Published in final edited form as:

Cogn Psychol. 2015 June ; 79: 68–101. doi:10.1016/j.cogpsych.2015.04.001.

Single-Word Predictions of Upcoming Language During Comprehension: Evidence from the Cumulative Semantic Interference Task

Daniel Kleinman¹, Elin Runnqvist², and Victor S. Ferreira¹

Daniel Kleinman: kleinman@gmail.com; Elin Runnqvist: elin_runnqvist@yahoo.es; Victor S. Ferreira: vferreira@ucsd.edu

¹Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, USA

²Laboratoire de Psychologie Cognitive, UMR 7290, CNRS & Aix-Marseille Université, 3 place Victor Hugo, 13331 Marseille CEDEX 3, France

Abstract

Comprehenders predict upcoming speech and text on the basis of linguistic input. How many predictions do comprehenders make for an upcoming word? If a listener strongly expects to hear the word “sock”, is the word “shirt” partially expected as well, is it actively inhibited, or is it ignored? The present research addressed these questions by measuring the “downstream” effects of prediction on the processing of subsequently presented stimuli using the cumulative semantic interference paradigm. In three experiments, subjects named pictures (*sock*) that were presented either in isolation or after strongly constraining sentence frames (“After doing his laundry, Mark always seemed to be missing one...”). Naming *sock* slowed the subsequent naming of the picture *shirt* – the standard cumulative semantic interference effect. However, although picture naming was much faster after sentence frames, the interference effect was not modulated by the context (bare vs. sentence) in which either picture was presented. According to the only model of cumulative semantic interference that can account for such a pattern of data, this indicates that comprehenders pre-activated and maintained the pre-activation of best sentence completions (*sock*) but did not maintain the pre-activation of less likely completions (*shirt*). Thus, comprehenders predicted only the most probable completion for each sentence.

Keywords

Prediction; Sentence comprehension; Speech production; Word retrieval; Semantic interference

© 2015 Published by Elsevier Inc.

Corresponding author: Daniel Kleinman, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, USA, kleinman@gmail.com, Voice: +1-408-596-3757, Fax: +1-858-534-7190.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Language comprehenders and horror movie victims have something in common: Both would benefit from knowing what's going to happen next. The ability to anticipate upcoming events on the basis of current information is useful in a wide variety of situations, as it helps drivers brake for pedestrians who intend to cross the street, allows batters to hit baseballs thrown at high speeds, and increases the likelihood of successfully evading a hockey mask-wearing pursuer.

One domain in which anticipation is especially helpful is language processing. As a sentence unfolds over time, listeners must rapidly recognize each word and integrate it into the preceding context to recover the speaker's intended meaning. The difficulty of this process could be reduced if listeners were capable of generating expectations about words prior to hearing them. For example, consider this sentence fragment: "After doing his laundry, Mark always seemed to be missing one..." It is easy to see that the next word is likely to be an article of clothing; furthermore, it is the kind of article that is often misplaced. To the extent that listeners can make efficient use of this real-world knowledge, they might be able to anticipate (correctly) that the next word will be "sock", making it easier to recognize the word once they actually hear it.

Existing research suggests that listeners and readers do in fact engage in such anticipatory behavior, generating predictions of upcoming speech and text that can vary in scope from semantic features (Szewczyk & Schriefers, 2013) to the level of individual words (DeLong, Urbach, & Kutas, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004). These predictions may be either conscious and controlled in nature (as suggested above) or they may be more passive, with contextually relevant words rising and falling in accessibility as a sentence or discourse unfolds (Myers & O'Brien, 1998; Van Berkum, 2009). Either way, predicting a word affects processing by increasing the activation (i.e., accessibility) of its representation in the mental lexicon, often called its lemma. This increase facilitates the subsequent access of the predicted word when the prediction is correct.¹

However, prediction may also affect the activation of words other than the most likely candidate, including those that are semantically related to it (e.g., *shirt* for the "sock" sentence above). These related words might receive activation directly from the sentence, as they share overlapping conceptual representations with the most strongly predicted word and thus are likely to appear in the same kinds of contexts, or indirectly, via spreading activation. The present research focuses on how prediction affects these semantically related words on two different timescales. Specifically, how does the "sock" sentence affect the activation of the *shirt* lemma (a) as the sentence unfolds (the immediate effects of prediction), and (b) when similar contexts are encountered in the future (the "downstream" effects of prediction)?

¹Alternative accounts argue that some facilitation from context may be attributable to the ease with which a word can be integrated into that context (e.g., Hagoort, Baggio, & Willems, 2009). The current data will not be able to adjudicate this long-standing debate, but results will be framed in terms of pre-activation.

We will present a novel way of examining these effects by combining sentence comprehension with a paradigm from language production research. As a first step, we focus here on a situation that encourages comprehenders to make specific predictions about the identity of an upcoming word by using strongly constraining sentences. It is fair to note at this point that confining our investigation to a task context that encourages very specific predictions may limit the extent to which our conclusions can be generalized. We address this point in the General Discussion in light of the observed results, and speculate on how future research can use this paradigm to investigate the scope of prediction more generally, including discriminating between situations in which comprehenders do and do not make single-word predictions.

1.1. Effects of sentential constraint on non-target words

There are three possible ways that the “sock” sentence might affect the activation of non-target lemmas. As it seems likely that non-target lemmas become (either directly or indirectly) activated during sentence comprehension (Collins & Loftus, 1975; Dell, 1986; Roelofs, 1992), the first possibility is that this extra pre-activation lingers until the target is ultimately selected. In other words, the language system may recognize that multiple words are contextually appropriate (in addition to “sock”) and thereby predict multiple words for the same word slot, such as “shirt”, “jacket”, and “pants”. Although this would increase the likelihood of making a correct prediction, it also guarantees that (at a minimum) all but one of those predictions must be incorrect. Depending on the costs associated with incorrect predictions, making multiple predictions might be inefficient.

The second possibility is that the language system ultimately inhibits the lemmas of pre-activated non-target words. In other words, the “sock” sentence might initially boost the activation of both the *sock* and *shirt* lemmas, but as the evidence in favor of *sock* continues to accumulate, the language system may react by reducing the activation of *shirt* (as well as the activation of different-category lemmas) below its baseline level prior to engaging in lemma selection. The costs and benefits of this possibility are essentially inverted relative to those of the previous one. If a listener is highly certain of what an upcoming word will be, reducing the activation of competitors will, upon presentation of the target, facilitate processing of the target word even more. However, if the prediction is incorrect and the correct word is instead a close competitor of the expected word (“shirt”), the inhibition will need to be lifted before “shirt” can be accessed and integrated into the preceding context.

The third possibility is that the “sock” sentence has no net effect on the activation of non-target words. This account is similar to the previous one, except that when evidence accumulates suggesting that *shirt* is an unlikely target, its activation returns to its baseline level rather than being inhibited. In a sense, this compromise represents a hedged bet: The comprehender’s language system is sufficiently confident in its prediction of the target so as not to maintain the pre-activation of other potential completions, but not so confident as to double down on that prediction by actively suppressing them.

To address questions about the scope of prediction and how it affects the activation of lexical representations, researchers have studied the processing of words across contexts in which their predictability varies. The predictability of a particular word in a given context is

typically determined via a standard cloze task (Taylor, 1953), in which subjects are presented with a series of sentence fragments (such as the “sock” fragment above), each followed by a blank, and asked to fill in the blank with the word that comes to mind first or that best completes the sentence. Subsequently, responses are scored according to their probability and sentences are scored according to their response distribution. For each sentence, the *cloze probability* of a response is the probability that it will be produced as a completion of that sentence, with high-cloze responses (e.g., “sock”) produced more often than low-cloze responses (“shirt”, “quarter”). The *constraint* of a sentence is linked to the probability of its highest-cloze response. For example, a strongly constraining sentence might be completed with the same word by 90% of subjects, whereas for a weakly constraining sentence (“Today I saw a ____”), it might be only 20%. These measurements are generally collected off-line from one group of subjects and then used in a task with a second group of subjects to determine how sentential constraint and cloze affect the processing of individual words.

Prior research along these lines has largely relied on comprehension tasks, with the ease of processing a particular word assessed either via reaction times, reading times, or electrophysiological measures. Schwanenflugel and colleagues’ investigation into the scope of sentential constraint is representative of these behavioral studies (Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985). They used a lexical decision (word/non-word judgment) task in which critical words were presented as high- or low-cloze completions of strongly or weakly constraining sentences, or after a neutral condition (either a string of Xs or a sentence that could be completed by any word). Relative to the neutral conditions, responses to high-cloze completions were facilitated by strongly constraining sentences and, to a lesser degree, by weakly constraining sentences as well. In contrast, responses to low-cloze completions were facilitated only by weakly constraining sentences, and even then only when they were semantically related to the high-cloze completion (e.g., “Hank reached into his pocket to get the” followed by the low-cloze completion “coin”, related to the high-cloze completion “money”). In fact, responses to low-cloze completions were sometimes slowed by strongly constraining sentences. Other studies that used lexical decision and word naming tasks supported the notion that strongly constraining sentences have a narrow scope of facilitation, finding that they either did not generally affect the processing of low-cloze completions under normal conditions (Stanovich & West, 1979, 1981) or that they slowed such processing (Fischler & Bloom, 1979, 1985; Forster, 1981; Schubert & Eimas, 1977).

Eye-tracking studies that have investigated reading in more natural contexts have also demonstrated effects of contextual constraint on word processing. In one such study (Ehrlich & Rayner, 1981), participants read strongly and weakly constraining paragraphs containing a target that was either a predicted word (e.g., *horse*) or an orthographic neighbor (i.e., a misspelling) of the predicted word (*house*). When the target word was predicted, participants fixated it less often and for less time in strongly constraining paragraphs than in weakly constraining paragraphs. Interestingly, when the target word was misspelled – and thus not predicted – the average fixation duration was the same across strongly and weakly constraining paragraphs (though misspellings were fixated less often in (mis-)predictive

contexts). A similar contrast was reported by Ashby, Rayner, and Clifton (2005), who presented participants with (among other kinds of stimuli) sentences that strongly predicted an unrepresented word (“The gambler visited the *island* as part of his vacation.”) and sentences that were contextually neutral (“After bowling, the *local* bar is always crowded.”). The extent to which eye movement measures differed between these contexts depended on several factors including reading skill and word frequency, but – most relevant for present purposes – differences were not always apparent: When reading high-frequency words, there was essentially no difference in reading times between mis-predictive and neutral contexts (*island* vs. *local*), either for skilled readers or for average readers (summing across gaze durations and spillover effects). These results, which demonstrate that strongly constraining sentences do not always affect the processing of unlikely words, are consistent with those observed in the lexical decision and word naming tasks.

In addition to reaction time and eye movement measures, a wealth of electrophysiological studies, in which subjects typically read or listen to sentences passively while EEG is recorded, have investigated the effects of sentential constraint and expectedness on word processing through the examination of ERP components. Perhaps the most widely-researched such component is the N400 (Kutas & Hillyard, 1980, 1984), which is generally taken to index the goodness of fit between a word and its context. Kutas and Hillyard (1984) found that when the sentence-final word was a high-cloze completion, the amplitude of the N400 component was modulated by the degree of constraint, with smaller amplitudes for more strongly constraining sentences. In contrast, the N400 elicited by a low-cloze completion was unaffected by constraint except if it was semantically related to the best completion (e.g., “He liked lemon and sugar in his coffee.”), in which case it elicited a reduced N400 relative to semantically unrelated low-cloze completions (“Don’t touch the wet dog.”). Similarly, words from different semantic categories than the highest-cloze completion (e.g., “tulips”, when “palms” is expected) elicit a robust N400 regardless of constraint, whereas words from the same semantic category (“pines”) elicit a smaller N400, especially after a strongly constraining sentence (Federmeier & Kutas, 1999). These and other results (e.g., Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Lau, Almeida, Hines, & Poeppel, 2009) dovetail nicely with the conclusions from behavioral experiments while refining them further, confirming that some on-line language comprehension processes (as indexed by the N400) are acutely sensitive to the goodness of fit between a word and its context (with potentially better fits for more constraining contexts) but less sensitive to the badness of fit. When electrophysiological responses to semantically unrelated words *are* modulated by constraint, such effects often emerge later in the form of an increased parietal positivity, suggesting that strongly constraining sentences may affect processing in several different ways across multiple time points (see Van Petten & Luka, 2012 for a review). Thus, there appear to be both benefits for correct predictions and costs for incorrect predictions, even if they are spatially and temporally distinct.

The existence of both benefits and costs means it may be possible to determine whether comprehenders can predict several words simultaneously. If the language system can make multiple, graded predictions – say, by strongly pre-activating the *sock* lemma and weakly pre-activating the *shirt* lemma until lemma selection – both benefits and costs may be

incurred by the exact same sentence context and completion. However, this may be difficult to determine via methodologies (like those described above) that only permit researchers to assess the effect of constraint on the processing of one word per trial.

It may be easier to measure the different effects of multiple predictions within a single sentence at different times (Federmeier et al., 2007; Van Petten & Luka, 2012). If the language system constantly retunes itself on the basis of the success and failure of past predictions via incremental learning mechanisms (e.g., Chang, Dell, & Bock, 2006; Jaeger & Snider, 2013; Oppenheim, Dell, & Schwartz, 2010; Pickering & Garrod, 2013), the effects of a prediction may be realized before, while, *and after* its success is evaluated. For example, consider a reader who predicts the word “sock” as the completion of the sentence about misplaced laundry. If the word “sock” is indeed presented, the reader may pre-activate its lemma even more strongly in similar contexts in the future, while pre-activating other possible completions (like *shirt*) less than before. Thus, the immediate benefit of prediction – the facilitation of *sock* on the current trial – is distinct from the future benefits and costs that derive from the accuracy (or inaccuracy) of that prediction.

1.2. Assessing downstream effects of prediction using the cumulative semantic interference paradigm

To the best of our knowledge, sentential constraint research has not previously assessed both the predictive benefits and costs incurred by a single complete sentence or measured possible downstream effects of prediction. This study does both by combining highly constraining sentences with a paradigm from language production research, the cumulative semantic interference task (Howard, Nickels, Coltheart, & Cole-Virtue, 2006). Subjects in this paradigm name a sequence of pictures, some of which are members of specific semantic categories (e.g., *sock*, *house*, *cow*, *shirt*, *pen*, *cup*, *church*, *horse*, *glove*). For example, the preceding sequence contains pictures depicting (among other categories) articles of clothing, buildings, and animals. The basic finding is that naming latencies are slower for each extra picture previously named in the same semantic category by an approximately linear amount. So, for example, the second clothing picture (*shirt*) will be named slower than the first (*sock*), the third clothing picture (*glove*) will be named slower than the second by the same amount, and so on. This slowing is referred to as the *cumulative semantic interference effect*.

The inferences about the nature of prediction that can be drawn using this paradigm depend on what causes the interference; as such, it is necessary to describe in detail the models of the effect that have been proposed. According to Howard et al. (2006), the effect arises due to a confluence of three properties: shared activation, competitive lemma selection, and (long-term) priming. When a subject is presented with a picture of a sock, the semantic representation of *sock* (referred to henceforth as <SOCK> according to conventional notation) becomes activated. Some of this activation is shared with proximal semantic representations (e.g., <SHIRT> and <GLOVE>). These representations feed their activation forward to their respective word lemmas in proportion to the strength of their respective semantic-lexical connections. While *sock* is the most active lemma, *shirt* and *glove* are still more active than lemmas in other semantic categories. Howard et al. assume that lemma selection is a competitive process in which each lemma inhibits other lemmas in proportion

to its own activation level; selection is only completed when one lemma (ideally the target, *sock*) reaches an absolute activation threshold. Thus, the duration of lemma selection is shorter when target lemma activation is higher (because it takes fewer timesteps to reach the selection threshold), non-target (i.e., competitor) lemma activation is lower (because the target gets inhibited by competitors less at each timestep), or both.

After *sock* is selected, the connection between its semantic representation <SOCK> and its lexical representation *sock* is strengthened in proportion to that connection's current weight. This increased weight will facilitate the production of *sock* in the future, resulting in repetition priming, as its semantic representation will feed more activation forward to the lexical level the second time around. However, this weight increase has an important side effect: When a subject later attempts to name a different same-category picture (e.g., *shirt*), the semantic representation <SOCK> – which receives shared activation due to its similarity to <SHIRT> – feeds more of that activation forward to the lemma *sock*. Due to the competitive nature of lemma selection, this means that *sock* inhibits *shirt* more than it would have if *sock* had not been named. The extra inhibition slows the selection of *shirt*, leading to cumulative semantic interference. Subsequent pictures (e.g., *glove*) receive even more inhibition because all previously named same-category picture lemmas (*sock*, *shirt*) compete more strongly.

Howard et al. (2006) only described how to account for the semantic interference that accumulates when naming a picture in isolation, but the paradigm used in this study will also assess semantic interference that accumulates when a picture is named at the end of a strongly constraining sentence. Because of its high predictability, this picture's lemma will be selected much faster than after a weakly constraining sentence (Griffin & Bock, 1998) or after no sentence at all. What effect does this shorter duration of processing have on cumulative semantic interference?

In Howard et al.'s model, there is a structural relationship between the duration of lemma selection and the degree of semantic interference that is experienced on the current trial. When lemma selection is faster, competitors have less time to inhibit the target prior to selection. Because cumulative semantic interference in Howard et al.'s model manifests as an increase in inhibition, the effect of this interference on selection will be smaller when the target receives inhibition from competitors for a shorter amount of time – in other words, when selection is fast. As a result, faster lemma selection should lead to less interference.

To illustrate this important point, consider two lemmas: *sock* (high-frequency) and *corset* (low-frequency). If lexical frequency affects a lemma's baseline level of activation, *sock* should have a higher baseline level of activation than *corset*; in turn, *sock* will require less additional activation to reach the selection threshold and can thus be selected more quickly. The inhibition each word receives from competitors will slow the rate at which that word advances toward the selection threshold. If *sock* normally takes half as long to reach the threshold as *corset*, the total amount of inhibition *sock* receives from competitors during selection will be only half as large – and any modulation of that inhibition will affect it half as much. As a result, cumulative semantic interference, which affects how much competitors inhibit the target, will slow selection less when naming *sock* than when naming *corset*.

Given this structural relationship between the duration of lemma selection and the degree of semantic interference, a prediction naturally follows: If a strongly constraining sentence greatly pre-activates its high-cloze word lemma, interference on the current trial should decrease *regardless* of how the sentence affects the activation of non-target word lemmas (as long as lemma selection is still faster overall with the sentence than without).²

An alternative model of cumulative semantic interference was proposed by Oppenheim et al. (2010), who argued that it is not necessary to assume that lemma selection is competitive to account for the effect. According to one version of their model, there is no competition between lemmas; instead, selection proceeds according to a horse race: A word (e.g., *sock*) is selected when its (boosted) activation level crosses a certain threshold. This selection triggers an error-based, incremental learning process in which connections between active semantic representations and the target lemma (e.g., <CLOTHING>-*sock*) are strengthened in inverse proportion to the target's activation level, while connections between active semantic representations and non-target lemmas (<CLOTHING>-*shirt*, <CLOTHING>-*glove*) are weakened in direct proportion to those non-targets' activation levels. That weakening leads to the emergence of cumulative semantic interference: When a subject later names a picture with a weakened connection (e.g., *shirt*), the target lemma receives less activation from the shared semantic representation (<CLOTHING>) and thus has a lower activation level when selection begins than it would have if the subject had not previously selected *sock*. As a result, it takes longer for the target lemma to reach threshold. Selecting *shirt* weakens <CLOTHING>-*glove* for a second time, thereby increasing the amount of time it will take to select *glove* in the future (hence the cumulative nature of the effect).

As with Howard et al. (2006), implementing sentential constraint was outside the scope of Oppenheim et al.'s (2010) model. However, unlike Howard et al.'s model, Oppenheim et al.'s provides a way to tease apart the possible effects of reading strongly constraining sentences. If such sentences increase target lemma activation (*sock*), the target will cross the activation threshold for selection sooner, leading to faster picture naming latencies. Separately, if the sentences cause a net increase in non-target lemma activation (*shirt*, *glove*), the connections between the shared semantic representation (<CLOTHING>) and those lemmas will be weakened more (a consequence of the fact that reweighting is proportional to the activation of those lemmas). As this weakening causes semantic interference on future trials in which the non-target lemmas become targets, increasing non-target lemma activation ultimately leads to a larger cumulative semantic interference effect. Conversely, if strongly constraining sentences cause a net decrease in non-target lemma activation, the amount of cumulative semantic interference will decrease. If there is no change in the net activation of non-target lemmas, the amount of cumulative semantic interference will not change.³

²For another model in which cumulative semantic interference results from competition during lemma selection, see Belke (2013). We will not consider this model here in detail because key aspects have not yet been computationally implemented; however, because it claims that competition is integral to the existence of cumulative semantic interference, we tentatively conclude that Belke (2013) should predict the same relationship between lemma selection duration and interference as Howard et al. (2006).

³In another version of their model, Oppenheim et al. (2010) account for cumulative semantic interference by positing competition during lemma selection accompanied by weight strengthening, but no weight weakening. Given the structural similarity between this version of their model and Howard et al.'s (2006), we believe it would likely share that model's predictions regarding the effects of sentential constraint.

The question of which model best accounts for cumulative semantic interference is relevant to several other theoretical issues, including the nature of lemma selection and of the learning mechanism that gives rise to the interference. If Howard et al. (2006) are correct, words inhibit each other during selection, a form of competition (see also Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992). Furthermore, according to Howard et al., the learning mechanism triggered by selection is not sensitive to the difficulty of selection or to the activation of target and non-target representations, and all weight changes increase connection strengths (i.e., connections are not weakened). In contrast, if Oppenheim et al. (2010) are correct, cumulative semantic interference cannot be taken as evidence that lemma selection is a competitive process; as such, their model is consistent with other, non-competitive accounts (see, e.g., Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Navarrete, Del Prato, Peressotti, & Mahon, 2014). In addition, under their model, learning can both strengthen and weaken connection weights, and the amount of learning that takes place at the interface between semantic and lexical representations is proportional to the amount of error (i.e., the difference between the actual and ideal activation levels for a given lemma while producing the target). In these respects, the learning mechanism is similar to those that have been claimed to operate on syntactic representations (Chang, Dell, & Bock, 2006; Jaeger & Snider, 2013) and to underlie alignment between conversational partners (Pickering & Garrod, 2013), as well as those in a wide variety of non-linguistic domains such as Pavlovian conditioning in rats (Rescorla & Wagner, 1972; see Oppenheim et al., 2010 for other examples). The issues at stake are not limited to language production: The learning mechanism that underlies cumulative semantic interference may extend to any system in which a potential response is selected from a cohort of co-activated responses (Frazer, O'Séaghdha, Munoz-Avila, & Roessler, 2014; Howard et al., 2006; Mulatti, Peressotti, Job, Saunders, & Coltheart, 2012; Oppenheim et al., 2010; Runnqvist, Strijkers, Alario, & Costa, 2012).

1.3. Logic of the present experiments

By combining strongly constraining sentences with the cumulative semantic interference task, it may be possible to jointly inform theoretical issues in both language comprehension and language production, potentially shedding light on two theoretical debates: how prediction affects the activation of expected and unexpected lemmas, and how speakers tune their language systems to optimize them for future acts of production. If strongly constraining sentences speed lemma selection (thereby reducing naming latencies) while leading to decreased semantic interference, it would be consistent with both models of cumulative semantic interference described above. Oppenheim et al. (2010) could claim that sentences increased target lemma activation (hence the lower naming latencies) while decreasing non-target lemma activation (hence the smaller interference effect). Howard et al. (2006) could also claim that sentences likely increased target lemma activation, but could not specify how sentences affected non-target lemma activation. However, if sentences speed picture naming latencies while leading to more semantic interference than in bare picture naming – perhaps because non-target activation lingers as the language system predicts multiple words for the same slot – or leading to the same amount of semantic interference, only Oppenheim et al.'s model could account for the data. As such, it could be used to diagnose the separable effects of sentences on target and non-target lemmas (and,

perhaps, the effects of sentences on category representations as well). Furthermore, such a pattern of data would also provide evidence that it is not necessary to posit competition between lemmas during selection to account for cumulative semantic interference effects, and that the underlying learning mechanism is error-based.

2. Experiment 1

Previous research has shown that naming *sock* in a bare context (i.e., when it is presented in isolation) interferes with subsequently naming *shirt* in a bare context (e.g., Howard et al., 2006). Experiment 1 was designed to determine whether naming *sock* in a sentence context (i.e., after reading a strongly constraining sentence) interferes with naming *shirt* in a sentence context and whether this interference differs in magnitude from that observed in bare contexts.

2.1. Method

2.1.1. Subjects—Eighty members of the University of California, San Diego community participated in Experiment 1. Subjects received class credit for their participation. All reported that they were native English speakers.

2.1.2. Apparatus—Stimuli were presented on an iMac computer running PsyScope X (Build 53; Cohen, MacWhinney, Flatt, & Provost, 1993; <http://psy.ck.sissa.it/>). A Shure SM10A headworn microphone connected to the button box measured voice onset latencies.

2.1.3. Materials—The pictures were 94 line drawings of common objects. Nearly all of these (92) were selected from the International Picture Naming Project picture database (Bates et al., 2003); the other two were found online and drawn in a similar style. Of the 94 pictures, 60 critical pictures formed 12 categories of 5 items each (see Appendix A). All categories, and 90% of the target names in each category, were used by Howard et al. (2006). Categories were chosen to minimize conceptual overlap between them (e.g., there was one category of farm animals, but no categories of fish or shellfish). None of the 34 filler pictures belonged to any of the 12 categories with the possible exception of *igloo*, which may belong to the category <BUILDING>.

For each picture (both critical and filler), a sentence was constructed in which the final word was the name of the picture. Not counting the final word, these sentences varied in length from 6 to 19 words (mean: 11.7 words), generally did not mention any words that belonged to any critical category, and were designed to be strongly constraining such that subjects would primarily use the picture name (or an acceptable alternative) to complete the sentence; e.g., “John turns into a werewolf whenever there is a full *moon*.” To confirm our intuitions regarding sentential constraint, a norming study was conducted in which the sentences for all 60 critical pictures were presented to 100 subjects who did not participate in any primary experiments. Sentences were presented in one of two fixed random orders. Each had its final word removed and was followed by a blank in which subjects were instructed to write “the single word that you think best completes the sentence.” Twenty-three responses that were illegible, were left blank, or consisted of multiple words were discarded; the other 5,977 were scored according to whether they matched an acceptable

name for the target picture using the same criteria that determined naming accuracy in all three experiments. Completions matched acceptable target names 85.6% of the time, with 53 of the 60 critical sentences eliciting higher than 70% name agreement (see Appendix A).

2.1.4. Design—Subjects were presented with two blocks of 94 trials each (188 trials total) in an order determined by one of 20 stimulus lists; however, as only the first block was designed to examine the effects of strongly constraining sentences on non-target activation, all analyses and further description of counterbalancing will be restricted to the first block.⁴ Experiment 1 included three factors of interest: ordinal position (the picture was the first, second, third, fourth or fifth member of a semantic category), context (the picture was presented either in isolation or after a strongly constraining sentence), and their interaction. Each subject named all of the pictures in half of the categories, as well as half of the fillers, in each context. In other words, for each category, all five pictures were presented in isolation or all five pictures were presented after sentences. On both critical and filler sentence trials, the picture was always preceded by the matching sentence. Across subjects, every critical picture was presented eight times in every combination of ordinal position and context.

In keeping with past experiments that used the cumulative semantic interference paradigm (e.g., Howard et al., 2006), the number of intervening pictures between category members (*lag*) was manipulated as well. Within each category, pictures at consecutive ordinal positions were separated by 2, 4, 6 or 8 intervening pictures, with each of those lags represented once per category. Thus, the first and fifth picture from a category were always separated by exactly 23 intervening pictures (e.g., 6+1+2+1+4+1+8). Each category was assigned to one of twelve unique lag sequences (out of a possible 24; 4!). Each subject named three pictures at each combination of lag and ordinal position (not counting the first picture presented in each category, which by definition does not have a lag); however, as lag has previously been shown not to affect cumulative semantic interference (Howard et al., 2006) – at least, when short lags are present (Schnur, 2014) – lag was not systematically manipulated across pictures or categories. The order of categories was counterbalanced such that across the 80 subjects, each category was presented 8 times in each of 10 positions.

2.1.5. Procedure—Each trial began with a cue, presented in 24-point Times New Roman font for 1000 ms, that indicated whether the next picture would be presented in isolation (a cross: +) or after a strongly constraining sentence (five ampersands: &&&&&). After the offset of the cue, a 750 ms delay was followed by either the picture (on bare trials) or the first word in the sentence (on sentence trials). On sentence trials, the sentence was displayed one word at a time using rapid serial visual presentation, with each word presented for 285 ms and followed immediately by the next word. The last printed word was followed

⁴The second blocks of both Experiments 1 and 2 were designed to measure the effect of context on the degree of repetition priming to test a prediction on which the models of Howard et al. (2006) and Oppenheim et al. (2010) were believed to differ. In the second block of Experiment 1, all pictures were presented in bare contexts. In the second block of Experiment 2, half of the filler pictures and all of the critical pictures from half of the categories were presented in bare contexts; the other pictures were presented after weakly constraining sentence fragments (e.g., “The next image you will see is a ____”). In Experiment 1, subjects generally took longer to name pictures that were previously named in strongly constraining sentence contexts; in Experiment 2, the reverse pattern was obtained. As these inconsistent results do not directly bear on the question of how strongly constraining sentences affect non-target activation, they will not be discussed further.

immediately by the picture. On both bare and sentence trials, the picture was displayed until the voice key registered a response. All stimuli were presented in the center of the screen.

After the subject responded, the experimenter coded both the accuracy of the vocal response (according to a list of acceptable picture names) and (when appropriate) the presence of a voice key error, which arose when the microphone mistakenly recorded a response that was earlier or later than the actual onset of the name of the picture. The next trial began 1500 ms after the experimenter coded the response.

Subjects were not familiarized with the materials beforehand and practiced only one trial of each type before beginning. However, the first six trials of the experiment always contained filler pictures.

2.1.6. Analyses—The same data analysis procedure was used for all three experiments. Trials were excluded when a subject provided an inappropriate name for the picture, when the voice key was not triggered at response onset (e.g., due to overt disfluencies or microphone errors), or when the subject responded faster than 300 ms or slower than 3000 ms.

Prior to analysis, the remaining data were transformed to approximate a normal distribution. A Box-Cox test (Box & Cox, 1964) performed on models fitted separately to all usable data from each context from each experiment revealed that the mean lambda values were -1.06 for bare trials and -0.72 for sentence trials; for consistency, inverse RTs (corresponding to a lambda value of -1) were used for every experiment (Ratcliff, 1993). In addition, following Baayen and Milin (2010), the resulting values were multiplied by -10000 so that the model coefficients would have the same sign as if they had been fitted to untransformed data and would be large enough to allow the models to converge.⁵ For example, reaction times of 800 and 1200 ms were transformed into $-10000 * (800^{-1}) = -12.50$ and $-10000 * (1200^{-1}) = -8.33$, respectively. (Note that the latter value is still larger than the former, so effects that increase reaction time will still have positive slopes.)

The transformed data for each experiment were submitted to a mixed-effects model (Baayen, Davidson, & Bates, 2008). In general, ordinal position (1–5; a continuous variable), context (Bare or Sentence; represented in tables as Context_n), and the interaction between ordinal position and context were always included as fixed factors of theoretical interest (though not for all analyses in Experiment 3). To account for potential switch costs between conditions (e.g., see Belke, 2013, Experiment 4), the context of trial $n-1$ (Bare or Sentence; represented in tables as Context_{n-1}) and its interaction with the context of trial n were also included as covariates. Furthermore, to ensure that effects of ordinal position did not simply reflect experiment-wide slowing, trial number was included as a continuous covariate (Alario & Moscoso del Prado Martín, 2010).

Subjects, semantic categories, and pictures were treated as random factors. Whenever possible, a maximal random effects structure was used (Barr, Levy, Scheepers, & Tily,

⁵In addition to this multiplier, predictors were sometimes linearly scaled to facilitate model convergence. All reported data show de-scaled estimates and standard errors.

2013) in which every fixed main effect and interaction (except trial number, due to convergence issues) was allowed to vary by every random factor. If a model did not converge, all random slopes of covariates were removed to facilitate convergence; such exceptions are noted. Finally, to identify outliers, separate models were initially fit to the data from bare and sentence contexts for each experiment to ensure equivalent data retention across conditions. Each model contained all factors listed above (except trial n context and its interactions, as contexts were considered separately). As recommended by Baayen and Milin (2010), data points with absolute standardized residuals greater than 2.5 standard deviations were removed as outliers. All models reported here were fit on the remaining data.

In accordance with common practice for large data sets, t values are treated as z values for the purpose of determining statistical significance (cf. Baayen, 2008). As such, absolute t values greater than or equal to 1.96 are taken to be significant; t values greater than or equal to 1.65 but less than 1.96 are taken to be marginally significant. All predictors were centered.

To determine whether decisions regarding RT transformations and covariate inclusions affected the results, three models were fit for every analysis. One set of models included all effects listed above and was fit using inverse RTs; these are reported for each experiment. A second set of models included the same fixed and random effects but was fit using untransformed (raw) RTs. A third set of models included only effects of theoretical interest (generally the main effects of ordinal position and context as well as their interaction, each varied by all random factors) and was fit using inverse RTs. The statistical significances of nearly every fixed effect of theoretical interest were identical across all three sets of analyses; all exceptions are noted in the text.

2.2. Results

Two pictures (*drill* and *speaker*) were removed from analyses for all experiments due to extremely high error rates (49% and 34% naming errors, respectively, across the first blocks of experiments without a pre-exposure phase). Trials on which these pictures were named are omitted entirely from further discussion and trial counts.

The 80 subjects provided data for 4,640 trials, of which 87.1% (4,043) were analyzed. Trials were excluded when a subject provided an inappropriate name for the picture (136 from bare contexts, 85 from sentence contexts) or when the voice key was not triggered at response onset (303). Trials were also excluded when a subject responded to the picture faster than 300 ms (8) or slower than 3000 ms (15), or if the naming latency was determined to be an outlier according to the exclusion procedure described above (87). (Note that some trials violated multiple criteria.)

The model is summarized in Table 1. Transformed subject means for theoretically relevant conditions are shown in Figure 1, and untransformed means are shown in Figure B.1. Naming latencies were slower for pictures at higher ordinal positions and for pictures that were named in a bare context, as indicated by significant effects of ordinal position and trial n context, respectively. No other main effects or interactions were significant. In particular,

the interaction between ordinal position and trial n context was not significant, $t = -0.038$; nor was it significant with untransformed RTs, $\beta = -7.28$, $SE = 11.20$, $t = -0.65$, or with inverse RTs and no covariates, $\beta = -0.0064$, $SE = 0.16$, $t = -0.04$.

To determine whether the effect of ordinal position was statistically significant for both bare and sentence contexts, separate models were fit to the data from each context. Each model contained the same fixed and random effects as the model reported above (except trial n context and its interactions). The effect of ordinal position was significant for both bare contexts, $\beta = 0.18$, $SE = 0.067$, $t = 2.65$, and sentence contexts, $\beta = 0.16$, $SE = 0.059$, $t = 2.67$.

Although these effects were not significantly different from each other, the effect of ordinal position in sentence contexts (approximately 8.8 ms per position) was only 59% as large as the effect of ordinal position in bare contexts (approximately 15.0 ms per position). Thus, it is conceivable that a majority of subjects or items showed larger effects of ordinal position in bare contexts than in sentence contexts but that the data were simply too noisy for a significant interaction to emerge. This possibility will be addressed (and, to foreshadow the results, dismissed) in the meta-analysis.

2.3. Discussion

Experiment 1 showed that semantic interference accumulates within semantic categories regardless of whether named pictures are presented in isolation or after highly constraining sentences. Furthermore, the amount of interference that accumulated in the two contexts did not significantly differ despite the fact that pictures presented in sentence contexts were named 208 ms faster than pictures presented in bare contexts.

As noted, a model in which cumulative semantic interference arises due to inhibition during lemma selection (Howard et al., 2006) is hard-pressed to account for a pattern of data in which a manipulation (sentential context) facilitates lemma selection but does not decrease the quantity of interference (although one possible way to do so is considered in Experiment 3). In contrast, Oppenheim et al.'s (2010) error-based learning model can account for the data provided that strongly constraining sentences increase target lemma activation while having no net effect on non-target lemma activation. Thus, these data favor an account of lexical prediction in which the language system only maintains the pre-activation of a single word: the most likely candidate. Furthermore, they support a model of cumulative semantic interference in which the interference does not arise as the result of inhibition between lemmas during selection, and in which the learning mechanism that tunes connection weights between semantic and lexical representations is error-based.

3. Experiment 2

In Experiment 1, pictures in sentence contexts were not only named faster than pictures in bare contexts, they were also named more accurately (as evidenced by error rates of 3.7% and 5.9%, respectively). This is most likely because subjects confronted with pictures that they would be unable to name in isolation were able to use semantic information provided by the sentence to constrain the range of possible responses. As the analyses were restricted

to correct responses, this could have caused more difficult pictures to be included in analyses more often in sentence contexts than bare contexts. In addition, because sentences were always predictive of the following object, subjects may have been able to occasionally prepare a response in advance and simply produce it when the picture appeared. (The existence of an equivalent cumulative semantic interference effect within sentence contexts indicates that this represents an unlikely possibility, but it is a possibility nonetheless.)

To alleviate these potential problems, the design of Experiment 2 differed from the design of Experiment 1 in two key ways. First, to reduce the difference in naming accuracy between conditions, subjects were familiarized with the pictures and their correct names before the experiment. Second, to ensure that subjects had to process the pictures before naming them, the cue validity of sentences was decreased by presenting filler pictures after mismatching sentences (e.g., “Matt couldn’t open the lock because he was using the wrong” followed by the picture *sandwich*).

3.1. Method

3.1.1. Subjects—Sixty new subjects from the same population as Experiment 1 participated in Experiment 2.

3.1.2. Apparatus—The same apparatus was used as in Experiment 1, with two differences: The experimental software used was PsyScope X Build 57 (Cohen et al., 1993; <http://psy.ck.sissa.it/>), and the microphone was connected to the button box indirectly via a Marantz PMD661 voice recorder.

3.1.3. Materials—All pictures and sentences were identical to those used in Experiment 1. However, on filler trials only, the pairings between pictures and sentences were different (see below for details).

3.1.4. Design—Prior to the experiment, subjects were familiarized with all 94 pictures in an order determined by one of 60 stimulus lists, which counterbalanced the order of critical pictures. On each familiarization trial, a fixation point was presented for 1000 ms, the screen remained blank for 750 ms, and then a picture was presented. After the voice key registered a response, the name of the picture was presented in 30-point Times New Roman font immediately below the picture for 1500 ms. Trials were separated by a 1000-ms inter-trial interval. Although subjects were instructed to name each picture as it was presented and to use the written name for the rest of the experiment, the same criteria were used to determine naming accuracy as in Experiment 1.

During the experiment, subjects were presented with two blocks of 94 trials each (188 trials total) in an order determined by one of 60 stimulus lists; however, as with Experiment 1, the analysis and description of counterbalancing will be restricted to the first block (see Footnote 4). The only difference from Experiment 1 was that on 30 of the 34 filler trials, the picture was paired with a strongly constraining sentence corresponding to a different filler picture. The first six trials of each list, which were always fillers, included two of these mismatched sentence trials, two matching sentence trials, and two bare trials.

As in Experiment 1, there were three factors of interest: ordinal position, trial n context, and their interaction. Each subject named all of the pictures in half of the categories in each context. Across subjects, every critical picture was presented six times in every combination of ordinal position and context and six times in every combination of lag and context, and each category was presented 10 times in each of 6 positions.

3.1.5. Procedure—After the familiarization phase, the procedure was identical to that used in Experiment 1. Pre-trial cues distinguished between bare and sentence trials, but did not identify whether the sentence would match the subsequent picture or not.

3.2. Results

The 60 subjects provided data for 3,480 trials, of which 95.8% (3,335) were analyzed. Trials were excluded when a subject provided an inappropriate name for the picture (29 from bare contexts, 14 from sentence contexts) or when the voice key was not triggered at response onset (52). Trials were also excluded when a subject responded to the picture faster than 300 ms (1) or slower than 3000 ms (1), or if the naming latency was determined to be an outlier according to the exclusion procedure described in Experiment 1 (57). (Note that some trials violated multiple criteria.)

The model is summarized in Table 2. Transformed subject means for theoretically relevant conditions are shown in Figure 2, and untransformed means are shown in Figure B.2. Naming latencies were slower for pictures at higher ordinal positions, pictures that were named in a bare context, and pictures presented later in the block, as indicated by significant effects of ordinal position, trial n context, and trial number, respectively. No other main effects or interactions were significant. In particular, the interaction between ordinal position and trial n context was not significant, $t = -0.054$; nor was it significant with untransformed RTs, $\beta = -4.31$, $SE = 9.69$, $t = -0.44$, or with inverse RTs and no covariates, $\beta = -0.0096$, $SE = 0.18$, $t = -0.05$. Separate models fit to the data from each context indicated that the effect of ordinal position was significant for both bare contexts, $\beta = 0.18$, $SE = 0.069$, $t = 2.54$, and sentence contexts, $\beta = 0.14$, $SE = 0.060$, $t = 2.32$.

3.3. Discussion

As expected, adding a pre-exposure phase and mismatching filler sentences reduced the effect of context (from 208 ms to 129 ms) and decreased the error rate (from 4.8% to 1.2%) relative to Experiment 1. However, these changes did not affect the results: Experiment 2 replicated Experiment 1, as semantic interference accumulated within semantic categories in both bare and sentence contexts and the amount of interference in the two contexts did not significantly differ.

4. Experiment 3

Experiments 1 and 2 showed that naming *sock* in a sentence context interferes with naming *shirt* in a sentence context as much as naming *sock* in a bare context interferes with naming *shirt* in a bare context, suggesting that strongly constraining sentences increased target lemma activation while having no net effect on non-target lemma activation. However, neither experiment tested for transfer of interference between contexts. In other words, does

naming *sock* in a sentence context interfere with naming *shirt* in a bare context, and vice-versa? How does the amount of interference transferred across contexts compare to that accumulated within each context?

Given the results of Experiments 1 and 2, models of cumulative semantic interference make different predictions with respect to how much should transfer. According to Oppenheim et al.'s (2010) model of cumulative semantic interference, if presenting *sock* in a sentence context does not change the activation of *shirt* (at the time of selection) relative to presenting *sock* in a bare context, the <CLOTHING>-*shirt* connection should be weakened by the same amount in both contexts. This means that when *shirt* is subsequently produced, the amount of semantic interference is the same regardless of the contexts in which either *sock* or *shirt* are presented. Thus, interference should transfer equally across contexts.

As noted, Howard et al.'s (2006) model cannot straightforwardly account for equivalent interference effects in bare and sentence contexts because increasing the activation of a target presented after a sentence should be accompanied by a reduction in semantic interference, which was not observed. However, there is at least one modification that could potentially allow Howard et al. to account for these results. One distinguishing feature of their model, relative to Oppenheim et al.'s (2010), is that the mechanism that reweights the connections between a target's semantic representation and its lexical representation always does so by the same amount, multiplying the existing weight by a fixed parameter that is invariant to target (or non-target) lemma activation. In theory, this parameter could be allowed to vary based on context, so that naming a picture after a strongly constraining sentence would strengthen the relevant semantic-lexical connection more than naming that picture in isolation. This would cause pictures named in sentence contexts to be stronger competitors on future trials than pictures named in bare contexts, leading to more semantic interference. With the right parameter values, that could potentially cancel out the decrease in interference that results from increasing target activation (the other effect of sentence presentation), thereby causing *sock* named in a sentence context to interfere with *shirt* named in a sentence context just as much as *sock* named in a bare context interferes with *shirt* named in a bare context.

If such an account is correct, this adapted version of Howard et al.'s (2006) model would predict that *sock* should compete more strongly with *shirt* when *sock* was named after a strongly constraining sentence than when *sock* was named in isolation, regardless of which condition *shirt* is named in. This contrasts with the prediction of Oppenheim et al. (2010) that, if strongly constraining sentences do not affect non-target lemma activation, naming *sock* should slow the naming of *shirt* equally regardless of which condition either picture is named in.

Experiment 3 was designed to address these discrepant predictions by presenting different pictures from the same semantic category in bare and sentence contexts; e.g., presenting *sock* in a bare context and *shirt* in a sentence context, and vice-versa. This makes it possible to determine whether naming latencies for *shirt* are modulated not only by the number of previously named same-category members, but also by the contexts in which those previous same-category members were named. If semantic interference is greater when generated by

pictures named in sentence contexts than pictures named in bare contexts, Howard et al. (2006) can account for the results of Experiments 1 and 2, which would make it impossible to tease apart the effect of sentential constraint on target and non-target lemmas. If the magnitude of semantic interference is unaffected by its source, only Oppenheim et al.'s (2010) model can account for the results, which would reinforce the previous conclusions regarding the nature of lemma selection and the learning mechanisms underlying the interference, as well as the effects of constraint on target and non-target activation.

4.1. Method

4.1.1. Subjects—Eighty new subjects from the same population as Experiments 1 and 2 participated in Experiment 3.

4.1.2. Apparatus—The same apparatus was used as in Experiment 1.

4.1.3. Materials—All pictures and sentences were identical to those used in Experiments 1 and 2. As in Experiment 1, all filler sentences were paired with their matching pictures.

4.1.4. Design—Subjects were presented with two blocks of 94 trials each (188 trials total) in an order determined by one of 20 stimulus lists. However, the analysis and description of counterbalancing will be restricted to the first block for two reasons. First, this will increase comparability with Experiments 1 and 2. Second, those experiments revealed that the context in which a picture was named in Block 1 (inconsistently) affected naming latencies in Block 2 (see Footnote 4). Given that Block 1 naming context and Block 2 naming context were confounded in Experiment 3 – pictures presented in a bare context in Block 1 were presented in a sentence context in Block 2, and vice-versa – it would be difficult to interpret data from Block 2.

In Experiments 1 and 2, all five pictures within a category were presented in the same context to a given subject. In Experiment 3, four pictures within a category were presented in one context (the “standard” condition), and the other picture – presented either at ordinal position 3 or ordinal position 4 – was presented in the opposite context (the “deviant” condition). This design, borrowed from Navarrete, Mahon, and Caramazza (2010, Experiment 3), makes it possible to determine whether naming a picture in a bare trial slows the subsequent naming of a same-category picture in a sentence trial and vice-versa by breaking down the overall effect of ordinal position into separate contributions from previously named pictures in bare and sentence contexts. Of the 12 categories presented to each subject, six categories contained four bare pictures and one sentence picture, and six categories contained four sentence pictures and one bare picture; within each group, the deviant picture was presented at the third ordinal position in three categories and at the fourth ordinal position in three categories. Half of the fillers were presented in each context. Across subjects, every critical picture was presented equally often in every ordinal position, and each category was presented 8 times in each of 10 positions.

4.1.5. Procedure—The procedure was identical to that used in Experiment 1.

4.2. Results

Different facets of the data were examined using three separate analyses. First, an analysis was conducted to determine whether pictures named in standard (i.e., non-deviant) conditions showed effects of ordinal position (e.g., whether naming *sock* in a bare context slows the subsequent naming of *shirt* in a bare context, and whether naming *sock* in a sentence context slows the subsequent naming of *shirt* in a sentence context). This analysis (henceforth the “standard analysis”), which is largely comparable to those presented in Experiments 1 and 2, included pictures named in bare contexts for which a majority of previously named same-category pictures were also named in bare contexts, and pictures named in sentence contexts for which a majority of previously-named same-category pictures were also named in sentence contexts.

Second, an analysis was conducted to directly compare pictures named in deviant and standard conditions to determine whether semantic interference generated by previously named pictures was modulated by the context in which those pictures were named (e.g., whether *shirt* is named equally slowly regardless of the context in which *sock* was previously named). This analysis (henceforth the “deviancy analysis”) was restricted to pictures named at the third and fourth ordinal positions, which are the only two positions at which deviant pictures were presented. If, as Howard et al. (2006) must predict to account for the pattern of data observed in Experiments 1 and 2, pictures named in sentence contexts generate more semantic interference on subsequent trials than pictures named in bare contexts, this comparison should reveal an interaction between deviancy and context: Among pictures named in bare contexts, deviant trials (preceded by same-category pictures named in sentence contexts) should be slower than standard trials, whereas among pictures named in sentence contexts, deviant trials (preceded by same-category pictures named in bare contexts) should be faster than standard trials.

Finally, an analysis was conducted to determine whether naming latencies in bare and sentence contexts are equally slowed by previously named pictures (e.g., whether naming *sock* in a bare context slows the naming of *shirt* equally in bare and sentence contexts, and whether naming *sock* in a sentence context slows the naming of *shirt* equally in bare and sentence contexts). This analysis (henceforth the “omnibus analysis”) included all data from all conditions.

4.2.1. All analyses—The 80 subjects provided data for 4,640 trials. Of these, 92.8% (4,307) were eligible for inclusion in the analyses below. Trials were excluded when a subject provided an inappropriate name for the picture (97 from bare contexts, 47 from sentence contexts) or when the voice key was not triggered at response onset (190). Trials were also excluded when a subject responded to the picture faster than 300 ms (9) or slower than 3000 ms (20). Outliers were identified and excluded separately for each analysis via the same method used for Experiments 1 and 2.

4.2.2. Standard analysis—The goal of this analysis, which was restricted to pictures named in standard conditions, was to replicate the results of Experiments 1 and 2 by determining whether interference accumulates within each context. As in those experiments,

the factors of interest were ordinal position, trial n context, and their interaction; covariates included trial $n-1$ context, its interaction with the trial n context, and trial number. All fixed effects of interest were allowed to vary by all random factors (subjects, semantic categories, and pictures), but due to convergence issues, covariates were not included in the random effects structure.

Of the 3,728 trials on which a picture was named in a standard condition, 91.3% (3,402) were included in this analysis. Trials were excluded if they failed to meet the criteria listed above (262) or if the naming latency was determined to be an outlier according to the exclusion procedure described in Experiment 1 (64).

The model is summarized in Table 3. Transformed subject means for theoretically relevant conditions are shown in Figure 3, and untransformed means are shown in Figure B.3. (The data in this analysis are represented in the figures by the points linked by solid best-fit lines.) Naming latencies were slower for pictures at higher ordinal positions, pictures named in a bare context, and pictures presented later in the block, as indicated by significant effects of ordinal position, trial n context, and trial number, respectively. Furthermore, when the picture on trial $n-1$ was presented in a sentence context instead of a bare context, the picture on trial n was slowed marginally more when it was presented in a bare context than in a sentence context, as indicated by a marginally significant interaction between the contexts of trial $n-1$ and trial n . No other main effects or interactions were significant. In particular, the interaction between ordinal position and trial n context was not significant, $t = 0.36$; nor was it significant with untransformed RTs, $\beta = -7.75$, $SE = 7.39$, $t = -1.05$, or with inverse RTs and no covariates, $\beta = 0.03$, $SE = 0.11$, $t = 0.29$.

Separate models fit to the data from each context indicated that the effect of ordinal position was significant for sentence contexts, $\beta = 0.21$, $SE = 0.070$, $t = 2.99$. It failed to reach significance for bare contexts, $\beta = 0.12$, $SE = 0.079$, $t = 1.47$; however, it was significant both with untransformed RTs, $\beta = 15.00$, $SE = 5.54$, $t = 2.71$, and with inverse RTs and no covariates, $\beta = 0.18$, $SE = 0.077$, $t = 2.37$. The apparent tenuousness of this effect may be due here to the reduction in power from omitting half of the data in the third and fourth ordinal positions (i.e., the deviant conditions). Given that it was observed in every analysis in both Experiments 1 and 2 and trended in the right direction here, the effect of ordinal position within bare contexts is most likely real. Thus, these results should be taken to indicate that in Experiment 3, cumulative semantic interference emerged in both bare and sentence contexts when pictures were named in the same context as most other same-category members. This replicates the results of Experiments 1 and 2.

4.2.3. Deviancy analysis—The goal of this analysis, which was restricted to pictures presented at the third and fourth ordinal positions, was to determine whether pictures named in sentence contexts generate a different amount of semantic interference on subsequent trials than pictures named in bare contexts. The factors of interest were trial n context, deviancy (“deviant” or “standard”), and their interaction; covariates included trial $n-1$ context, its interaction with trial n context, and trial number. (Ordinal position was not included as a factor in this analysis because it was not relevant to the theoretical question under consideration and because the data set was restricted to only two levels of that

continuous variable.) All fixed effects of interest were allowed to vary by all random factors (subjects, semantic categories, and pictures), but due to convergence issues, covariates were not included in the random effects structure.

Of the 1,856 trials on which a picture was named in the third or fourth ordinal position, 91.3% (1,694) were included in this analysis. Trials were excluded if they failed to meet the criteria listed above (132) or if the naming latency was determined to be an outlier according to the exclusion procedure described in Experiment 1 (30).

The model is summarized in Table 4, and the data in this analysis are represented in Figures 3 (transformed means) and B.3 (untransformed means) by the points at the third and fourth ordinal positions. Naming latencies were slower for pictures named in a bare context, as indicated by significant effect of trial n context. No other main effects or interactions were significant, including the main effect of deviancy, $t = -0.57$ (numerically trending toward slower naming latencies for standard trials) and the interaction between deviancy and context, $t = 0.38$ (numerically, relative to deviant trials, standard trials slowed naming latencies more in bare contexts than sentence contexts).

Separate models fit to the data from each context indicated that the effect of deviancy was not significant for either bare contexts, $\beta = -0.29$, $SE = 0.26$, $t = -1.08$, or sentence contexts, $\beta = -0.01$, $SE = 0.34$, $t = -0.03$. The effect on bare contexts was marginally significant when analyzed using untransformed RTs, $\beta = -41.45$, $SE = 22.82$, $t = -1.82$ (trending toward slower naming latencies for standard trials), but not when inverse RTs were used without covariates, $\beta = -0.28$, $SE = 0.26$, $t = -1.08$.

These results indicate that the naming latencies for pictures in a given context are largely unaffected by the context in which same-category pictures were previously named. To the extent that any effect of prior context emerged at all, it was in the direction of slower naming latencies for pictures in bare contexts when more same-category pictures were previously named in bare (as opposed to sentence) contexts; however, this result was only marginally significant, and even then only in one of three analyses. Thus, there was no empirical support for the possibility that naming *sock* in a sentence context causes a greater reweighting of the connection between the semantic representation <SOCK> and the lexical representation *sock*, thereby leading to greater semantic interference than naming *sock* in a bare context. As such, Howard et al. (2006) ultimately cannot account for the observed pattern of data in which strongly constraining sentences facilitate picture naming without modulating semantic interference.

4.2.4. Omnibus analysis—The goal of this analysis, which included all usable data collected in the first block of the experiment, was to determine whether the increase in naming latencies caused by a previously named picture differs depending on the context of a subsequently named same-category picture, and thus whether interference transfers between contexts. In this analysis, ordinal position was replaced by two variables: bare ordinal position and sentence ordinal position. For a given picture, bare ordinal position represents the number of same-category pictures (plus the present one, if applicable) that have been named in a bare context. Similarly, sentence ordinal position represents the number of same-

category pictures (plus the present one, if applicable) that have been named in a sentence context. For example, for a picture presented in a sentence context, if two same-category pictures were previously named in sentence contexts and another was named in a bare context, its bare ordinal position would be 1 and its sentence ordinal position would be 3. (Note that bare ordinal position and sentence ordinal position always sum to ordinal position as defined in Experiments 1 and 2 and in the standard analysis of Experiment 3.) Thus, there were five factors of interest: bare ordinal position, sentence ordinal position, trial n context, the interaction of bare ordinal position and trial n context, and the interaction of sentence ordinal position and trial n context. Covariates included trial $n-1$ context, its interaction with trial n context, and trial number. All fixed effects except trial number were allowed to vary by all random factors (subjects, semantic categories, and pictures).

Of the 4,640 trials in the first block, 91.2% (4,231) were included in this analysis. Trials were excluded if they failed to meet the criteria listed above (333) or if the naming latency was determined to be an outlier according to the exclusion procedure described in Experiment 1 (76). The model is summarized in Table 5, and the data in this analysis are represented by all points in Figures 3 (transformed means) and B.3 (untransformed means). Naming latencies were slower when more pictures were previously named in either bare contexts or sentence contexts, as indicated by significant effects of bare ordinal position and sentence ordinal position, respectively. Naming latencies were also slower when pictures were named in a bare context and were marginally slower for pictures presented later in the block, as indicated by a significant effect of trial n context and a marginally significant effect of trial number, respectively. No other main effects or interactions were significant, including the interactions between bare ordinal position and context, $t = 0.57$, and between sentence ordinal position and context, $t = 0.25$ (both numerically trending toward larger effects of ordinal position for sentence contexts).

Separate models fit to the data from each context indicated that the effect of bare ordinal position was significant for both bare contexts, $\beta = 0.15$, $SE = 0.071$, $t = 2.14$, and sentence contexts, $\beta = 0.25$, $SE = 0.10$, $t = 2.47$. The effect of sentence ordinal position, however, was less consistent. For bare contexts, it was not significant, $\beta = 0.085$, $SE = 0.085$, $t = 0.99$, but this varied across analyses: It reached significance when untransformed RTs were used, $\beta = 11.98$, $SE = 5.89$, $t = 2.03$, and was marginally significant when inverse RTs were used but covariates were excluded, $\beta = 0.17$, $SE = .089$, $t = 1.95$. Given these discrepancies, the lack of significance in the main analysis may have been due to substantial collinearity between the variance accounted for by the measures of bare ordinal position and sentence ordinal position, which correlated at $r = 0.573$. In support of this hypothesis, the full bare context model reported above fit the data significantly better than a model with the same random effects structure and all of the same fixed effects except sentence ordinal position, $\chi^2(1) = 4.55$, $p = .033$, suggesting that the effect of sentence ordinal position accounted for variance above and beyond that explained by the effect of bare ordinal position. For sentence contexts, the effect of sentence ordinal position was significant, $\beta = 0.23$, $SE = 0.10$, $t = 2.22$, but this also varied across analyses: It was not significant when untransformed RTs were used, $\beta = 8.51$, $SE = 5.60$, $t = 1.52$, but reached significance when inverse RTs were used and covariates were excluded, $\beta = 0.22$, $SE = 0.095$, $t = 2.34$. Given the results of the

main analysis and the presence of an effect of sentence ordinal position on sentence contexts in two of three omnibus analyses in Experiment 3 (as well as in all analyses in Experiments 1 and 2), it seems reasonable to conclude that this effect is real as well.

4.3. Discussion

The standard analysis replicated Experiments 1 and 2 in showing that interference accumulates within semantic categories when most pictures are named in the same context, regardless of which context that is. The deviancy analysis demonstrated that, holding the context of the current picture constant, there is no effect of the context in which a same-category picture was previously named. The omnibus analysis showed that, holding the context of the previously-named same-category pictures constant, there is no effect of the context of the current picture on the amount of interference it receives. Collectively, these results indicate that cumulative semantic interference transfers fully between contexts: Naming *sock* slows the subsequent naming of *shirt* equally regardless of the context in which either picture is named.

Experiments 1 and 2, as well as the standard analysis of Experiment 3, showed that sentence contexts facilitate picture naming (in part by speeding lemma selection, a point that will be discussed further in the General Discussion), but that naming *sock* in a sentence context slows the subsequent naming of *shirt* in a sentence context as much as naming *sock* in a bare context slows the subsequent naming of *shirt* in a bare context. To account for this pattern of data, Howard et al. (2006) could claim that picture naming in a sentence context yields more learning, and hence more semantic interference on subsequent trials, than picture naming in a bare context, but that this extra learning is offset by target facilitation. However, this prediction was not borne out. The deviancy analysis showed no effect of prior context on the naming latencies of pictures presented in sentence contexts, and to the extent that any effect of prior context was observed for pictures presented in bare contexts, it was in the opposite direction from the one that Howard et al. would have to predict. Thus, models of cumulative semantic interference in which the interference arises from inhibition during lemma selection are unable to account for the data. This means that strongly constraining sentences increased target lemma activation while having no net effect on non-target lemma activation, as this is the only set of circumstances under which any model of cumulative semantic interference (Oppenheim et al., 2010) can account for the observed pattern of data.

5. Meta-analysis

All data used in the Experiment 1, Experiment 2, and Experiment 3 omnibus analyses were combined to increase statistical power to detect a potential interaction between ordinal position and context. The 220 subjects provided data for 12,760 trials, of which 91.0% (11,609) were analyzed. (Trials were included if and only if they were included in the appropriate single-experiment analysis; no additional trimming procedures were performed.) The model included all fixed effects and random effects reported for Experiments 1 and 2, as well as an additional random factor – experiment – that, like all other random factors, was crossed with every fixed effect except trial number.

The model is summarized in Table 6. Transformed subject means for theoretically relevant conditions are shown in Figure 4, and untransformed means are shown in Figure B.4. Naming latencies were slower for pictures at higher ordinal positions, pictures that were named in a bare context, pictures presented later in the block, and pictures for which the previous trial included a sentence context, as indicated by significant effects of ordinal position, trial n context, trial number, and trial $n-1$ context, respectively. No interactions were significant. In particular, the interaction between ordinal position and trial n context was not significant, $t = 0.017$; nor was it significant with untransformed RTs, $\beta = -6.95$, $SE = 8.03$, $t = -0.87$, or with inverse RTs and no covariates, $\beta = -0.0042$, $SE = 0.12$, $t = -0.036$. Separate models fit to the data from each context indicated that the effect of ordinal position was significant for both bare contexts, $\beta = 0.17$, $SE = 0.055$, $t = 3.05$, and sentence contexts, $\beta = 0.16$, $SE = 0.040$, $t = 4.07$.

Although no interaction was observed between ordinal position and context despite naming latencies being 191 ms faster in sentence contexts than bare contexts (based on the untransformed RT analysis), the interference effect was still numerically smaller (by 7 ms) in sentence contexts. To determine whether an interaction could be detected using non-parametric statistics, an interference effect was computed for every combination of subject and context from the data in the untransformed RT analysis. Of the 220 subjects, 116 subjects (52.7%) showed larger semantic interference effects in bare contexts than in sentence contexts, a number that did not significantly differ from chance (50%) according to an exact binomial test, $p = 0.46$. The same result was obtained when effects were computed for each picture, as 32 of 58 pictures (55.2%) showed larger semantic effects in bare contexts than in sentence contexts, $p = 0.51$. Thus, these non-parametric statistics reinforce the conclusion that the effect of ordinal position was not modulated by context.

6. General Discussion

6.1. Summary

In three experiments, subjects named pictures presented either in isolation or after strongly constraining sentences. Every experiment showed the same pattern of data: Naming a picture in a sentence context slowed the subsequent naming of a different same-category picture in a sentence context as much as if both pictures were named in bare contexts. Furthermore, as Experiment 3 showed, this cumulative semantic interference fully transferred between contexts. In short, producing “sock” slowed the subsequent production of “shirt” by the same amount regardless of the context in which either picture was presented.

A model of cumulative semantic interference claiming that the interference emerges from inhibition during lemma selection (Howard et al., 2006) cannot account for such a pattern of data in its present form. According to this model, increasing target activation should simultaneously facilitate lemma selection and reduce the overall influence of non-target lemma activation on selection. If, as Howard et al. claim, cumulative semantic interference emerges because non-target lemmas compete more strongly, speeding lemma selection should yield less interference (as long as non-target lemmas still have nonzero activation and thus generate a nonzero amount of interference). Contrary to this prediction, although

sentence contexts facilitated lemma selection, the amount of interference was the same in bare and sentence contexts. Furthermore, one potential assumption that could allow Howard et al. to account for the results – that naming pictures in sentence contexts yields more learning than naming pictures in bare contexts – was not supported by the Experiment 3 deviancy analysis.

In contrast, Oppenheim et al. (2010) can accommodate the data very straightforwardly. According to their model, the fact that subjects named pictures faster in sentence contexts could mean that target lemma activation was increased, and the equivalent semantic interference between contexts means non-target lemma activation at the moment of lemma selection was the same as in bare contexts. This in turn means that the activation of relevant category representations (e.g., <CLOTHING>) was likely also the same between contexts, because changes in category activation should have affected the activation of non-target lemmas. These results support the conclusion that strongly constraining sentences in the present experiments had a narrow scope of facilitation, and are consistent with other behavioral and electrophysiological studies demonstrating no effect of sentential constraint on the processing of incongruous words (Federmeier et al., 2007; Kutas & Hillyard, 1984; Lau et al., 2009; Stanovich & West, 1979, 1981). Furthermore, by supporting Oppenheim et al.'s account, the data lead to several other conclusions about the nature of word production: It is not necessary to assume that lemma selection is a competitive process to account for cumulative semantic interference effects, and the learning that underlies this interference is error-based, bringing it in line with learning mechanisms that have been hypothesized to operate at other levels of linguistic representation and in other domains entirely (e.g., Chang et al., 2006; Jaeger & Snider, 2013; Pickering & Garrod, 2013; Rescorla & Wagner, 1972).

6.2 Necessary assumptions

The logic of the experiments, according to which the model of cumulative semantic interference proposed by Howard et al. (2006) was ruled out, depends on the assumption that strongly constraining sentences facilitate processing during lemma selection. If, instead, sentences only facilitate stages of picture naming that occur before or after lemma selection (e.g., visual recognition or phoneme selection), the duration of lemma selection should be unaffected and hence it should show the same amount of interference regardless of naming context (which it does); thus, Howard et al. could account for the results and no conclusions could be drawn about the effects of sentential constraint on target and non-target lemma activation. In contrast, if at least some of the facilitation from strongly constraining sentences speeds lemma selection, the experimental logic is valid.

Prior research supports the assumption that strongly constraining sentences affect lemma selection. As Griffin and Bock (1998) noted, sentential constraint affects performance even in object-identification tasks that do not require phonological wordform retrieval (e.g., Kroll, 1990), and a recent study showed that strongly constraining sentences can actually cause speakers to initiate lexical access of a picture before it is presented (Piai, Roelofs, & Maris, 2014). A dual-task experiment has also shown that a constraint manipulation in a picture naming task affects the processing of a second, unrelated task, which it would likely not do if the stage of word production facilitated by constraint occurred after phoneme selection

(Ferreira & Pashler, 2002). Furthermore, the fact that effects of sentential constraint on picture naming are modulated by the frequency of the picture name (Griffin & Bock, 1998; Piai et al., 2014) indicates that constraint does not solely affect early, pre-lexical processing. Thus, the assumption that constraint affects lemma selection is valid, as are the inferences that depend on that assumption.

A second key assumption inherent in these experiments is that the effects of sentential constraint and picture presentation on lemma activation are additive rather than interactive. If presenting a to-be-named picture were to ‘wash out’ the effect of a preceding sentence on the activation of underlying representations, our conclusions about the nature of constraint would necessarily be more circumscribed. Although the effect of constraint on picture processing in a comprehension task is similar to its effect on word processing (Ganis, Kutas, & Sereno, 1996; Nigam, Hoffman, & Simons, 1992), this possibility cannot be fully ruled out at present. Given that constraint affects lemma selection, though, our conclusions about models of cumulative semantic interference would be unchanged, as the models were evaluated according to whether or not they could account for an observed pattern of data.

6.3. Is the scope of facilitation too narrow?

It is not surprising that the results suggest a narrow scope of facilitation for strongly constraining sentences, but generally speaking, the scope may seem *too* narrow. Kutas and Hillyard (1984) and Fedemeier et al. (1999) showed that strongly constraining sentences may facilitate the processing of low-cloze words that are semantically related to the best completion. The pictures in the present experiments clearly shared enough semantic features to co-activate each other during naming, which is what gave rise to the interference in the first place. However, despite their semantic relatedness, either the sentence that preceded the *sock* picture did not increase the activation of *shirt*, or – if it did – the language system did not maintain that extra activation as the sentence continued to unfold. This result seems at odds with those of Kutas and colleagues. What accounts for the discrepancy?

One possibility is that task demands caused our subjects to behave differently than they would during other comprehension tasks. In Experiments 1 and 3, every sentence was followed by a matching picture, making it so that subjects only needed to predict a single word to correctly perform picture naming, their central task (assuming their best prediction was correct). Although filler pictures in Experiment 2 were preceded by mismatching sentences, these mismatches were so absolute that it would have been impossible for subjects to correctly predict anything about the target picture, so pre-activating multiple potential picture names still would not have been productive. In contrast, everyday reading materials likely afford fewer sentences that are so strongly predictive of a single word, and predictions during normal comprehension are less likely to be as correct or as immediately useful as they were here. Thus, it may be premature to draw conclusions on the basis of these data about the scope of predictions that comprehenders make during everyday reading, except insofar as to say that they *can* maintain single-word predictions when it seems reasonable to do so.

A second, more intriguing possibility is that the degree of semantic relatedness required to generate or maintain joint pre-activation after a strongly constraining sentence may be

greater than that required to generate cumulative semantic interference. Consistent with this hypothesis, the categories used by Federmeier and Kutas (1999) were defined relatively narrowly. For example, they considered cars, public transportation, and aircraft to be three different categories, whereas a single, broader semantic category in the present experiments (transport) contained exemplars from all three. Similarly, dishes and utensils, which constituted two separate categories in their study, were merged here within the tableware category. Readers are clearly sensitive to fine-grained categorical distinctions, as evidenced by the fact that a sentence for which the best completion was *palms* reduced the N400 elicited by another tree name (*pinus*) but not by another (non-tree) plant name (*tulips*). Thus, it may be that the categories in the present experiments were not defined narrowly enough for the strongly constraining sentences to elicit and maintain the activation of multiple words.

A third possibility is that the sentences pre-activated the target lemmas via concepts that were not shared by other same-category members. To take a particularly extreme example, the sentence that preceded the picture *lighthouse* (a building) was “The sailors narrowly avoided being shipwrecked thanks to the beacon on the ____.” Although the target word was produced more than three times as often than any other completion in the norming study, subjects produced 26 other unique completions as well, not one of which was a building (“ship”, “shore”, “horizon”, “left”, etc.). Given that the most contextually relevant feature of the target was its function (beacon-bearer) rather than its form (building), it would not be surprising if the preceding sentence failed to pre-activate other buildings. In contrast, the sentence that preceded the picture *finger*, which belonged to the category of body parts, was, “Reaching to pick up the files, Lauren got a paper cut on her ____.” In addition to the target, subjects produced three other unique completions, all of which were body parts (“hand”, “thumb”, “head”). This suggests that the sentence for *finger* may have been more likely to activate category-level semantic features, pre-activating the names of other body parts and, in turn, generating cumulative semantic interference.

To determine whether the amount of interference that accumulated within each category could be traced back to the degree of co-activation among category members, we computed, for each picture in the sentence norming study, the proportion of responses that did not match the target but belonged to the same semantic category as the target (see Appendix A). This proportion was averaged for the pictures within each category to determine that category’s *shared feature activation*, and a median split on this proportion determined whether categories had a low or high shared feature activation (means: 3.1% and 10.2%, respectively). Using the data from the untransformed RT analysis, a cumulative semantic interference effect was computed for each combination of category and context within each experiment, then averaged across experiments and within median split bins. Unsurprisingly, when pictures were named in bare contexts, shared feature activation – a measure derived from responses to sentences – had only a small effect on interference: Categories with low and high shared feature activation generated 14.5 ms and 17.5 ms of interference per ordinal position, respectively. When pictures were named in sentence contexts, however, this difference was even smaller: Categories with low and high shared feature activation generated 10.4 and 10.3 ms of interference per ordinal position, respectively. (The numeric

difference between interference effects for each context is consistent with the estimates of the non-significant interaction term from the untransformed RT analyses reported for each experiment individually and in the meta-analysis, though as noted, the numbers reported here represent simple category means.) This indicates a total lack of support for the hypothesis that variation in the activation of category-relevant semantic features can explain the absence of additional semantic interference for pictures named in sentence contexts.

This is curious because all models of cumulative semantic interference predict that increasing the activation of shared semantic features, and thus of same-category non-target lemmas, should increase interference. It is true that shared feature activation as defined here is inextricably confounded with target response probability (they cannot sum to greater than 1) and that it represents at best a rough measure of the underlying construct, so this conclusion should be taken with a grain of salt. Nevertheless, it raises the possibility that if these strongly constraining sentences do indeed pre-activate word lemmas and that activation is maintained until selection, the incremental learning mechanism responsible for the cumulative semantic interference effect may be able to distinguish between different sources of activation. As such, it could essentially “tag” the activation that originated with the sentence and leave it out of the equation altogether when reweighting connections. However, this scenario would directly conflict with recent research suggesting that “conceptual-lexical links to targets and competitors are modulated during picture naming regardless of the intrinsic or extrinsic basis of their co-activation” (Frazer et al., 2014).

A fourth possibility is that the invariance of semantic interference to shared feature activation could be taken to indicate that lemmas were not pre-activated at all, and that the facilitation of naming latencies in sentence contexts simply reflected increased ease of integration (e.g., Hagoort et al., 2009). However, this too seems unlikely, as unambiguous evidence suggests that comprehenders can pre-activate words from strongly constraining sentence contexts (e.g., DeLong, Urbach, & Kutas, 2005). The logic of the present experiments holds as long as *some* of the context benefit from sentences is attributable to lemma pre-activation (i.e., as long as that benefit does not solely reflect easier integration) because the extra activation conferred by the sentence should interact with the mechanisms that underlie semantic interference even though the source of pre-activation is external to the picture (Frazer et al., 2014).

More work is needed to determine which of these scenarios (if any) accurately explains why the scope of facilitation was narrower than might have been expected. As previous research has shown, many experimental parameters may affect the nature of predictions that comprehenders make, including task demands, the distribution of cloze probabilities, and the extent to which likely non-target words belong to the same category as the target. The degree of cumulative semantic interference should be sensitive to these manipulations as well. For example, if comprehenders strategically adjust the breadth of their predictions, less constraining sentences should lead to broader predictions, which should in turn generate more interference. Future research along these lines would help determine when comprehenders maintain single-word vs. multiple-word predictions, and thus whether the present results represent the rule or the exception.

6.4. Conclusions

Comprehenders generate expectations about upcoming speech and text, pre-activating words that might come next. They may even make multiple, graded predictions for the same word slot, though each failed prediction comes with a cost that may not be evident until later (Van Petten & Luka, 2012). The cumulative semantic interference paradigm potentially provides a way to evaluate multiple predictions for the same exact sentence by assessing both the immediate benefits and long-term costs of successful and failed predictions, respectively. In the present experiments, sentential constraint facilitated picture naming but did not modulate semantic interference, suggesting that the only net effect of strongly constraining sentences was to increase target activation. Thus, comprehenders can maintain single-word predictions of upcoming language when it seems reasonable to do so.

Acknowledgments

D.K. and V.S.F. were supported by the NIH under Grant number R01 HD051030 (PI: Ferreira), and D.K. was additionally supported by Grant number R01 NIDCD011492 (PI: Gollan). E.R. was supported by a postdoctoral grant from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement n° PIEF-GA-2012-329339. We are grateful to Gary Oppenheim for helpful discussions, to two anonymous reviewers for feedback on an earlier version of this manuscript, and to Sarah Bae, Kristi Cheng, Cameron Hays, Elizabeth Jimenez, Danielle Lew, Cheryl Ma, Jessica Ma, Gabriela Meckler, Brittany Nielsen, Shivani Patel, and Kurina Wolff for data collection.

References

- Alario FX, Moscoso del Prado Martín F. On the origin of the “cumulative semantic inhibition” effect. *Memory & Cognition*. 2010; 38:57–66.10.3758/MC.38.1.57 [PubMed: 19966239]
- Ashby J, Rayner K, Clifton CJ. Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology*. 2005; 58A: 1065–1086.10.1080/02724980443000476 [PubMed: 16194948]
- Baayen, RH. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press; 2008.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory & Language*. 2008; 59:390–412.10.1016/j.jml.2007.12.005
- Baayen RH, Milin P. Analyzing reaction times. *International Journal of Psychological Research*. 2010; 3:12–28.
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68:255–278.10.1016/j.jml.2012.11.001
- Bates E, D’Amico S, Jacobsen T, Székely A, Andonova E, Devescovi A, Tzeng O. Timed picture naming in seven languages. *Psychonomic Bulletin & Review*. 2003; 10:344–380.10.1016/j.jml.2004.03.002 [PubMed: 12921412]
- Belke E. Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*. 2013; 69:228–256.10.1016/j.jml.2013.05.008
- Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*. 1964; 26:211–252.
- Chang F, Dell GS, Bock K. Becoming syntactic. *Psychological Review*. 2006; 113:234–272.10.1037/0033-295X.113.2.234 [PubMed: 16637761]
- Cohen J, MacWhinney B, Flatt M, Provost J. *PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers*. *Behavior Research Methods, Instruments, & Computers*. 1993; 25:257–271.10.3758/BF03204507
- Collins AM, Loftus EF. A spreading-activation theory of semantic processing. *Psychological Review*. 1975; 82:407–428.10.1037/0033-295X.82.6.407

- Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93:283–321.10.1037/0033-295X.93.3.283 [PubMed: 3749399]
- Dell GS, Chang F. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2014; 369(1634):20120394.10.1098/rstb.2012.0394
- DeLong KA, Urbach TP, Kutas M. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*. 2005; 8:1117–1121.10.1038/nn1504
- Ehrlich SF, Rayner K. Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*. 1981; 20:641–655.10.1016/S0022-5371(81)90220-6
- Federmeier KD, Wlotko EW, De Ochoa-Dewald E, Kutas M. Multiple effects of sentential constraint on word processing. *Brain Research*. 2007; 1146:75–84.10.1016/j.brainres.2006.06.101 [PubMed: 16901469]
- Federmeier KD, Kutas M. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*. 1999; 41:469–495.10.1006/jmla.1999.2660
- Ferreira VS, Pashler H. Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:1187–1199.10.1037/0278-7393.28.6.1187
- Fischler IS, Bloom PA. Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*. 1979; 18:1–20.10.1016/S0022-5371(79)90534-6
- Fischler IS, Bloom PA. Effects of constraint and validity of sentence contexts on lexical decisions. *Memory & Cognition*. 1985; 13:128–139.10.3758/BF03197005 [PubMed: 4033415]
- Forster KI. Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. 1981; 33:465–495.10.1080/14640748108400804
- Frazer, AK.; O'Séaghdha, PG.; Munoz-Avila, H.; Roessler, N. Competitor activation and semantic interference: Evidence from combined phonological and semantic similarity. In: Bello, P.; Guarini, M.; McShane, M.; Scassellati, B., editors. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; 2014. p. 1108-1113.
- Ganis G, Kutas M, Sereno MI. The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*. 1996; 8:89–106.10.1162/jocn.1996.8.2.89 [PubMed: 23971417]
- Griffin ZM, Bock K. Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*. 1998; 38:313–338.10.1006/jmla.1997.2547
- Hagoort, P.; Baggio, G.; Willems, RM. Semantic unification. In: Gazzaniga, MS., editor. *The Cognitive Neurosciences*. 4. Boston, MA: MIT Press; 2009. p. 819-936.
- Howard D, Nickels L, Coltheart M, Cole-Virtue J. Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*. 2006; 100:464–482.10.1016/j.cognition.2005.02.006 [PubMed: 16413014]
- Jaeger TF, Snider NE. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*. 2013; 127:57–83.10.1016/j.cognition.2012.10.013 [PubMed: 23354056]
- Kroll JF. Recognizing words and pictures in sentence contexts: A test of lexical modularity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1990; 16:747–759.10.1037/0278-7393.16.5.747
- Kutas M, Hillyard SA. Reading senseless sentences: Brain potential reflect semantic incongruity. *Science*. 1980; 207:203–205.10.1126/science.7350657 [PubMed: 7350657]
- Kutas M, Hillyard SA. Brain potential during reading reflect word expectancy and semantic association. *Nature*. 1984; 307:161–163.10.1038/307161a0 [PubMed: 6690995]
- Lau E, Almeida D, Hines PC, Poeppel D. A lexical basis for N400 context effects: Evidence from MEG. *Brain & Language*. 2009; 111:161–172.10.1016/j.bandl.2009.08.007 [PubMed: 19815267]

- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 1999; 22:1–39.10.3115/992628.992631 [PubMed: 11301520]
- Mahon BZ, Costa A, Peterson R, Vargas KA, Caramazza A. Lexical selection is not by competition: A reinterpretation of interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007; 33:503–535.10.1037/0278-7393.33.3.503
- Mulatti C, Peressotti F, Job R, Saunders S, Coltheart M. Reading aloud: The cumulative lexical interference effect. *Psychonomic Bulletin & Review*. 2012; 19:662–667.10.3758/s13423-012-0269-z [PubMed: 22623265]
- Myers JL, O'Brien EJ. Accessing the discourse representation during reading. *Discourse Processes*. 1998; 26:131–157.10.1080/01638539809545042
- Navarrete E, Del Prato P, Peressotti F, Mahon BZ. Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*. 2014; 76:253–272.10.1016/j.jml.2014.05.0030749-596X/ [PubMed: 25284954]
- Navarrete E, Mahon BZ, Caramazza A. The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*. 2010; 134:279–289.10.1016/j.actpsy.2010.02.009 [PubMed: 20347062]
- Nigam A, Hoffman JE, Simons RF. N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*. 1992; 4:15–22.10.1162/jocn.1992.4.1.15 [PubMed: 23967854]
- Oppenheim GM, Dell GS, Schwartz MF. The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*. 2010; 114:227–252.10.1016/j.cognition.2009.09.007 [PubMed: 19854436]
- Piai V, Roelofs A, Maris E. Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*. 2014; 53:146–156.10.1016/j.neuropsychologia.2013.11.014 [PubMed: 24291513]
- Pickering MJ, Garrod S. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*. 2013; 36:377–392.10.1017/S0140525X12003238 [PubMed: 24049786]
- Ratcliff R. Methods for dealing with reaction time outliers. *Psychological Bulletin*. 1993; 114:510–532.10.1037/0033-2909.114.3.510 [PubMed: 8272468]
- Rescorla, RA.; Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH.; Prokasy, WF., editors. *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts; 1972. p. 64-99.
- Roelofs A. A spreading-activation theory of lemma retrieval in speaking. *Cognition*. 1992; 42:107–142.10.1016/0010-0277(92)90041-F [PubMed: 1582154]
- Runqvist E, Strijkers K, Alario FX, Costa A. Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*. 2012; 66:850–869.10.1016/j.jml.2012.02.007
- Schnur TT. The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*. 2014; 75:27–44.10.1016/j.jml.2014.04.006
- Schuberth RE, Eimas PD. Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*. 1977; 3:27–36.10.1037/0096-1523.3.1.27
- Schwanenflugel PJ, LaCount KL. Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1988; 14:344–354.10.1037/0278-7393.14.2.344
- Schwanenflugel PJ, Shoben EJ. The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*. 1985; 24:232–252.10.1016/0749-596X(85)90026-9
- Stanovich KE, West RF. Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*. 1979; 7:77–85.10.3758/BF03197588
- Stanovich KE, West RF. The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*. 1981; 7:658–672.10.1037/0096-1523.7.3.658

- Szewczyk JM, Schriefers H. Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*. 2013; 68:297–314.10.1016/j.jml.2012.12.002
- Taylor WL. “Cloze procedure”: A new tool for measuring readability. *Journalism Quarterly*. 1953; 30:415–433.
- Van Berkum, JJA. The neuropragmatics of “simple” utterance comprehension: An ERP review. In: Sauerland, U.; Yatsushiro, K., editors. *Semantics and pragmatics: From experiment to theory*. Basingstoke, UK: Palgrave Macmillan; 2009. p. 276-316.
- Van Berkum JJA, Brown CM, Zwitserlood P, Kooijman V, Hagoort P. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2005; 31:443–467.10.1037/0278-7393.31.3.443
- Van Petten C, Luka BJ. Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*. 2012; 83:176–190.10.1016/j.ijpsycho.2011.09.015 [PubMed: 22019481]
- Wicha NYY, Bates EA, Moreno EM, Kutas M. Potato not pope: Human brain potential to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*. 2003; 346:165–168.10.16/S0304-3940(03)00599-8 [PubMed: 12853110]
- Wicha NYY, Moreno EM, Kutas M. Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*. 2003; 39:483–508.10.1016/S0010-9452(08)70260-0 [PubMed: 12870823]
- Wicha NYY, Moreno EM, Kutas M. Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*. 2004; 16:1272–1288.10.1162/0898929041920487 [PubMed: 15453979]

Appendix A. Critical Stimuli Used in Experiments 1–3

Semantic category	Picture	Sentence frame	Exact match	Concept match	Other same-category responses
Audio-visual equipment	headphones	When he wanted to listen to music, Heath took out his iPod and plugged in his	0.60	0.89	0.08
	microphone	When Allison sings in her room, she uses her hairbrush as a	1.00	1.00	0.00
	radio	Vince always finds the best music when browsing stations on the	0.91	0.91	0.00
	speaker	At the rock concert, the bass boomed from the huge	0.78	0.80	0.09
	television	Flipping through the channels, Katherine couldn't find anything good to watch on	1.00	1.00	0.00
Body parts	ear	After swimming, Jaclyn tried to shake out the	0.73	0.73	0.22

Semantic category	Picture	Sentence frame	Exact match	Concept match	Other same-category responses
		water stuck in her			
	eye	Eric preferred glasses because he always had a hard time putting a contact lens in his right	0.99	0.99	0.01
	finger	Reaching to pick up the files, Lauren got a paper cut on her	0.84	0.84	0.16
	hand	As a lefty, Kevin held the pencil in his left	1.00	1.00	0.00
	nose	Rudolph the Reindeer's most prominent feature is his bright red	1.00	1.00	0.00
Buildings	castle	In medieval times, a moat deterred would-be attackers from sieging a	0.66	0.66	0.04
	church	Every Sunday, the whole family attended Mass at the local	0.97	0.97	0.03
	house	When she finally had enough money, Sarah moved to the suburbs and bought her own	0.86	0.86	0.10
	lighthouse	The sailors narrowly avoided being shipwrecked thanks to the beacon on the	0.39	0.39	0.00
	windmill	It is impossible to travel more than a few miles in the Dutch countryside without seeing a	0.20	0.20	0.02
Celestial phenomena	cloud	The sky was completely blue except for one fluffy white	1.00	1.00	0.00
	lightning	The dogs howled and ran indoors because of the thunder and	0.85	0.85	0.15
	moon	John turns into a werewolf whenever there is a full	1.00	1.00	0.00

Semantic category	Picture	Sentence frame	Exact match	Concept match	Other same-category responses
Clothes	rainbow	Leprechauns are known for hiding a pot of gold at the end of a	0.99	0.99	0.00
	star	Alyssa makes a wish whenever she looks up and sees a shooting	1.00	1.00	0.00
	glove	While dancing, Michael Jackson was famous for wearing a single sequined	0.76	0.76	0.24
	jacket	After stepping into the cold night air, Liz ran back inside to grab a	0.52	0.94	0.04
	pants	Because he had eaten too much over winter break, Owen had a hard time fitting into a pair of	0.73	0.99	0.01
Farm animals	skirt	As part of her schoolgirl uniform, Kristin was required to wear a white pleated	0.72	0.72	0.28
	sock	After doing his laundry, Mark always seemed to be missing one	0.99	0.99	0.01
	cow	On the class field trip, the students got to milk a	0.92	0.92	0.04
	donkey	In the story of Winnie the Pooh, Eeyore is the name of the gloomy	0.86	0.89	0.06
	horse	Colin went to the Kentucky Derby and bet \$100 on his favorite	0.89	0.93	0.00
	pig	In the book Charlotte's Web, a spider becomes friends with a	0.88	0.89	0.01
	sheep	Wool is made from the fleece of a	0.86	0.96	0.01
Furniture	bed	After traveling abroad for weeks, Nathan couldn't wait to	0.98	0.98	0.00

Semantic category	Picture	Sentence frame	Exact match	Concept match	Other same-category responses
		get home and sleep in his own			
	chair	When thinking about a difficult problem, Zach would often rock back in his	0.90	0.90	0.06
	desk	When the students finished their exams, they were to place them on the teacher's	0.99	0.99	0.01
	stool	Since there wasn't an open booth, Blake sat at the counter on a	0.71	0.76	0.15
House parts	table	After cooking supper, Jan set the food on the dinner	0.99	0.99	0.00
	balcony	On sunny days, Ron likes to go up to the second floor and read a book outside on his	0.54	0.82	0.11
	chimney	Children believe that Santa brings them presents by climbing down the	0.97	0.97	0.02
	door	When Jessica knocked, Max opened the front	1.00	1.00	0.00
	roof	To replace the fallen shingles, Emily climbed a ladder to get to the	0.78	0.79	0.03
Musical instruments	window	The room was getting stuffy, so Cody opened a	0.97	0.97	0.03
	drum	As a member of the school band, Steven played the snare	0.81	0.81	0.16
	guitar	Hailey and her friends sang around the campfire while she played her	0.88	0.88	0.11
	piano	Marissa's family has a 9-foot Steinway grand	0.72	0.72	0.00
	trumpet	Louis Armstrong was famous for his	0.21	0.21	0.76

Semantic category	Picture	Sentence frame	Exact match	Concept match	Other same-category responses
Tableware	violin	ability to sing and play the Joshua lifted the bow high into the air and touched it to the strings of his	0.47	0.48	0.03
	cup	When Lilah was little, she drank her juice from a sippy	0.99	0.99	0.01
	fork	To eat the pasta, Tina wound it around her	0.95	0.95	0.03
	glass	Because it was so fragile, Katie was always extra careful when washing the wine	0.92	0.92	0.05
	knife	Since Ricky ordered steak, the server brought him an extra sharp	0.99	0.99	0.00
	spoon	Because his guests were going to have soup, Alex made sure to give each of them a	0.96	0.96	0.02
Tools	ax	It is rumored that George Washington chopped down a cherry tree with his	0.94	0.94	0.01
	drill	Kurt wanted to hang a painting, so he bored a hole with a	0.29	0.30	0.17
	hammer	To help with his woodworking, Travis bought some nails and a	0.94	0.94	0.06
	saw	The carpenter cut the wood in half using a	0.85	0.85	0.14
	screwdriver	To open the vent, Diane needed a flathead	0.82	0.82	0.12
	bus	The football team traveled to the away game together on a	0.93	0.93	0.03
Transport	car	When Rita ran out of gas, she had to ask people to help push her	0.97	0.98	0.01

Semantic category	Picture	Sentence frame	Exact match	Concept match	Other same-category responses
	helicopter	Due to the severity of the accident, the skier was airlifted off the mountain by a	0.95	0.96	0.04
	plane	The nice flight attendant gave the boy an extra bag of pretzels on the	0.77	0.77	0.00
	truck	As a driver for a shipping company, Jeremy logged thousands of miles every year in his	0.60	0.63	0.21

Note. “Exact match” represents the proportion of responses that matched the name given in the “Picture” column, “Concept match” represents the proportion that matched an acceptable name for the picture (e.g., “jeans” for “pants”), and “Other same-category responses” represents the proportion of responses that belonged to the same category as the target but did not match the concept.

Appendix B. Figures with Untransformed Means **Figure B.1.**

Experiment 1 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an untransformed RT scale. Model fits for linear effects of ordinal position are denoted by best-fit lines. Error bars represent 95% confidence intervals.

Figure B.2.

Experiment 2 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an untransformed RT scale. Model fits for linear effects of ordinal position are denoted by best-fit lines. Error bars represent 95% confidence intervals.

Figure B.3.

Experiment 3 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an untransformed RT scale. Each point represents a unique combination of bare ordinal position (the number of pictures previously named in bare contexts, plus the current one if applicable), sentence ordinal position (the equivalent measure for sentence contexts), and context. Number pairs next to each point represent its (bare, sentence) ordinal positions. Categories for which four of five pictures were named in bare contexts are denoted by points with black numbers (Bare (Standard) and Sentence (Deviant) conditions); categories for which four of five pictures were named in sentence contexts are denoted by points with grey numbers (Sentence (Standard) and Bare (Deviant) conditions). Model fits for linear effects of ordinal position are denoted by solid best-fit lines (for standard points only). Error bars represent 95% confidence intervals.

Figure B.4.

Experiment 1–3 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an untransformed RT scale. Model fits for linear effects of ordinal position are denoted by best-fit lines. Error bars represent 95% confidence intervals.

Highlights

- We assess the immediate and “downstream” effects of prediction.
- Pictures were named in isolation or after strongly constraining sentences.
- Semantic interference accumulated at the same rate in both contexts.
- Cumulative semantic interference is not caused by competition during word selection.
- When appropriate, comprehenders predict only the most probable completion for each sentence.

Figure 1.

Experiment 1 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an inverse scale to match statistical analyses, such that a fixed vertical distance along the Y-axis always represents the same difference in inverse RT space. Model fits for linear effects of ordinal position are denoted by best-fit lines. Error bars represent 95% confidence intervals.

Figure 2. Experiment 2 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an inverse scale to match statistical analyses. Model fits for linear effects of ordinal position are denoted by best-fit lines. Error bars represent 95% confidence intervals.

Figure 3.

Experiment 3 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an inverse scale to match statistical analyses. Each point represents a unique combination of bare ordinal position (the number of pictures previously named in bare contexts, plus the current one if applicable), sentence ordinal position (the equivalent measure for sentence contexts), and context. Number pairs next to each point represent its (bare, sentence) ordinal positions. Categories for which four of five pictures were named in bare contexts are denoted by points with black numbers (Bare (Standard) and Sentence (Deviant) conditions); categories for which four of five pictures were named in sentence contexts are denoted by points with grey numbers (Sentence

(Standard) and Bare (Deviant) conditions). Model fits for linear effects of ordinal position are denoted by solid best-fit lines (for standard points only). Error bars represent 95% confidence intervals.

Figure 4.

Experiment 1–3 picture naming latencies shown as a function of ordinal position and context. Latencies are averaged and presented on an inverse scale to match statistical analyses. Model fits for linear effects of ordinal position are denoted by best-fit lines. Error bars represent 95% confidence intervals.

Table 1

Experiment 1 results and effect sizes derived from a mixed-effects model (N trials = 4,043).

Fixed effect	β	Approximate effect size (ms)	SE	t
Intercept	-12.98	770.27	0.36	-35.99 *
Trial number	0.0029	0.17	0.0021	1.37
Context _n (Sentence)	-3.44	-208.05	0.28	-12.18 *
Ordinal position	0.18	10.50	0.067	2.65 *
Ordinal position * Context _n	-0.0063	-6.18	0.17	-0.038
Context _{n-1} (Sentence)	0.23	13.45	0.23	1.00
Context _{n-1} * Context _n	0.055	-3.95	0.47	0.12

Note. Naming latencies were inversely transformed and multiplied by -10,000. All fixed effects except trial number were allowed to vary by all random factors (subjects, semantic categories, pictures). Approximate effect sizes represent slopes computed on back-transformed condition means, and may not always have the same sign as their corresponding beta estimates for non-significant effects. Words in parentheses denote factor levels for which positive beta estimates would indicate slower naming latencies.

[†] $p < .10$.

* $p < .05$.

Table 2
Experiment 2 results and effect sizes derived from a mixed-effects model (N trials = 3,335).

Fixed effect	β	Approximate effect size (ms)	SE	t
Intercept	-14.73	678.77	0.30	-49.07 *
Trial number	0.0065	0.30	0.002	3.22 *
Context _n (Sentence)	-2.77	-128.87	0.27	-10.18 *
Ordinal position	0.15	7.13	0.04	3.91 *
Ordinal position * Context _n	-0.0095	-3.19	0.18	-0.054
Context _{n-1} (Sentence)	0.26	11.71	0.25	1.04
Context _{n-1} * Context _n	-0.41	-23.48	0.54	-0.76

Note. Naming latencies were inversely transformed and multiplied by -10,000. All fixed effects except trial number were allowed to vary by all random factors (subjects, semantic categories, pictures). Approximate effect sizes represent slopes computed on back-transformed condition means, and may not always have the same sign as their corresponding beta estimates for non-significant effects. Words in parentheses denote factor levels for which positive beta estimates would indicate slower naming latencies.

[†] $p < .10$.

* $p < .05$.

Table 3
Experiment 3 standard analysis results and effect sizes derived from a mixed-effects model (N trials = 3,402).

	β	Approximate effect size (ms)	SE	t
Intercept	-13.50	740.96	0.38	-35.75 *
Trial number	0.0076	0.42	0.0025	3.01 *
Context _n (Sentence)	-3.78	-212.79	0.48	-7.86 *
Ordinal position	0.16	9.03	0.051	3.20 *
Ordinal position * Context _n	0.04	-2.92	0.11	0.36
Context _{n-1} (Sentence)	0.054	2.96	0.11	0.49
Context _{n-1} * Context _n	-0.38	-24.41	0.22	-1.75 †

Note. Naming latencies were inversely transformed and multiplied by -10,000. Fixed effects of context_n, ordinal position, and their interaction were allowed to vary by all random factors (subjects, semantic categories, pictures). Approximate effect sizes represent slopes computed on back-transformed condition means, and may not always have the same sign as their corresponding beta estimates for non-significant effects. Words in parentheses denote factor levels for which positive beta estimates would indicate slower naming latencies.

† $p < .10$.

* $p < .05$.

Table 4

Experiment 3 deviancy analysis results and effect sizes derived from a mixed-effects model (N trials = 1,856).

	β	Approximate effect size (ms)	SE	t
Intercept	-13.32	750.66	0.42	-31.66 *
Trial number	-0.0015	-0.083	0.0041	-0.36
Context _n (Sentence)	-3.89	-224.76	0.45	-8.59 *
Deviancy (Deviant)	-0.13	-7.07	0.22	-0.57
Deviancy * Context _n	0.17	14.43	0.44	0.38
Context _{n-1} (Sentence)	0.14	7.71	0.19	0.73
Context _{n-1} * Context _n	-0.47	-33.22	0.38	-1.23

Note. Naming latencies were inversely transformed and multiplied by -10,000. Fixed effects of context_n, deviancy, and their interaction were allowed to vary by all random factors (subjects, semantic categories, pictures). Approximate effect sizes represent slopes computed on back-transformed condition means, and may not always have the same sign as their corresponding beta estimates for non-significant effects. Words in parentheses denote factor levels for which positive beta estimates would indicate slower naming latencies.

[†] $p < .10$.

* $p < .05$.

Table 5

Experiment 3 omnibus analysis results and effect sizes derived from a mixed-effects model (N trials = 4,231).

	β	Approximate effect size (ms)	SE	t
Intercept	-13.41	745.63	0.39	-34.43 *
Trial number	0.0048	0.27	0.0025	1.94 †
Context _n (Sentence)	-3.78	-215.32	0.31	-12.27 *
Bare ordinal position	0.21	11.66	0.072	2.87 *
Bare ordinal position * Context _n	0.12	0.30	0.21	0.57
Sentence ordinal position	0.19	10.69	0.072	2.64 *
Sentence ordinal position * Context _n	0.056	-2.95	0.22	0.25
Context _{n-1} (Sentence)	0.11	6.02	0.25	0.44
Context _{n-1} * Context _n	-0.42	-28.49	0.49	-0.85

Note. Naming latencies were inversely transformed and multiplied by -10,000. All fixed effects except trial number were allowed to vary by all random factors (subjects, semantic categories, pictures). Approximate effect sizes represent slopes computed on back-transformed condition means, and may not always have the same sign as their corresponding beta estimates for non-significant effects. Words in parentheses denote factor levels for which positive beta estimates would indicate slower naming latencies.

† $p < .10$.

* $p < .05$.

Table 6

Experiment 1–3 results and effect sizes derived from a mixed-effects model (N trials = 11,609).

Fixed effect	β	Approximate effect size (ms)	SE	t
Intercept	-13.73	728.57	0.62	-22.21 *
Trial number	0.0057	0.31	0.0012	4.85 *
Context _n (Sentence)	-3.33	-179.68	0.17	-19.41 *
Ordinal position	0.17	8.99	0.056	3.04 *
Ordinal position * Context _n	0.0036	-4.30	0.22	0.017
Context _{n-1} (Sentence)	0.21	10.93	0.06	3.44 *
Context _{n-1} * Context _n	-0.25	-19.49	0.33	-0.76

Note. Naming latencies were inversely transformed and multiplied by -10,000. All fixed effects except trial number were allowed to vary by all random factors (experiment, subjects, semantic categories, pictures). Approximate effect sizes represent slopes computed on back-transformed condition means, and may not always have the same sign as their corresponding beta estimates for non-significant effects. Words in parentheses denote factor levels for which positive beta estimates would indicate slower naming latencies.

[†] $p < .10$.

* $p < .05$.