaning, but is a more complex process affected by multiple factors, which continues throughout adulthood. Together, these findings suggest that listeners can learn from recent experiences in different ways and are not solely influenced by the most recent encounter. This seems to provide a balance among the influences of word usage patterns across a range of timescales, such that listeners can dynamically retune and update their lexical-semantic representations in response to on-going experience. It certainly seems that adults do not have a permanent, stable, preferred word-meaning interpretation, but can modulate their preferences in accordance with their life experience. This allows listeners to capitalise on experience in order to reflect the most likely meaning of words and maximise comprehension efficacy.

References

Aardman. (2010). Aardman Animations Ltd.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814-823. doi: https://dx.doi.org/10.1111/j.1467-9280.2006.01787.x

Adobe Systems. (2016). Photoshop Software.

Armstrong, B. C., Tokowicz, N., & Plaut, D. C. (2012). eDom: Norming software and relative meaning frequencies for 544 English homonyms. *Behavior research methods, 44*(4), 1015-1027. doi: https://doi.org/10.3758/s13428-012-0199-8

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi: https://dx.doi.org/10.1016/j.jml.2007.12.005

Bainbridge, J. V., Lewandowsky, S., & Kirsner, K. (1993). Context effects in repetition priming are sense effects. *Memory & Cognition, 21*(5), 619-626. doi: https://dx.doi.org/10.3758/BF03197194

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

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*(67), 1-48. doi: https://dx.doi.org/10.18637/jss.v067.i01

Becker, S., Behrmann, M., Moscovitch, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning Memory and Cognition, 23*(5), 1059-1082. doi: https://dx.doi.org/10.1037/0278-7393.23.5.1059

Binder, K. S., & Morris, R. K. (1995). Eye Movements and Lexical Ambiguity Resolution: Effects of Prior Encounter and Discourse Topic. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 21(5), 1186-1196. doi: https://doi.org/10.1037/0278-7393.21.5.1186

Binder, K. S., & Morris, R. K. (2011). An Eye-Movement Analysis of Ambiguity Resolution: Beyond Meaning Access. *Discourse Processes, 48*(5), 305-330. doi: https://dx.doi.org/10.1080/0163853X.2011.577391

Blumstein, S. E., Milberg, W., & Shrier, R. (1982). Semantic processing in aphasia: Evidence from an auditory lexical decision task. *Brain and Language, 17*(2), 301-315. doi: https://doi.org/10.1016/0093-934X(82)90023-2

Bock, M., & Klinger, E. (1986). Interaction of emotion and cognition in word recall. *Psychological Research, 48*(2), 99-106. doi: https://dx.doi.org/10.1007/BF00309323

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review, 6*(4), 635-640. doi: https://doi.org/10.3758/bf03212972

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature, 382*(6588), 252-255. doi: https://dx.doi.org/10.1038/382252a0

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., . . . McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences of the United States of America, 93*(15), 8016-8021. doi: https://dx.doi.org/10.1073/pnas.93.15.8016

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

Carroll, J. M., & Snowling, M. J. (2004). Language and phonological skills in children at high risk of reading difficulties. *Journal of Child Psychology and Psychiatry, 45*(3), 631-640. doi: https://doi.org/10.1111/j.1469-7610.2004.00252.x

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*(6), 2156-2160. doi: https://doi.org/10.1016/j.chb.2013.05.009

Cauldron. (www.gorilla.sc, 2017). Gorilla experiment software.

Chen, L., & Boland, J. E. (2008). Dominance and context effects on activation of alternative homophone meanings. *Memory and Cognition, 36*(7), 1306-1323. doi: https://dx.doi.org/10.3758/MC.36.7.1306

Chomsky, N. (2002). *On nature and language*: Cambridge University Press.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology, 69*(5), 817-828. doi: https://doi.org/10.1080%2F17470218.2015.1134603

Colbert-Getz, J., & Cook, A. E. (2013). Revisiting effects of contextual strength on the subordinate bias effect: Evidence from eye movements. *Memory and Cognition, 41*(8), 1172-1184. doi: https://dx.doi.org/10.3758/s13421-013-0328-3

Copland, D. A. (2006). Meaning selection and the subcortex: Evidence of reduced lexical ambiguity repetition effects following subcortical lesions. *Journal of Psycholinguistic Research, 35*(1), 51-66. doi: https://dx.doi.org/10.1007/s10936-005-9003-6

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*(3), 453-459. doi: https://doi.org/10.3758/bf03193787

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

Delaney, P. F., Spirgel, A. S., & Toppino, T. C. (2012). A deeper analysis of the spacing effect after “deep” encoding. *Memory & Cognition, 40*(7), 1003-1015. doi: https://dx.doi.org/10.3758/s13421-012-0207-3

Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review, 93*(3), 283-321. doi: https://dx.doi.org/10.1037/0033-295X.93.3.283

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief Wakeful Resting Boosts New Memories Over the Long Term. *Psychological Science, 23*(9), 955-960. doi: https://doi.org/10.1177/0956797612441220

Dholakia, A., Meade, G., & Coch, D. (2016). The N400 elicited by homonyms in puns: Two primes are not better than one. *Psychophysiology, 53*(12), 1799-1810. doi: https://dx.doi.org/10.1111/psyp.12762

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. doi: https://dx.doi.org/10.1016/0749-596X(88)90066-6

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science (0956-7976), 18*(1), 35-39. doi: https://doi.org/10.1111/j.1467-9280.2007.01845.x

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808-816. doi: https://doi.org/10.1080/17470218.2017.1310261

Durso, F. T., & Johnson, M. K. (1979). Facilitation in naming and categorizing repeated pictures and words. *Journal of Experimental Psychology: Human Learning and Memory, 5*(5), 449-459. doi: https://dx.doi.org/10.1037/0278-7393.5.5.449

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R(2) Statistic for Fixed Effects in the Linear Mixed Model. *Statistics in medicine, 27*(29), 6137-6157. doi: https://doi.org/10.1002/sim.3429

Fang, X., & Perfetti, C. A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia, 99*, 270-278. doi: https://dx.doi.org/10.1016/j.neuropsychologia.2017.03.015

Fine, A. B., & Jaeger, T. F. (2013). Evidence for Implicit Learning in Syntactic Comprehension. *Cognitive Science, 37*(3), 578-591. doi: https://dx.doi.org/10.1111/cogs.12022

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition, 2*(2), 337-339. doi: https://dx.doi.org/10.3758/BF03209005

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680-698. doi: https://dx.doi.org/10.1037/0278-7393.10.4.680

Foss, D. J. (1970). Some effects of ambiguity upon sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 9*(6), 699-706. doi: https://dx.doi.org/10.1016/S0022-5371(70)80035-4

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP). *Memory & Cognition, 28*(7), 1140-1151. doi: https://doi.org/10.3758/bf03211815

Gale, W. A., Church, K. W., & Yarowsky, D. (1992). *One sense per discourse*. Paper presented at the Proceedings of the workshop on Speech and Natural Language, Harriman, New York.

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

Geis, M. F., & Winograd, E. (1974). Norms of semantic encoding variability for fifty homographs. *Bulletin of the Psychonomic Society, 3*(6), 429-431. doi: https://dx.doi.org/10.3758/BF03333516

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating Differences in General Comprehension Skill. *Journal of experimental psychology. Learning, memory, and cognition, 16*(3), 430-445. doi: https://psycnet.apa.org/doi/10.1037/0278-7393.16.3.430

Geukes, S., Gaskell, M. G., & Zwitserlood, P. (2015). Stroop effects from newly learned color words: effects of memory consolidation and episodic context. *Frontiers in Psychology, 6*(278). doi: https://doi.org/10.3389/fpsyg.2015.00278

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and Readers Generalize Their Experience With Word Meanings Across Modalities. *J Exp Psychol Learn Mem Cogn*. doi: https://doi.org/10.1037/xlm0000532

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15*(1), 1-16. doi: https://dx.doi.org/10.1016/S0022-5371(76)90002-5

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making, 26*(3), 213-224. doi: https://doi.org/10.1002/bdm.1753

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual review of psychology, 66*, 877-902. doi: https://doi.org/10.1146/annurev-psych-010814-015321

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American psychologist, 59*(2), 93-104.

Gotts, S. J., Chow, C. C., & Martin, A. (2012). Repetition Priming and Repetition Suppression: A Case for Enhanced Efficiency Through Neural Synchronization. *Cognitive neuroscience, 3*(3-4), 227-237. doi: https://dx.doi.org/10.1080/17588928.2012.670617

Greene, R. L. (1989). Spacing Effects in Memory: Evidence for a Two-Process Account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(3), 371-377. doi: https://dx.doi.org/10.1037/0278-7393.15.3.371

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 12*(2), 229-238. doi: https://dx.doi.org/10.1016/S0022-5371(73)80013-1

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical journal, 50*(3), 346-363. doi: https://doi.org/10.1002/bimj.200810425

Huettig, F., Gaskell, M. G., & Quinlan, P. T. (2004). *How speech processing affects our attention to visually similar objects: Shape competitor effects and the visual world paradigm.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia, 45*(5), 1019-1028. doi: https://doi.org/10.1016/j.neuropsychologia.2006.09.004

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434-446. doi: https://dx.doi.org/10.1016/j.jml.2007.11.007

JASP Team. (https://jasp-stats.org/, 2017). Statistical software.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13*(2), 278-305. doi: https://doi.org/10.1016/0010-0285(81)90011-6

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1250. doi: https://dx.doi.org/10.1037/a0023436

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133*(3), 450-467. doi: https://dx.doi.org/10.1037/0096-3445.133.3.450

Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the timecourse of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 601. doi: https://doi.org/10.1037/0096-1523.14.4.601

Khanna, M. M., & Cortese, M. J. (2011). Age of acquisition estimates for 1,208 ambiguous and polysemous words. *Behavior research methods, 43*(1), 89-96. doi: https://doi.org/10.3758/s13428-010-0027-y

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20*(1), 1-24. doi: https://doi.org/10.1016/j.jneuroling.2006.02.001

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 554. doi: https://dx.doi.org/10.1037/0278-7393.2.5.554

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American psychologist, 53*(9), 1017-1031.

Kurdziel, L. B. F., Mantua, J., & Spencer, R. M. C. (2017). Novel word learning in older adults: A role for sleep? *Brain and Language, 167*, 106-113. doi: https://doi.org/10.1016/j.bandl.2016.05.010

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review, 76*(1), 82. doi: https://dx.doi.org/10.1037/h0026746

Leinenger, M., & Rayner, K. (2013). Eye movements while reading biased homographs: Effects of prior encounter and biasing context on reducing the subordinate bias effect. *Journal of Cognitive Psychology, 25*(6), 665-681. doi: https://dx.doi.org/10.1080/20445911.2013.806513

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

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 608. doi: https://dx.doi.org/10.1037/a0029243

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods, 49*(2), 433-442. doi: https://doi.org/10.3758/s13428-016-0727-z

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22*(1), 1-35. doi: https://dx.doi.org/10.1016/0010-0285(90)90002-L

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology, 28*(6), 684-706. doi: https://dx.doi.org/10.1080/20445911.2016.1181637

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*(6), 828-835. doi: http://doi.org/10.1016/S0022-5371(69)80050-2

Martin, C., Vu, H., Kellas, G., & Metcalf, K. (1999). Strength of Discourse Context as a Determinant of the Subordinate Bias Effect. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 52*(4), 813-839. doi: https://doi.org/10.1080/713755861

Masson, M. E. J., & Freedman, L. (1990). Fluent Identification of Repeated Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 355-373. doi: https://psycnet.apa.org/doi/10.1037/0278-7393.16.3.355

MATLAB. (R2013b, 2013).

MATLAB. (R2015a, 2015).

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. doi: https://doi.org/10.1016/j.jml.2017.01.001

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419. doi: https://doi.org/10.1037/0033-295X.102.3.419

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375. doi: https://psycnet.apa.org/doi/10.1037/0033-295X.88.5.375

McKenna, K. Y. A., & Bargh, J. A. (2000). Plan 9 From Cyberspace: The Implications of the Internet for Personality and Social Psychology. *Personality and Social Psychology Review, 4*(1), 57-75. doi: https://doi.org/10.1207/S15327957PSPR0401_6

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 596-606. doi: https://dx.doi.org/10.1016/S0022-5371(70)80107-4

Mensink, G. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology, 33*(2), 172-186. doi: http://dx.doi.org/10.1016/0022-2496(89)90029-1

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234. doi: https://dx.doi.org/10.1037/h0031564

Microsoft Clipart (Producer). (2016).

Monsell, S., & Hirsh, K. W. (1998). Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1495-1520. doi: https://psycnet.apa.org/doi/10.1037/0278-7393.24.6.1495

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal*

Behavior, 16(5), 519-533. doi: https://doi.org/10.1016/S0022-5371(77)80016-9

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16*(2), 151-171. doi: https://dx.doi.org/10.1016/S0022-5371(77)80044-3

Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior research methods, 37*(3), 398-416. doi: https://doi.org/10.3758/bf03192709

Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior research methods, 44*(4), 934-945. doi: https://doi.org/10.3758/s13428-011-0176-7

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology, 90*(2), 142-171. doi: https://doi.org/10.1016/j.jecp.2004.11.003

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204-238. doi: https://dx.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. doi: https://doi.org/10.3758/bf03196957

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227-252. doi: https://dx.doi.org/10.1016/j.cognition.2009.09.007

Paivio, A. (1974). Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 497-511. doi: https://dx.doi.org/10.1016/S0022-5371(74)80002-2

Parks, R., Ray, J., & Bland, S. (1998). Wordsmyth English Dictionary-Thesaurus. [On-line]. Available: https://www.wordsmyth.net/ [1999, February 1], University of Chicago

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive*

Science, 29(4), 559-586. doi: https://dx.doi.org/10.1207/s15516709cog0000_14

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*(2), 101. doi: https://doi.org/10.1037/1076-898X.14.2.101

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*(3), 280-291. doi: https://doi.org/10.1016/j.cognition.2011.10.004

Poort, E. D., Warren, J. E., & Rodd, J. M. (2016). Recent experience with cognates and interlingual homographs in one language affects subsequent processing in another language. *Bilingualism: Language and Cognition, 19*(1), 206-212. doi: https://dx.doi.org/10.1017/S1366728915000395

Prolific Academic Ltd. (www.prolific.ac, 2016). Prolific recruitment platform.

Qualtrics Inc. (www.qualtrics.com, 2016). Qualtrics Inc. survey software.

R Core Team. (2016). R: A language and environment for statistical

computing. Vienna, Austria: R Foundation for Statistical Computing.

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: application of the SAM model. *Cognitive Science, 27*(3), 431-452. doi: https://dx.doi.org/10.1016/S0364-0213(03)00007-7

Rayner, K., Binder, K. S., & Duffy, S. A. (1999). Contextual Strength and the Subordinate Bias Effect: Comment on Martin, Vu, Kellas, and Metcalf. *The Quarterly Journal of Experimental Psychology Section A, 52*(4), 841-852. doi: https://doi.org/10.1080/713755868

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

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of Prior Encounter and Global Discourse Bias on the Processing of Lexically Ambiguous Words: Evidence From Eye Fixations. *Journal of Memory and Language, 33*(4), 527-544. doi: https://dx.doi.org/10.1006/jmla.1994.1025

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. doi: https://dx.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. doi: https://dx.doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89-104. doi: https://dx.doi.org/10.1016/j.cogsci.2003.08.002

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

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

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 487-494. doi: https://doi.org/10.1016/S0022-5371(70)80091-3

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. doi: https://doi.org/10.1016/0010-0285(82)90017-2

Shadmehr, R., & Brashers-Krug, T. (1997). Functional Stages in the Formation of Human Long-Term Motor Memory. *The Journal of Neuroscience, 17*(1), 409-419. doi: https://doi.org/10.1038/382252a0

Shea, C. H., Lai, Q., Black, C., & Park, J. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19*(5), 737-760. doi: https://dx.doi.org/10.1016/S0167-9457(00)00021-X

Shook, A., & Marian, V. (2012). Bimodal bilinguals co-activate both languages during spoken comprehension. *Cognition, 124*(3), 314-324. doi: https://doi.org/10.1016/j.cognition.2012.05.014

Simpson, G. B., & Kang, H. (1994). Inhibitory processes in the recognition of homograph meanings. In D. D. T. H. C. hibitory processes in the recognition of homograph meanings (Ed.), *Inhibitory processes in attention, memory, and language* (pp. 359-381). San Diego, CA, US: Academic Press.

Simpson, G. B., & Kellas, G. (1989). Dynamic Contextual Processes and Lexical Access. In D. S. Gorfein (Ed.), *Resolving Semantic Ambiguity* (pp. 40-56). New York, NY: Springer New York.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174-215. doi: https://psycnet.apa.org/doi/10.1037/0278-7393.6.2.174

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua, 157*, 124-152. doi: https://doi.org/10.1016/j.lingua.2014.12.004

Stringaris, A. K., Medford, N., Giora, R., Giampietro, V. C., Brammer, M. J., & David, A. S. (2006). How metaphors influence semantic relatedness judgments: The role of the right frontal cortex. *NeuroImage, 33*(2), 784-793. doi: https://doi.org/10.1016/j.neuroimage.2006.06.057

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. doi: https://doi.org/10.1016/S0022-5371(79)90355-4

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27*(3), 324-340. doi: https://doi.org/10.1016/0749-596X(88)90058-7

Tabossi, P., & Zardon, F. (1993). Processing Ambiguous Words in Context. *Journal of Memory and Language, 32*(3), 359-372. doi: https://doi.org/10.1006/jmla.1993.1019

Tamminen, J., & Gaskell, M. G. (2013a). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology, 66*(5), 1001-1025. doi: https://dx.doi.org/10.1080/17470218.2012.724694

Tamminen, J., & Gaskell, M. G. (2013b). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology, 66*(5), 1001-1025. doi: https://doi.org/10.1080/17470218.2012.724694

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience, 30*(43), 14356-14360. doi: https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 427-440. doi: https://doi.org/10.1016/S0022-5371(79)90237-8

Thios, S. J. (1972). Memory for words in repeated sentences. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 789-793. doi: https://dx.doi.org/10.1016/S0022-5371(72)80014-8

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

Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 573-580. doi: https://dx.doi.org/10.1016/S0022-5371(70)80104-9

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190. doi: https://dx.doi.org/10.1080/17470218.2013.850521

Vitello, S., & Rodd, J. M. (2015). Resolving Semantic Ambiguities in Sentences: Cognitive Processes and Brain Mechanisms. *Language and Linguistics Compass, 9*(10), 391-405. doi: https://dx.doi.org/10.1111/lnc3.12160

Vu, H., & Kellas, G. (1999). Contextual Strength Modulates the Subordinate Bias Effect: Reply to Rayner, Binder, and Duffy. *The Quarterly Journal of Experimental Psychology Section A, 52*(4), 853-855. doi: https://doi.org/10.1080/713755863

Wacom. (2016). Bamboo Stylus.

Warren, J. E., Vitello, S., Devlin, J., & Rodd, J. M. (in preparation).

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time‐course of novel word learning and lexical competition in adults and children. *Brain and Language, 167*, 13-27. doi: https://doi.org/10.1016/j.bandl.2016.07.010

Welford, A. T. (1952). The 'Psychological Refractory Period' and the timing of high-speed performance - A review and a theory. *British Journal of Psychology. General Section, 43*(1), 2-19. doi: https://dx.doi.org/10.1111/j.2044-8295.1952.tb00322.x

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of Spoken Word Production by Priming a Semantic Competitor. *Journal of Memory and Language, 33*(3), 332-356. doi: https://doi.org/10.1006/jmla.1994.1016

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. doi: https://dx.doi.org/10.1080/17470218.2016.1251470

Wittgenstein, L. (Ed.). (1953). *Philosophical Investigations*: Oxford, UK: Blackwell.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ, 3*, e1058. doi: https://doi.org/10.7717/peerj.1058

Zipf, G. K. (1949). Human Behavior and the Principle of Least-Effort: Cambridge, MA: Addison Wesley.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language Comprehenders Mentally Represent the Shapes of Objects. *Psychological Science (0956-7976), 13*(2), 168. doi: https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*(4), 954-958. doi: https://doi.org/10.3758/bf03196557

Appendices

Appendix A

Stimuli from Experiment 1

Table I. List of the 60 experimental ambiguous words used in Experiment 1.

Ambiguous Words			
Appendix	Coach	Key	Pupil
Arms	Cold	Lace	Race
Ball	Craft	Landing	Racket
Band	Crane	Letter	Record
Bar	Cricket	Mark	Ring
Bark	Deck	Mould	Spade
Bolt	Drill	Mouse	Spring
Bonnet	Figure	Note	Staff
Break	Gear	Nut	Step
Cabinet	Gum	Organ	Stitch
Cap	Habit	Palm	Straw
Case	Interest	Panel	Strike
Change	Iron	Pipe	Temple
Cheek	Issue	Pride	Trailer
Chest	Jam	Punch	Watch

Appendix B

Stimuli from Experiments 2 & 3

Table II. List of the 88 ambiguous words used in Experiments 2 and 3.

Ambiguous Words			
Appendix	China	Joint	Racket
Ball	Coach	Key	Record
Band	Cold	Knight	Ring
Bar	Craft	Lace	See
Bark	Cricket	Landing	Sign
Bat	Cross	Letter	Sink
Bed	Cup	Mark	Skip
Blew	Deck	Match	Son
Bonnet	Drawer	Mould	Spade
Bow	Fan	Mouse	Speaker
Bowl	Fence	Nail	Spring
Box	Figure	Note	Staff
Break	Flour	Organ	Step
Bulb	Gear	Pair	Stitch
Button	Glasses	Palm	Straw
Cabinet	Gum	Panel	Strike
Calf	Hand	Park	Temple
Cap	Hare	Pen	Toast
Card	Interest	Pipe	Trailer
Case	Iron	Plug	Trunk
Change	Issue	Punch	Watch
Chest	Jam	Pupil	Wave

Appendix C

Additional results from Experiment 4 – correlations and regressions

Below are the correlation and multiple regression (including R^2) alternatives to the mixed effects models on comparing different measures of word-meaning dominance (from Analysis Stage 2 in the Results section of Experiment 4). For the regression analyses, all predictors were included in the same simultaneous regression to measure their contributions while taking into account the other predictors.

Correlations - dominant meaning

Table III, below, which only includes the dominant meaning of each word, shows the correlations between all factors. One aim was to see whether the three predictors (word association scores, eDom scores, and picture quality) were correlated with one another. Word association and eDom dominance scores were significantly positively correlated indicating similarity between these measures. Picture quality was not significantly correlated with word association or eDom dominance measures. Another aim was to see which predictors were correlated with semantic relatedness performance. As expected, word association was significantly negatively correlated with both semantic relatedness measures (an increase in dominance should be associated with faster reaction times and more accurate responses). Faster reaction times and lower error rates were associated with the more dominant of the dominant meanings. Picture quality was also significantly negatively correlated with both semantic relatedness measures. However, eDom showed a slightly different pattern: a significant negative correlation with semantic relatedness RTs but not with errors.

Table III. Correlation matrix for Experiment 4 data for the dominant word meanings, including both dominance measures (word association (WA) and eDom), the picture quality for the dominant meaning and performance measures from the semantic relatedness task (RT and Error). The mean dominance score is provided for each measure, with standard deviations in parentheses. The mean values for each measure are in the following units: proportion for word association and eDom; absolute rating for picture quality (rating scale 1-5); milliseconds for semantic relatedness RTs; percentage for semantic relatedness error. Significance level emphasised with asterisks (* <.05, ** < .01, ***<.001).

	Mean (SD)		**WA**	**eDom**	**PicQuality**	**SemRelRT**	**SemRelError**
WA	0.728 (0.230)	Pearson's r	-	0.650***	0.231	-0.487***	-0.269*
		p value	-	< .001	.064	< .001	.030
eDom	0.610 (0.168)	Pearson's r		-	0.180	-0.346**	-0.223
		p value		-	.152	.005	.074
PicQuality	4.903 (0.120)	Pearson's r			-	-0.305*	-0.332**
		p value			-	.014	.007
SemRel RT	595ms (69ms)	Pearson's r				-	-
		p value				-	-
SemRel Error	2.9% (6%)	Pearson's r					-
		p value					-

R^2 from multiple regression - dominant meaning RTs and errors

Table IV. Multiple regression analysis results for dominant meaning reaction times for Experiment 4.

Predictors	**SE**	**Standardised *β***	***p***
Word Association	0.043	-0.415	.006
eDom	0.059	-0.040	.782
Picture Quality	0.065	-0.202	.077
F (3,61) = 7.809, p < .001			
R^2 = .277			

Table V. Multiple regression analysis results for dominant meaning error rates for Experiment 4.

Predictors	**SE**	**Standardised *β***	***p***
Word Association	0.041	-0.158	.318
eDom	0.056	-0.069	.657
Picture Quality	0.061	-0.283	.023
F (3,61) = 3.646, p = .017			
R^2 = .152			

Correlations - subordinate meaning

The pattern of significance in Table VI, below, which only concerns the subordinate meaning of each word is entirely consistent with the pattern for the dominant meaning. One aim was to see whether the three predictors (word association scores, eDom scores, and picture quality) were correlated with one another. Word association and eDom dominance scores were significantly positively correlated indicating similarity between these measures. Picture quality was not significantly correlated with word association or eDom. Another aim was to see which predictors were correlated with semantic relatedness performance. As before, word association was significantly negatively correlated with both semantic relatedness measures, where slower and less accurate responses were associated with the more subordinate of the subordinate meanings. However, once again, eDom was only significantly negatively correlated with semantic relatedness RTs, not errors. Picture quality was significantly negatively correlated with both semantic relatedness measures.

Table VI. Correlation matrix for Experiment 4 data for the subordinate word meanings, including both dominance measures (word association (WA) and eDom), the picture quality for the subordinate meaning and performance measures from the semantic relatedness task (RT and Error). The mean dominance score is provided for each measure, with standard deviations in parentheses. The mean values for each measure are in the following units: proportion for word association and eDom; absolute rating for picture quality (rating scale 1-5); milliseconds for semantic relatedness RTs; percentage for semantic relatedness error. Significance level emphasised with asterisks (** < .01, ***<.001).

	Mean (SD)		**WA**	**eDom**	**PicQuality**	**SemRelRT**	**SemRelError**
WA	0.216 (0.202)	Pearson's r	-	0.617***	0.062	-0.533***	-0.365**
		p value	-	< .001	0.625	< .001	.003
eDom	0.372 (0.164)	Pearson's r		-	0.050	-0.338**	-0.223
		p value		-	0.695	.006	.074
PicQuality	4.855 (0.194)	Pearson's r			-	-0.429***	-0.361**
		p value			-	< .001	.003
SemRel RT	707ms (101ms)	Pearson's r				-	-
		p value				-	-
SemRel Error	10.5% (11.5%)	Pearson's r					-
		p value					-

R^2 from multiple regression - subordinate meaning RTs and errors

Table VII. Multiple regression analysis results for subordinate meaning reaction times for Experiment 4.

Predictors	**SE**	**Standardised *β***	***p***
Word Association	0.061	-0.503	< .001
eDom	0.075	-0.008	.949
Picture Quality	0.050	-0.397	< .001
F (3,61) = 16.050, $p < .001$			
R^2 = .441			

Table VIII. Multiple regression analysis results for subordinate meaning error rates for Experiment 4.

Predictors	**SE**	**Standardised *β***	***p***
Word Association	0.080	-0.350	.016
eDom	0.099	0.009	.948
Picture Quality	0.066	-0.340	.003
F (3,61) = 6.708, p = < .001			
R^2 = .248			